

**Grażyna Szyling**

Uniwersytet Gdański

## **„Być może będzie w przyszłości ozdobą polskiej piłki nożnej?”. Wybrane aspekty trafności oceny końcowej osiągnięć ucznia**

Zacznijmy od sytuacji szkolnej. Zbliża się koniec roku szkolnego, a wraz z nim wystawianie uczniom III klas gimnazjum ocen końcowych, ocen dużo bardziej znaczących dla ucznia niż dotychczas przez niego otrzymywane w trakcie procesu kształcenia. W założeniu mają one być zgodne z obowiązującymi w szkole systemami oceniania: wewnątrzszkolnym i przedmiotowymi, co powinno wydatnie ograniczyć dylematy nauczycieli i pomóc im w jak najbardziej obiektywnym określeniu, na jakim poziomie uczeń spełnił wymagania programowe.

Trudno jednak przewidzieć, jak postąpi nauczyciel, gdy dobrze mu znany, acz szczególnie przypadkiem ucznia nie da się wtłoczyć w ramy przyjętych uprzednio kryteriów, a kurczowe trzymanie się regulaminów wydaje się być w jakimś sensie krzywdzące dla ucznia. Czy podniesie przedmiotową ocenę jego osiągnięć, wierząc, że: „Być może będzie on w przyszłości ozdobą polskiej piłki nożnej?”<sup>1</sup>. Czy może nie udzieli uczniowi promocji, bo na to wskazują i oceny cząstkowe, i jego postawa wobec przedmiotu?

Jeśli przyjmiemy, że ocena szkolna zawiera nie tylko porównanie wyniku sprawdzania z wymaganiami programowymi, ale też osobiste ustosunkowanie się nauczyciela do tego wyniku (Niemierko, 2002: 184), to podejmowane w tej materii decyzje nauczyciela rozpatrywać należy w kategoriach kompromisu między zalecanymi procedurami oceniania a znaczeniem i wartościami, jakie przypisuje on ocenie osiągnięć

---

<sup>1</sup> Ten cytat, jak i pozostałe przytoczone w tej pracy wypowiedzi nauczycieli, pochodzi z badań przeprowadzonych przez autorkę (czerwiec/lipiec 2003).

ucznia. Wielkość tego kompromisu, nierzadko wymuszana warunkami pracy szkoły, ujawniać będą w coraz większym stopniu wyniki egzaminów zewnętrznych.

Przedstawiony problem wiąże się z szeroko rozumianym pojęciem **trafności oceny szkolnej**, którego wybrane aspekty zostaną rozpatrzone w niniejszym studium. Podstawę empiryczną rozważań stanowić będą wyniki badań przeprowadzonych przez autorkę w czerwcu i lipcu 2003 r.

## **Właściwości końcowej oceny osiągnięć ucznia na danym etapie edukacyjnym**

Przedmiotem zainteresowania uczyniono w tej pracy ocenę z danego przedmiotu uzyskaną przez ucznia na świadectwie ukończenia gimnazjum. Będąc swoistą, z konieczności uproszczoną do stopnia szkolnego, syntezą wyników wielu egzaminów powszednich o zróżnicowanej rzetelności i trafności, z samej swej natury jest wieloznacznym środkiem komunikacji. Dla wielu jej odbiorców nie jest jasne, co było oceniane, na jakiej zasadzie łączono wyniki i jakie normy zastosowano (Knight, 2000: 46–47).

Ocena ta jednak już przez sam fakt umieszczenia jej na świadectwie ukończenia gimnazjum staje się oceną doniosłą i jako taka jest poddawana wielorakim interpretacjom i wykorzystywana do różnych celów.

Po pierwsze – ocena ta zawiera istotną, nie tylko dla ucznia, wartościującą informację o osiągnięciu przez niego pewnego celu edukacyjnego i o aktualnym poziomie jego umiejętności. W tym sensie ma ona charakter sumujący, jest klamrą zamykającą pewien etap kształcenia z danego przedmiotu.

Po drugie – co nie znaczy, że mniej ważne dla uczniów i nauczycieli – w świetle obowiązującego prawa oświatowego ocena na świadectwie ukończenia gimnazjum i wyniki egzaminu gimnazjalnego zostają zastosowane do podejmowania decyzji o przyjęciu ucznia do szkoły wyższego szczebla (Rozporządzenie MEN z dnia 27 sierpnia 2001 r.), czyli wykorzystane do orientacji szkolnej i – w jakimś sensie – selekcji.

