

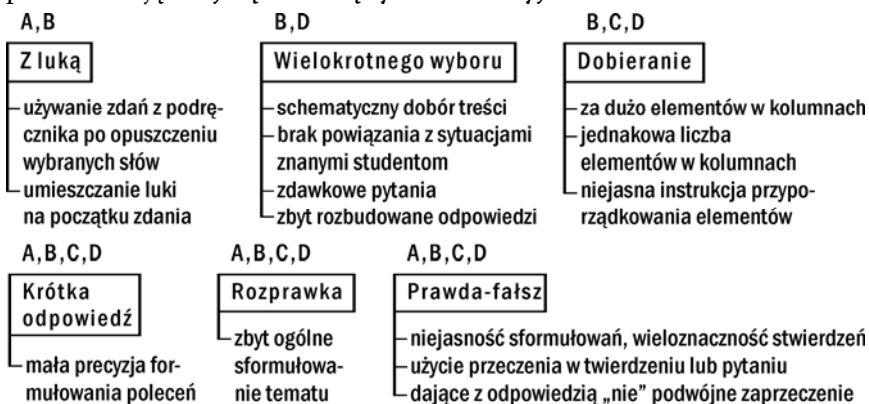
Przygotowanie narzędzi pomiaru dydaktycznego

Ocenianie jest procesem złożonym – przebiega w kilku etapach i poprzedzone jest dodatkowymi czynnościami:

- przygotowaniem wymagań programowych, których spełnienie będzie oceniane i mających służyć jako układ odniesienia dla wyniku uczenia się,
- ustaleniem norm wymagań,
- zbudowaniem skali osiągnięć studentów.

Ważne jest **powiązanie typu zadania pisemnego z jego przydatnością w realizacji określonych celów**. Zadania, które student otrzymuje do rozwiązania, mają na celu sprawdzenie poziomu jego zapamiętania (A) lub zrozumienia (B) wiedzy, umiejętności jej zastosowania w sytuacjach typowych (C) lub nietypowych (D)* [1,17]. Zapamiętanie to nie tylko umiejętność powtórzenia treści (reguły, zasady, prawidłowości itp.), lecz również niemylenie pojęć i elementarny poziom zrozumienia. Zrozumienie to przedstawienie wiadomości w innej formie, niż były początkowo podane, proste wnioskowanie, porządkowanie i streszczanie problemów. Stosowanie wiadomości w sytuacjach typowych polega na wykorzystaniu uprzednio zaobserwowanych wzorców do celów niezbyt odległych od przypadku szkoleniowego. Stosowanie wiadomości w sytuacjach nietypowych to formułowanie problemów i wyciąganie z nich wniosków: oryginalność, nowatorstwo i samodzielność.

Poniższy schemat (Rys. 5) prezentuje powyższe zależności; dodatkowo przedstawiając najczęstsze błędy konstrukcyjne:



Rys. 5. Powiązanie typu zadania pisemnego z jego przydatnością w realizacji określonych celów oraz z typowymi błędami konstrukcyjnymi

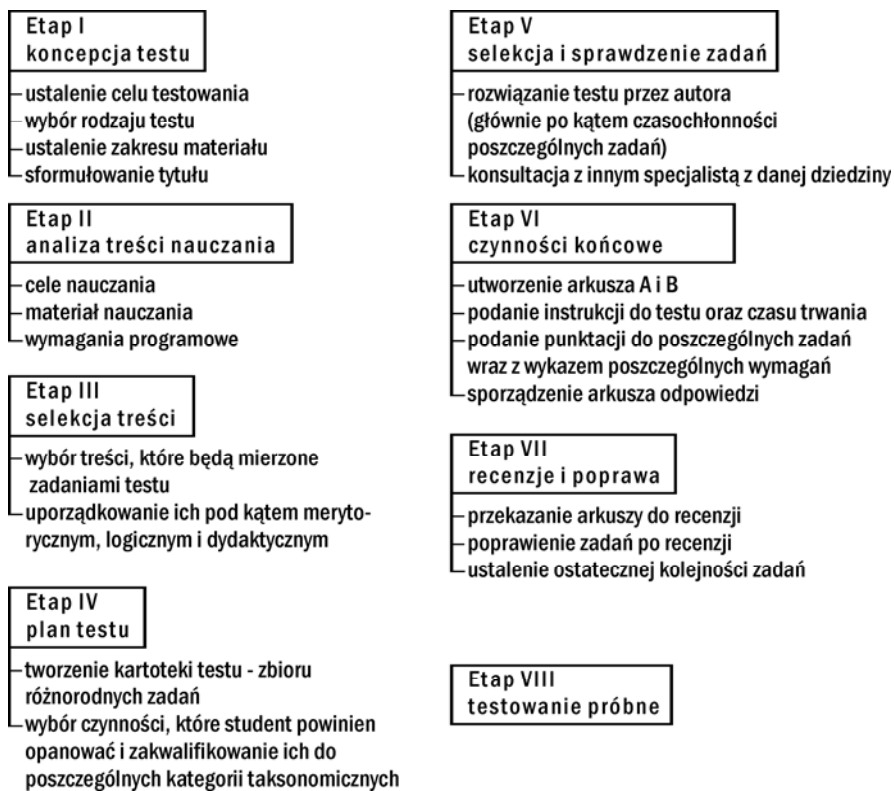
*Tzw. taksonomia celów nauczania ABC Niemierki.

Sprawdziany/kolokwia konstruuje się w różny sposób – w zależności od ich przeznaczenia. Wiąże się z tym ogromna różnorodność typów testów i różnorodność wymagań egzaminacyjnych związanych z nimi, co przedstawiono w końcowej części rozdziału „Dla chcących wiedzieć więcej” (Rys. 8). Szczególnie istotne wydaje się być rozróżnienie testów sprawdzających i różnicujących.

Testy sprawdzające wskazują na stopień spełnienia wymagań edukacyjnych przez studenta. Odmianą tego typu testu jest test diagnozujący dający wstępną informację czy studenci posiadają podstawową wiedzę, potrzebną do realizacji dalszej części materiału.

Test różnicujący pozwala na ułożenie studentów w listę rankingową. Uwypukla różnice pomiędzy studentami w testowanej grupie, ale nie wskazuje na stopień opanowania materiału przez jednostkę [5].

Konstrukcja testu dydaktycznego przebiega w kilku etapach, które przedstawiono na poniższym schemacie [1,19] (Rys. 6).



Rys. 6. Etapy konstrukcji testu dydaktycznego

Analiza narzędzi pomiaru dydaktycznego

Analiza narzędzia pomiaru dydaktycznego [3,5] (Rys. 7), choć rzadko stosowana w kręgach uniwersyteckich, powinna jednak wejść do kanonu działań

każdego nauczyciela akademickiego, ponieważ będąc informacją zwrotną o stopniu przyswojenia materiału przez studentów, jest dodatkowo zweryfikowana pod kątem trafności doboru zadań sprawdzających go (por. wskaźnik rzetelności testu). W rezultacie wykładowca otrzymuje wiadomość o efektywności nauczania i trafności wyboru zastosowanych metod przekazywania wiedzy, a więc o jakości kształcenia. Otrzymanie takiej informacji nie stanowi obecnie najmniejszego problemu (przy wykorzystaniu prostych programów komputerowych typu arkusz kalkulacyjny). Po utworzeniu szablonu z odpowiednimi formułami, wystarczy jedynie nanosić dane wynikowe grup studenckich. Analiza może dotyczyć pojedynczego zadania, jak i całego sprawdzianu/testu.



Rys. 7. Analiza narzędzia pomiaru dydaktycznego

Analiza pojedynczego zadania

Wskaźnik łatwości (p)/trudności zadania (q) informuje, czy zadanie było łatwe czy też nie. Tylko wartości w przedziale 0,2–0,8 dają gwarancję, że zadanie zostało poprawnie zbudowane:

$$p = \frac{f}{N} \quad q = 1 - p,$$

gdzie: f – liczba prawidłowych odpowiedzi na dane zadanie, a N – ogólna liczba piszących test. Wartości wskaźników rozkładają się w przedziale 0–1, zaś interpretacja poszczególnych wyników została zestawiona w Tabeli 1.