Po trzecie – zmiany w systemie szkolnictwa sprawiły, że oceny końcowe z danych przedmiotów są wykorzystywane jako jedno z ważnych kryteriów zewnętrznego aspektu trafności egzaminów doniosłych (Walczak, 2002; Dubiecka, 2002).

Po czwarte – szkoły wyższego szczebla mogą traktować ocenę końcową jako element diagnozy na wejściu i w tym kontekście nabiera ona znaczenia dydaktycznego – informuje o osiągnięciu przez ucznia dyspozycji do wykonywania zadań o określonym poziomie trudności.

Już ten, niepełny przecież, przegląd zastosowań oceny końcowej wskazuje na wielorakie konsekwencje, które pociągają za sobą jej wystawienie na świadectwie gimnazjalnym. Na ile są one przewidywane przez nauczycieli i na ile te przewidywania oparte są na interpretacji oceny potwierdzonej empirycznymi dowodami?

Od odpowiedzi na te pytania zależy,

[...] w jakim stopniu – jeśli w ogóle – na podstawie dowodów i rozumowych uzasadnień wyniki testu [oceny szkolne] powinny być interpretowane i wykorzystywane w zaproponowany sposób (Messick, 1989a: 9).

Jest to zagadnienie kluczowe dla **teorii trafności** sformułowanej przez Samuela Messicka (1989b), którego poglądy połączone z teorią nauczycielskiego oceniania (Brookhart, 1993) zostały zastosowane w tym studium do rozważań nad trafnością oceny szkolnej.

## Teoria trafności S. Messicka zastosowana do oceniania

Analizie i ocenie pod względem trafności podlega nie sama ocena szkolna, ale jej interpretacja i założone przez nauczyciela konsekwencje jej zastosowania. Ustalanie trafności (walidacja) jest procesem polegającym na gromadzeniu i porządkowaniu danych empirycznych oraz argumentów (racjonalnych uzasadnień), traktowanych jako równoprawne.

Teorię trafności Messicka (1989b) zastosowaną do oceniania (Brookhart, 1993) można przedstawić w postaci macierzy postępującej aspektów trafności.

Tabela 1.

Macierz postępująca aspektów trafności oceniania szkolnego

Uzasadnienie trafności	Funkcje oceny trafności	
	Interpretacja	Zastosowanie
Na podstawie dowodów	Trafność teoretyczna <i>Interpretacja potwierdzona empirycznymi dowodami</i> <i>Znaczenie pomyślane</i>	Trafność teoretyczna + stosowność i użyteczność <i>Wykorzystanie wyników do jakiegoś celu</i> <i>Zastosowanie oceny do opisanego ucznia</i>
Na podstawie konsekwencji	Trafność teoretyczna + stosowność i użyteczność + kształtowanie wartości <i>Oszacowanie wartości wywołanych interpretacją wyniku przez ucznia</i> <i>Uwzględnianie przez nauczyciela konsekwencji interpretowania oceny przez ucznia</i>	Trafność teoretyczna + stosowność i użyteczność + kształtowanie wartości + skutki społeczne <i>Oszacowanie konsekwencji społecznych (w szkole i poza nią)</i> <i>Uwzględnianie przez nauczyciela konsekwencji zastosowania oceny</i>

Cztery pola macierzy (tab. 1) odpowiadają czterem podstawowym zagadnieniom trafności. W takim ujęciu trafność jest jednolitym pojęciem, którego podstawę stanowi trafność teoretyczna potwierdzona empirycznymi dowodami. Kolejne kategorie macierzy rozwijają ten fundamentalny aspekt o jego zewnętrzne zastosowania służące różnym celom.

Stosując kategorie trafności Messicka do oceniania, można rozważyć zakres, w jakim oszacowanie wartości oceny i konsekwencji jej zastosowania jest częścią nauczycielskich refleksji nad własnym ocenianiem.