Tab. 1. Interpretacja wyników wskaźnika łatwości /trudności zadania

Przedział wartości wskaźnika p	Trudność/latwość zadania	
0,00 – 0,20	bardzo trudne	
0,21 – 0,40	trudne	wartości wskaźnika charakteryzujące zadania wchodzące w skład dobrego testu, sprawdzianu
0,41 – 0,60	średniej trudności	
0,61 – 0,80	łatwe	
0,81 – 1,00	bardzo łatwe	

Pozostałe wskaźniki dotyczące analizy pojedynczych zadań zostały omówione w dalszej części tego rozdziału „Dla chcących wiedzieć więcej” [3, 5].

Analiza całego testu

Wskaźnik rzetelności testu (R) oblicza się po to, by móc upewnić się, co do prawdziwości wyników testu przed wyciągnięciem wniosków, postawieniem

ocen. TYLKO wtedy, gdy współczynnik rzetelności testu przyjmuje wysokie wartości (zakres 0,8–1,0), wyniki uprawniają do postawienia studentom ocen. Jeżeli wartość współczynnika jest poniżej granicy podanego przedziału, należy wstrzymać się z wystawieniem not. Wskaźnik oblicza się ze wzoru:

$$R = \frac{ms^2 - M(m - M)}{(m - 1)s^2},$$

gdzie: s – odchylenie standardowe, m – liczba zadań w teście, M – średni wynik testu, s² – błąd standardowy (wariancja).

Dokładna analiza uzyskanych wyników wskaźnika rzetelności testu zawarta jest w Tabeli 2.

Tab. 2. Interpretacja wyników wskaźnika rzetelności testu

Przedział wartości wskaźnika R	Rzetelność testu	Dalsze postępowanie
0,00 – 0,49	nierzetelny	wyniki nie nadają się do dalszych analiz
0,50 – 0,79	mało rzetelny	wolno analizować wyniki grup, lecz nie powinno się oceniać poszczególnych studentów
0,80 – 0,89	rzetelny	wolno analizować wyniki indywidualne (także różnice pomiędzy studentami) oraz oceniać
0,90 – 1,00	bardzo rzetelny	wolno analizować różnice osiągnięć poszczególnych studentów w różnych zakresach

Warto pamiętać, że wzrastająca liczba zadań w teście wpływa na zwiększenie jego rzetelności.

Pozostałe wskaźniki dotyczące analizy całego testu zostały omówione w dalszej części tego rozdziału „Dla chcących wiedzieć więcej”.

DLA TYCH, KTÓRZY CHCĄ WIEDZIEĆ WIĘCEJ

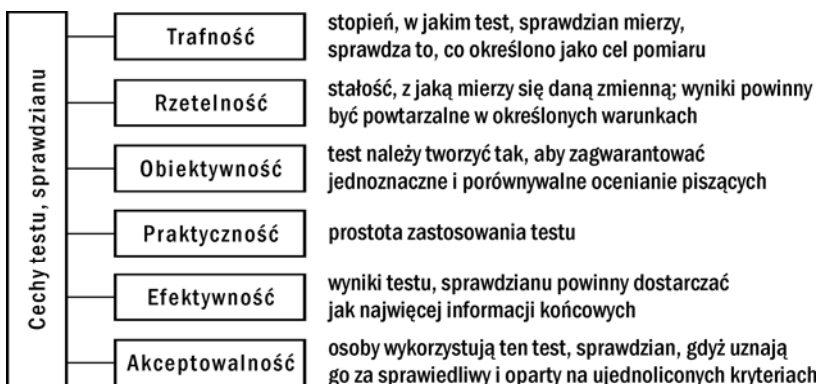
Jak już wspomniano, testy sprawdzające powinny być konstruowane w zależności od ich przeznaczenia (Rys. 8).