Istotne, zwłaszcza przy szacowaniu trafności oceny końcowej z przedmiotu, staje się pytanie, które aspekty trafności silniej oddziałują na decyzje nauczycieli w tej materii. Dominacja trafności konsekwencyjnej nad teoretyczną może bowiem sprawiać, że oceny końcowe uzależnione będą przede wszystkim od ingerencji kontekstu kształcenia i przyjętych w środowisku wartości społeczno-moralnych, co nierzadko wymusza stosowanie przez nauczycieli **drugiego układu wymagań programowych** (Niemierko, 2002: 207–222; Szyling, 2002). Z kolei ta „szara strefa oceniania”, ze względu na swój nie w pełni oficjalny charakter i trudne do rozpoznania zależności nią rządzące, wydatnie ogranicza możliwość interpretowania ocen w kontekście innym niż indywidualny bądź lokalny.

## Wyniki badań

Wykorzystane w pracy wyniki są zaczerpnięte z badań pilotażowych (Konarzewski, 2000: 46–47) prowadzonych przez autorkę tego tekstu w czerwcu i lipcu 2003 r. Na potrzeby niniejszego studium wybrano z nich do analizy tę część danych, które związane są z sytuacją wystawiania przez nauczycieli końcowej oceny umiejętności ucznia z danego przedmiotu.

### Dobór próbek

Badaniu poddano przypadkowo dobraną próbkę nauczycieli szkół ponadpodstawowych z różnych rejonów Polski: byli to nauczyciele z zaprzyjaźnionych szkół, słuchacze studiów podyplomowych dla nauczycieli, uczestnicy kursów metodycznych. Żadna z tych form doskonalenia nie była bezpośrednio związana z ocenianiem osiągnięć uczniów, choć – jak wynika z badań (Brookhart, 1993: 123) – wykształcenie pomiarowe zasadniczo nie zmienia nauczycielskich praktyk oceniania, a ogranicza jedynie skłonność nauczycieli do odwoływania się przy ocenianiu do wcześniejszych osiągnięć ucznia. Charakterystykę badanej próbki przedstawia tabela 2.

Próbka ta ze względu na swe wewnętrzne zróżnicowanie wydaje się uprawniać do ostrożnego wnioskowania na temat zakresu, w jakim sądy wartościujące i wiedza o wielu zastosowaniach ocen jest częścią nauczycielskich praktyk oceniania, zwłaszcza, że nie będą badane korelacje między opiniami nauczycieli a ich stażem pracy czy uczonym przez nich przedmiotem. Należy jednak pamiętać, że

[...] uogólnianie wyników z takiej próbki jest zawsze obarczone błędem, którego nie sposób oszacować (Konarzewski, 2000: 107).

Tabela 2.  
Charakterystyka próbki nauczycieli objętych badaniem

Staż pracy (% respondentów)	Szkoła*			Nauczany przedmiot*					
	gimnazjum	średnia ogólnokształcąca	średnia techniczna	język polski	historia	języki obce	matematyka, fizyka	geografia, biologia, chemia	zawodowe, inne
0–5 (23,2)	8	7	4	3	5	3	1	2	3
6–10 (18,3)	4	6	5	7		3	3	2	
11–20 (34,1)	11	9	8	6	8	2	3	5	4
21– (24,4)	3	12	5	1	6	1	5	3	4
SUMA	26	34	22	17	19	9	12	12	11
%	31,7	41,5	26,8	21,3	23,7	11,3	15,0	15,0	13,7

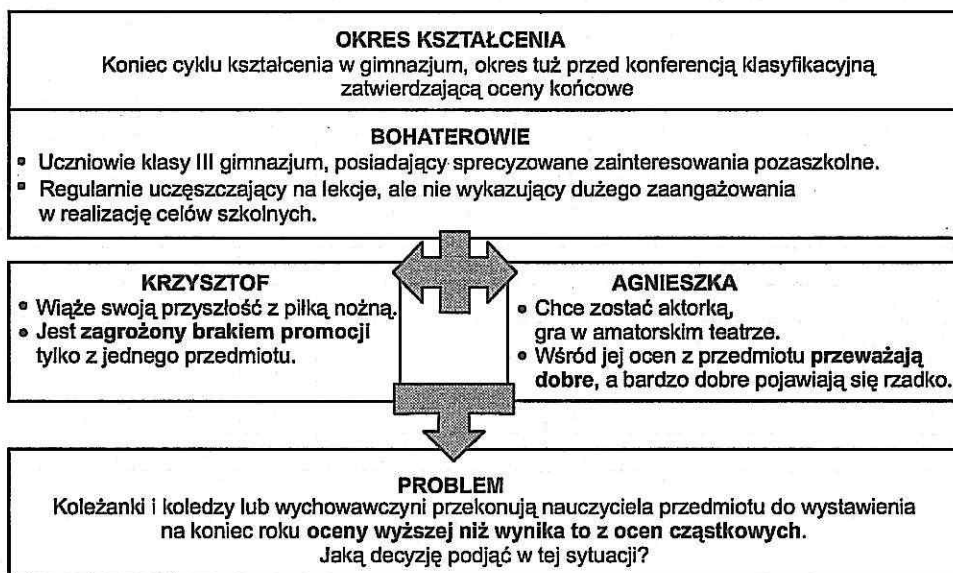
\* różnice w liczbach podanych w tych dwóch częściach tabeli wynikają z faktu, że nie wszyscy respondenci podali w metryczkach pełne dane