Rys. 8. Klasyfikacja testów

Test mocy powinien składać się z zadań o różnej trudności oraz zróżnicowanej treści. Nie jest konieczne szczególne ograniczenie czasowe, gdyż z założenia każdy student może ten test rozwiązać poprawnie. Skracanie czasu piszącym działa niekorzystnie, prowadząc do nerwowości podczas pisania oraz pływających, nieprzemyślanych odpowiedzi. Ograniczenie czasowe (czasem kilkuminutowe) obowiązuje podczas testów szybkości, które powinny składać się z wielu, prostych, jednorodnych zadań. Z założenia żaden z rozwiązujących nie powinien rozwiązać wszystkich zadań.

Należy pamiętać, że uzyskane wyniki pomiaru dydaktycznego mają wartość jedynie wtedy, jeżeli zastosowane do pomiaru narzędzie (sprawdzian, test, kolokwium) jest: rzetelne, trafne, obiektywne, praktyczne, efektywne i akceptowalne. Cechy dobrego sprawdzianu oraz ich krótka charakterystyka zostały przedstawione na Rys. 9.



Rys. 9. Cechy dobrego sprawdzianu

Analiza pojedynczego zadania

Procent prawidłowych rozwiązań (p) informuje o procentowym udziale dobrych odpowiedzi na dane zadanie:

$$p = \frac{f}{N} \cdot 100\%,$$

gdzie: f – liczba prawidłowych odpowiedzi na dane zadanie, N – ogólna liczba piszących test. Jest praktycznie tym samym, co wskaźnik łatwości testu.

Frakcja opuszczenia zadania (f) informuje o tym, ilu studentów opuściło zadanie podczas rozwiązywania testu:

$$f = \frac{n}{N},$$

gdzie n – liczba studentów, którzy opuścili dane zadanie, N – ogólna liczba piszących test.

Moc różnicująca (trafność pytania) (D) – pomaga w różnicowaniu studentów. Aby ją obliczyć, należy uporządkować wyniki studentów (w odniesieniu do danego zadania) wg malejącej liczby punktów i podzielić tę listę na dwie części: górną (dobrzy studenci) i dolną (słabi studenci):

$$D = \frac{2}{N} \cdot (L - S),$$

gdzie: N – parzysta liczba badanych studentów, L – liczba poprawnych odpowiedzi w górnej połowie wyników, S – liczba poprawnych odpowiedzi w dolnej połowie wyników.

Wyniki, które otrzymuje się po podstawieniu danych, mieszczą się w przedziale od -1 do 1. Ich analiza została przedstawiona w Tabeli 3.

Tab. 3. Interpretacja wyników mocy różnicującej zadania.

Przedział	Moc różnicująca zadania	Liczba zadań w teście	Wartość zadowalającej mocy różnicującej
-1,00 – 0,00	nie różnicujące, wadliwie skonstruowane	100	0,22
0,00 – 0,29	słabo różnicujące (rozwiązują je wszyscy studenci)	25	0,42
0,30 – 0,59	średnio różnicujące	10	0,60
0,60 – 1,00	dobrze i bardzo dobrze różnicujące (rozwiązują je tylko najlepsi studenci)	5	0,75

Moc różnicująca zadania jest przydatnym narzędziem dla nauczyciela akademickiego. Pozwala na wyeliminowanie z testu zadań, na które studenci odpowiadają w podobny sposób i otrzymują bardzo zbliżoną lub identyczną ilość punktów. Test zbudowany z zadań dobrze i średnio różnicujących pozwala na wyłonienie najlepszych studentów. W przypadku testu złożonego z zadań słabo różnicujących wyniki są tak zbliżone, iż niejednokrotnie potrzebna jest dodatkowa selekcja (np. w przypadku egzaminu, który ma za zadanie wyłonić określoną liczbę osób z całej grupy).

Analiza całego testu

Średni wynik testu (M) jest średnią arytmetyczną wyników wszystkich zdających:

$$M = \frac{\sum n}{N},$$

gdzie: n – wynik sprawdzianu każdego studenta, N – ogólna liczba piszących test.