## Narzędzia

W badaniu wykorzystano, wzorowaną na badaniach S. Brookhart (1993), otwartą odmianę metody epizodów (Konarzewski, 2000: 154–155), poszerzoną o możliwość wybrania przez respondenta w pytaniach zamkniętych opcji: „Podjąć inną decyzję. Jaką?”. Na potrzeby niniejszego studium wybrano dwa z siedmiu zastosowanych w badaniu scenariuszy o ocenianiu, które zbudowano według podobnego schematu (schemat 1). Uwzględnił on wszystkie grupy czynników sprzyjających występowaniu w szkole drugiego układu wymagań programowych: czynniki niezależne (koniec cyklu kształcenia, zdolności) i zależne od postawy ucznia (prośba osób trzecich, wkład pracy i zachowanie na lekcji) (Szyling, 2002: 81). Bohaterami omawianych scenariuszy uczyniono jednak uczniów różniących się od siebie osiągnięciami szkolnymi. Ponadto, by analogie między obydwoma przypadkami nie narzucały się respondentom w sposób oczywisty, rozdzielono je dwoma innymi scenariuszami przedstawiającymi odmienne sytuacje oceniania.

Pod każdym ze scenariuszy znajdowało się zadanie wielokrotnego wyboru zawierające wersje możliwych zachowań nauczyciela w danej sytuacji: zawyżenie oceny w stosunku do ocen częściowych lub wystawienie oceny zgodnej z ocenami częściowymi. Pytanie otwarte: „Dlaczego dokonałeś/eś takiego wyboru?” zapraszało respondentów do zaprezentowania osobistych racji i motywów, którymi się kierowali, podejmując decyzję.

**Schemat 1.**  
**Schemat scenariuszy dotyczących wystawiania oceny końcowej**



### Analiza wyników (zadania zamknięte)

W tabeli 3 zaprezentowano wybory dokonane przez nauczycieli w scenariuszach wykorzystanych w niniejszej pracy. W obydwu wskazać można, wyraźnie dominującą nad pozostałymi, najczęstszą odpowiedź, która jest jednak odmienna dla każdego z epizodów.

**Tabela 3.**  
**Decyzje dotyczące klasyfikacyjnych ocen dwojga uczniów podjęte przez nauczycieli**

Opis przypadku:	Decyzja nauczycieli o stopniu końcowym (w %)		
	A. Podnieść ocenę	B. Wystawić ocenę na podstawie ocen częściowych	C. Podjąć inną decyzję*
I. Krzysztof: uczeń zagrożony brakiem promocji tylko z jednego przedmiotu, piłkarz	60,9	23,0	16,1
II. Agnieszka: uczennica, wśród której ocen z przedmiotu przeważają <i>dobre</i> , występująca w amatorskim teatrze	16,1	73,6	10,3

Aż 60,9% respondentów jest skłonnych wystawić ze swojego przedmiotu na koniec III klasy gimnazjum dopuszczający uczniowi, wśród ocen którego przeważają niedostateczne. Natomiast za podniesieniem do bardzo dobrej oceny uczennicy, która w ciągu roku szkolnego otrzymywała głównie dobre, opowiada się zaledwie 16,1% respondentów. W tym drugim z przypadków odpowiedzią wybieraną najczęściej, bo aż przez 73,6% nauczycieli, jest: „Wystawić Agnieszce „dobry” na koniec roku, uznając, że otrzymane przez nią oceny jednoznacznie określają poziom jej umiejętności”. Za przyjęciem analogicznego rozwiązania w przypadku ucznia zagrożonego brakiem promocji opowiada się 23% badanych.

Analiza propozycji podanych przez nauczycieli przy wyborze opcji „Podjąć inną decyzję” wskazuje, że nie zmieniają one zasadniczo relacji frekwencyjnych między wersjami odpowiedzi A i B w obydwu scenariuszach. Respondenci albo wychodzą z założenia, że sytuacja prezentowana w scenariuszu nie mogłaby się przydarzyć w ich praktyce szkolnej, albo proponują „przepytanie uczniów”. To drugie z rozwiązań jest *de facto* unikaniem przez tę grupę nauczycieli ( $\approx 7\%$  respondentów) dokonania wyboru, ale wskazuje na ważną dla nich próbę silniejszego powiązania stopnia końcowego z aktualnymi osiągnięciami ucznia z danego przedmiotu.

Podsumowując: ze względu na znaczną przewagę frekwencyjną dwóch odpowiedzi wybranych przez respondentów w omawianych scenariuszach możemy pokusić się o ostrożne wnioskowanie na tej podstawie.

1. W sytuacji, gdy stopień ma decydować o uzyskaniu promocji przez ucznia, który ma sprecyzowane zainteresowania pozaszkolne, ale z reguły jest nieprzygotowany do lekcji, znaczna część nauczycieli jest skłonna zawyżyć ocenę końcową, obniżając okazjonalnie podstawowe wymagania programowe, czyli stosując drugi układ wymagań (por. Niemierko, 2002: 215–217). Niewątpliwie wpływa to niekorzystnie na zawartość treściową oceny czyli także jej „powiązanie z określoną treścią kształcenia, że o uczniu uzyskującym dany stopień można wnioskować, jakie elementy tej treści opanował, a jakich nie opanował” (tamże: 223).
2. Znacznie mniejszą gotowość nauczycieli do odstępstwa od określonych wcześniej wymagań obserwuje się przy ocenianiu osiągnięć ucznia wyższych niż podstawowe, co może wskazywać na silniejszy związek tych ocen z treścią kształcenia, ale nie wyklucza wpływu czynników zewnętrznych na ocenę.

Rozważając powyższe wnioski w kontekście pojęcia trafności, należy zauważyć, że stopień powiązania oceny końcowej z treścią kształcenia jest istotnym kryterium szacowania trafności teoretycznej. Dowodem empirycznym interpretacji oceny staje się w tym przypadku jej odniesienie do celów kształcenia, materiału i wymagań programowych.



### Analiza wyników (pytania otwarte)

Za wyborami, których dokonali respondenci w zadaniach zamkniętych umieszczonych pod scenariuszami, kryły się różne ich poglądy i argumenty wyrażone w swobodnych odpowiedziach na pytanie otwarte. Do analizy tych nauczycielskich refleksji o ocenianiu wykorzystano macierz postępującą aspektów trafności S. Messicka (tab. 1).

Wyniki analizy jakościowej tych pisemnych odpowiedzi nauczycieli, które były zaklasyfikowane jako reprezentatywne dla kategorii trafności teoretycznej, przedstawia tabela 4. Uwzględnione zostały w niej komentarze wskazujące wprost, co nauczyciele myślą o istocie oceny oraz jakie dowody empiryczne za tym przemawiają.

Tabela 4.  
Kategorie trafności teoretycznej

Trafność teoretyczna	Wybrane przykłady komentarzy wskazujące, co nauczyciele rozumieją pod znaczeniem oceny opartym na dowodach empirycznych (liczba/% odpowiedzi dotyczących trafności teoretycznej oceny)	Liczebność	
		Scenariusz I	Scenariusz II
A. Ocena jako wynagrodzenie za wykonaną pracę (39/76,5%)	1. Poziom wykonanej pracy (25/49%) • Ocena ma być adekwatna do poziomu wiedzy. • Uczennica wypracowała taką ocenę.	2	23
	2. Spełnienie kryteriów/wymagań (14/27,5%) • Istotnym elementem oceny wystawianej na koniec roku jest frekwencja ucznia na zajęciach (uczęszczanie na zajęcia) • Ocenę stawiam zgodnie z obowiązującymi kryteriami, także tymi zawartymi w WSO.	5	9
B. Ocena jako obliczony wynik punktowy (10/19,6%)	1. Odwołanie do bieżących ocen • Jest przewaga ocen dobrych, więc sytuacja jest klarowna. • Taka ocena wychodzi z ocen częściowych.	0	10
C. Ocena jako odzwierciedlenie zdolności (2/3,9%)	1. Osiągnięcia/zdolności • Nie każdy jest uzdolniony językowo.	2	0
Suma uzasadnień uwzględnionych w komentarzach respondentów:		9	43
% sumy wszystkich uzasadnień uwzględnionych w komentarzach respondentów:		27,1%	