Odchylenie standardowe średniej testu (S):

$$S = \frac{\sum (n - M)^2}{N(N - 1)},$$

gdzie: M – średni wynik testu, n – wynik sprawdzianu każdego studenta, N – ogólna liczba piszących test. Wielkość ta informuje, jak szeroko wyniki studentów są rozrzucone od średniego wyniku testu. Im mniejsza jest otrzymana wartość, tym bliżej wyniki oscylują dookoła średniej.

Wskaźnik trudności testu (ff):

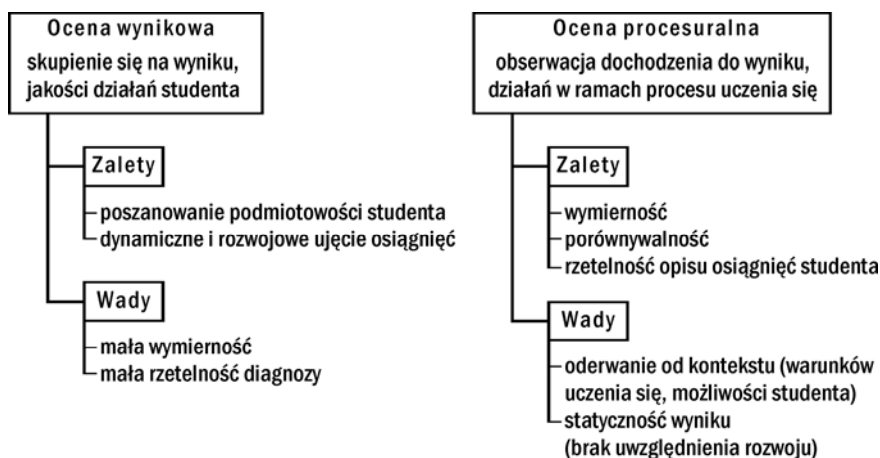
$$ff = \frac{M}{N},$$

gdzie: M – średni wynik testu, m - liczba zadań w teście.

W literaturze można znaleźć informacje, iż ocena może mieć postać stopnia, opisu lub komentarza ustnego. Niezależnie od tej formy, ocena spełnia różnorodne role i funkcje [3,8]:

- dydaktyczną – pomiar wiedzy i umiejętności;
- wychowawczą – kształtowanie postaw;
- społeczną – wskaźnik miejsca w grupie społecznej;
- informacyjno-kontrolną – pomiar stanu wiedzy i stopnia jej opanowania przez studenta, wskaźnik postępu w rozwoju umiejętności oraz planowaniu dalszego toku uczenia się;
- motywacyjną – motywuje, mobilizuje do dalszej pracy.

Inny podział ocen zestawiono w diagramie poniżej (Rys. 10).



Rys. 10. Nastawienia do działań studentów

LITERATURA CYTOWANA

1. B. Niemierko, *Pomiar wyników kształcenia*, WSiP, Warszawa 2000.
2. U. Zoller, *Innovative stes teaching towards scientific and technological literacy for all in the new millennium*, [w:] 3rd IOSTE Symposium For Central and East European Countries, Praga 15-18.06.2000, str. 14.
3. A. Burewicz, H. Gulińska, *Dydaktyka chemii*, Wyd. Naukowe UAM, Poznań 2002.
4. P. Chin, T. Overton, *Development of an electronic peer learning and assessment model*, [w:] *European Variety In Chemical Education*, Faculty of Chemistry, Jagiellonian University, Kraków 4-7.07.2005, Book of abstracts, str.38.
5. J. Mirecka, *Opracowanie sprawdzianów testowych*, [w:] materiały szkoleniowe „Ars docendi”, Uniwersytet Jagielloński, Kraków 2006.
6. U. Zoller, *Are lecture and learning compatible? Maybe for LOCS, unlikely for HOCS*, „Journal of Chemical Education”, 70(3) (1993) 195.
7. U. Zoller, *Assessment of beyond (just) „knowledge” – a doable practice in tertiary chemistry education*, [w:] *Proceedings of the 2nd European Variety in Chemistry Education*, Praga 2007, str. 14.