Dla wielu nauczycieli ocena jest rodzajem wynagrodzenia za wykonaną pracę, natomiast mniej liczne wypowiedzi wskazują na to, że ocena końcowa jest rodzajem obliczonego wyniku.

W tabeli 4 na szczególną uwagę zasługuje wyraźna dysproporcja w liczbie komentarzy dotyczących analizowanych scenariuszy. Aż 82,7% uzasadnień trafności



oceny, odwołujących się do dowodów empirycznych, dotyczy Scenariusza II, w którym respondenci w swej zdecydowanej większości odpowiedzieli, że nie podniosą uczennicy oceny końcowej do bardzo dobrego. Przytaczane przez nich argumenty głównie uzasadniają pewną rygorystyczność sumującej oceny poziomu umiejętności uczennicy.

Natomiast odpowiedź najczęściej udzielana w Scenariuszu I stoi w zdecydowanej sprzeczności z rozumieniem oceny jako odzwierciedlenia poziomu wykonanej pracy i obliczonego wyniku punktowego. Jednak zwolennicy wystawienia dopuszczającego uczniowi, wśród ocen którego przeważają niedostateczne, także formułują wypowiedzi związane z trafnością teoretyczną, a zwłaszcza z rozumieniem oceny jako spełnienia ustalonych kryteriów. Są to jednak kryteria, którym znacznie bliżej do drugiego układu wymagań programowych („Dopuszczający stawiam za uczęszczanie na lekcje”) niż do oficjalnych wymagań określających konieczne umiejętności ucznia.

Tabela 5.  
Kategorie stosowności i użyteczności

Stosowność i użyteczność	Wybrane przykłady komentarzy wskazujące, co nauczyciele myślą o znaczeniu oceny wystawionej uczniowi i gdzie szukają dowodów jej stosowności i użyteczności (liczba/% odpowiedzi dotyczących stosowności i użyteczności oceny)	Liczebność	
		Scenariusz I	Scenariusz II
A. Dowody uczniowskich osiągnięć (36/90%)	1. Dowody włożonego wysiłku (20/50%)	12	8
	<ul style="list-style-type: none"> <li>• <i>Jeśli uczeń aktywnie uczestniczyłby w lekcji, byłabym mu skłonna podnieść ocenę.</i></li> <li>• <i>Nudzenie się na lekcji świadczy, że uczeń nie chce się nauczyć czegokolwiek.</i></li> </ul>		
	2. Inne obszary aktywności ucznia (10/25%)	5	5
	<ul style="list-style-type: none"> <li>• <i>Uczeń ma talent w innej dziedzinie. Biorę to pod uwagę.</i></li> <li>• <i>Uwzględnię te zajęcia, bo uważam je za ważne/rozwijają ucznia.</i></li> </ul>		
	3. Zdolność do wykonania zadania (5/12,5%)	5	0
	<ul style="list-style-type: none"> <li>• <i>Skoro uczeń radzi sobie z innymi przedmiotami, powinien także nadrobić zaległości z mojego.</i></li> </ul>		
B. Dowody zewnętrzne wobec uczniów (4/10%)	1. Przypominanie/zachęcanie przez nauczyciela	4	0
	<ul style="list-style-type: none"> <li>• <i>Zakładam, że zostały wyczerpane przez nauczyciela wszystkie możliwości, by do takiej sytuacji nie doszło.</i></li> </ul>		
Suma uzasadnień uwzględnionych w komentarzach respondentów:		26	14
% sumy wszystkich uzasadnień uwzględnionych w komentarzach respondentów:		21,2%	

Tabela 5. przedstawia wyniki analizy jakościowej wszystkich pisemnych odpowiedzi nauczycieli, które były zaklasyfikowane jako reprezentatywne dla kategorii stosowności i użyteczności oceny. Uwzględnione zostały w niej komentarze wskazujące wprost, co nauczyciele myślą o znaczeniu oceny wystawionej uczniowi oraz gdzie szukają dowodów jej przydatności do opisanego osiągnięcia ucznia. Najchętniej wymieniane są dowody potwierdzające uczniowski wysiłek włożony w osiąganie celów szkolnych oraz – już znacznie rzadziej: inne obszary aktywności ucznia, przestrzeganie jego zdolności do wykonania zadania, a także jego reakcji na zachęty ze strony nauczyciela.

Wśród wypowiedzi respondentów analizowanych w ramach tego aspektu trafności oceny zaobserwować można pewną prawidłowość. Ponad 2/3 przytaczanych dowodów dotyczy kategorii nieosiągniętych przez ucznia, czyli braku jego zaangażowania w lekcje i zainteresowania przedmiotem, niewykorzystanych w pełni zdolności czy zlekceważonych przypomnień nauczyciela.

Takie rozumienie stosowności oceny jest nie tylko konsekwencją sytuacji przedstawionej w scenariuszach, ale także wynika z powiązania większości analizowanych komentarzy z wyborem opcji B. w zadaniach zamkniętych: „Wystawić niedostateczny [Sc. I] / dobry [Sc. II] na koniec roku, uznając, że otrzymane przez ucznia oceny jednoznacznie określają poziom jego umiejętności”.

Pisemne odpowiedzi zanalizowane w tabelach 4 i 5 ilustrowały, jak nauczyciele definiują znaczenie oceny końcowej i jakie dowody jej stosowności do opisanego osiągnięcia ucznia uznają za najważniejsze. Natomiast dwa kolejne aspekty trafności poszerzają te rozważania o szacowanie przez nauczyciela dwojakiego rodzaju oczekiwanych konsekwencji zastosowania oceny: skutków interpretowania oceny przez ucznia oraz idących znacznie dalej skutków społecznych.

Wyniki analizy jakościowej tych komentarzy respondentów, które były zaklasyfikowane jako reprezentatywne dla kategorii „kształtowanie wartości”, przedstawia tabela 6. Uwzględnione zostały w niej wypowiedzi o charakterze wartościującym, dotyczące przewidywanych przez nauczycieli skutków dokonanej przez uczniów interpretacji oceny oraz jej wpływu na ich przekonania i postawy.

W grupie komentarzy zawartych w tabeli 6 niewielką przewagą zdobyły te, które odwołują się do wyrozumiałości nauczyciela oraz jego skłonności do usprawiedliwiania postawy ucznia wobec przedmiotu czy uczenia się. Najbardziej znaczące dla formułowania tych kryteriów są poglądy respondentów, których istotę najpełniej ujmuje konstatacja: „Zawsze należy kierować się dobrem ucznia”.

Przy szacowaniu konsekwencji interpretowania oceny końcowej nauczyciele nieco rzadziej uwzględniają poczucie sprawiedliwości, choć ważne dla dużej ich części jest traktowanie wszystkich uczniów tak samo oraz konsekwentne przestrzeganie ustalonych wcześniej i znanych wymagań, „w których na pewno nie jest umieszczone wstawianie się wychowawczyni za uczniem”.