8. J.S. Francisco, M.B. Nakhleh, S.C. Nurrenbern, M.L. Miller, *Assessing Student Understanding of General Chemistry with Concept Mapping*, „Journal of Chemical Education”, 79(2) (2002) 248.
9. K. Stróżyński, *Ocenianie szkolne dzisiaj: poradnik dla nauczycieli*, Wydawnictwo szkolne PWN, Warszawa 2003.
10. R. Cole, *Computer aided assessment – how we use it*, [w:] *European Variety*, Book of abstracts, str. 42.
11. C.M. O’Connor, C.M. Mc Donnell, *It-supprted learning and assessment for first year undergraduate chemistry students*, [w:] *European Variety...*, Book of abstracts, str. 83.
12. E. Mc Crudden, O.E. Finlayson, *The use of PADs for continuous assessment*, [w:] *Proceedings of the 2nd European Variety...*, Praga 2007, str. 117.
13. P. Bernard, P. Broś, A. Migdał-Mikuli, *Electronic method of students assessment – the application of asynchronous PRS system*, [w:] *Variety In Chemistry Education 2008*, Dublin City University 28 -29.08.2008, Proceedings, str.10.
14. P. Bieniek, H. Gulińska, *Monitoring the user’s work – a New research tool In chemistry education*, [w:] *European Variety...*, Faculty of Chemistry, Jagiellonian University, Kraków 4-7.07.2005, Book of abstracts, str. 116.
15. S. Witkowski, J.P. Witkowski, M. Ruzsak, J. Pielaszek, *Projekt „Aria” powstanie i rozwój*, [w:] *Wykorzystanie technologii informatycznych w akademickiej dydaktyce chemii*, Wydział Chemii UJ, Kraków 2007, Materiały konferencyjne, str. 30.
16. P. Mimero, A. Smith, D. Cardin, J.A. Renuncio, K. Wähälä, M. Karayannis, T. Hase, *„EChemTest” a european evaluation tool to certify at university level both the academic and professional knowledge in chemistry*, [w:] *European Variety*, Book of abstracts, str. 113 <http://ectn-assoc.cpe.fr/echemtest/index.htm>.
17. Kalpachka, H.M., Boyanova Iordanova, L., Radoev, B., *Application of the „taxonomy method” for constructing tests on the course „Surface and colloid science” in the Sofia University*, [w:] *European Variety...*, Book of abstracts, str. 155.
18. M. Groenwald, *Odpowiedzialność nauczyciela jako konstruktora i użytkownika zadań testowych*, [w:] *Diagnostyka edukacyjna. Teoria i praktyka*, Polskie Towarzystwo Diagnostyki Edukacyjnej, Kraków 2004.
19. *Zadania testowe z chemii jako narzędzia sprawdzania i doskonalenia wiedzy uczących się*, red. A. Szternberg, R. Gmoch, Wyd. Skryptowe UO, Opole 1999.

LITERATURA UZUPEŁNIAJĄCA

- R.I. Arends, *Uczymy się nauczać*, WSiP, Warszawa 1998.
- K. Ciżkowicz, J. Ochenduszko, *Pomiar sprawdzający wielostopniowy*, WSiP, Warszawa 1998.
- J. Homplewicz, *Etyka pedagogiczna*, Wydawnictwo WSP, Rzeszów 1996.
- B. Niemierko, *Pomiar sprawdzający w dydaktyce*, WSiP, Warszawa 1990.
- T. Szaran, *Pomiar dydaktyczny*, WSiP, Warszawa 2000.
- D. Sołtys, M. Szmigiel, *Doskonalenie kompetencji nauczania w zakresie diagnozy edukacyjnej*, ZamKor, Kraków 2002.
- K. Stróżyński, M. Giermakowski, *Jak oceniać?*, Wyd. Nauczycielskie, Jelenia Góra 1999.
- T. Tyszka, *Psychologiczne pułapki oceniania i podejmowanie decyzji*, Gdańskie Wydawnictwo Psychologiczne, Gdańsk 2000.