**Tabela 6.**  
**Kategorie kształtowania wartości**

Kształtowanie wartości	Wybrane przykłady komentarzy wskazujące, jakie konsekwencje interpretowania oceny przez ucznia przewidują nauczyciele, gdy ją wystawiają (liczba/% odpowiedzi dotyczących kształtowania wartości przez ocenę)	Liczebność	
		Scenariusz I	Scenariusz II
A. Odwołanie do poczucia sprawiedliwości (27/45,8%)	1. Bycie sprawiedliwym/obiektywnym (13/22%)	6	7
	▪ <i>Jeśli podniesie się ocenę jednemu uczniowi, to trzeba też innym. Muszę uczniów traktować jednakowo.</i>		
	2. Konsekwencja nauczyciela (7/11,9%)	0	7
	▪ <i>Nie można ulegać presji wychowawcy czy innym naciskom tuż przed konferencją.</i>		
	3. Stworzenie możliwości (5/8,5%)	5	0
	▪ <i>Uczeń znał zakres wymagań na ocenę. Miał też możliwość poprawienia ocen.</i>		
	4. Rozróżnienie typu właściwy/błędny (2/3,4%)	1	1
▪ <i>Uważam za błędne podnoszenie ocen końcowych do bardzo dobrego. Jestem temu przeciwna.</i>			
B. Odwołanie się do wyrozumiałości (32/54,2%)	1. Usprawiedliwianie ucznia (12/20,3%)	11	1
	▪ <i>Jeżeli uczeń jest zdolny w jakimś kierunku, mogę mu darować pewne niedostatkі. Nie można być dobrym ze wszystkiego.</i>		
	▪ <i>Mój przedmiot nie musi być mu w życiu potrzebny.</i>		
	2. Nauczycielskie przekonania/poglądy (9/15,2%)	8	1
	▪ <i>Nigdy nie staję uczniowi na drodze.</i>		
	▪ <i>Szkoła nie powinna decydować o losie ucznia na całe życie.</i>		
	3. Współczucie/wyrozumiałość (6/10,9%)	4	2
▪ <i>Dzięki tej ocenie może ukończyć szkołę. Każdy powinien otrzymać szansę.</i>			
4. Wsparcie wysiłku ucznia (5/8,5%)	2	3	
▪ <i>Należy stworzyć uczniowi warunki do rozwijania zainteresowań pozaszkolnych.</i>			
Suma uzasadnień uwzględnionych w komentarzach respondentów:		37	22
% sumy wszystkich uzasadnień uwzględnionych w komentarzach respondentów:		31,4%	

Interesująca jest także występująca wśród analizowanych wypowiedzi respondentów pewna prawidłowość, którą ilustruje poniższa tabela.

Tabela 6a.  
Nauczycielskie kryteria kształtowania wartości a analizowane scenariusze

Scenariusz	Kategorie kształtowania wartości	
	A. Odwołanie się do poczucia sprawiedliwości	B. Odwołanie się do wyrozumiałości
Scenariusz I (o Krzysztofie)	34,4%	67,6%
Scenariusz II (o Agnieszce)	68,2%	31,8%

Można przeprowadzić prostą analogię między decyzjami nauczycieli dotyczącymi wysokości ocen klasyfikacyjnych dla Krzysztofa i Agnieszki (tab. 3) a szacowaniem skutków interpretowania tych ocen przez uczniów. Wyraźnie widoczne jest przeświadczenie respondentów, że podniesienie oceny końcowej pociągnie za sobą postrzeganie nauczyciela jako osoby wyrozumiałej i wspierającej ucznia, natomiast wystawienie stopnia zgodnego z ocenami bieżącymi implikować będzie poczucie sprawiedliwości i obiektywizmu. Dla wielu nauczycieli te dwie kategorie kształtowania wartości nie są wcale sobie przeciwstawne, zwłaszcza jeśli dotyczą ocen klasyfikacyjnych o różnej wartości.

Ostatni i – co za tym idzie – najszerzej rozumiany aspekt pojęcia trafności oceny obejmuje uwzględniane przez nauczycieli długofalowe konsekwencje zastosowania oceny. Wyniki analizy komentarzy nauczycieli związanych z przewidywanymi skutkami społecznymi przedstawia tabela 7.

Zdecydowana większość komentarzy w tabeli 7 związana jest z różnymi odmianami myślenia o wysiłku ucznia i jego postawie wobec uczenia się. W tym szczególnym przypadku, gdy szacowana jest trafność końcowej oceny z przedmiotu, konsekwencje szkolne siłą rzeczy dotyczą pracy ucznia w innej szkole.

Zaobserwować też można, że niemalże połowa przewidywanych przez nauczycieli skutków zastosowania oceny ma charakter weryfikujący poziom osiągnięć szkolnych ucznia w sposób dla niego niekorzystny: „Będzie zmuszony do wyrównania braków, bo szkoła średnia nie jest obowiązkowa”, „Nie ominą go egzaminy”. Może to mieć związek z faktem, że aż 79,8% sądów zaklasyfikowanych do tego aspektu trafności dotyczy Scenariusza I, w którym większość respondentów opowiedziała się za udzieleniem uczniowi promocji, mimo że oceny wskazywały, iż nie opanował on w sposób zadowalający podstawowych wymagań programowych.

Proporcjonalnie istotna dla przewidywanych skutków zastosowania oceny okazała się być dla nauczycieli atmosfera panująca w gronie pedagogicznym, co wskazuje na to, że respondenci biorą pod uwagę nie tylko konsekwencje, jakie ocena niesie z sobą dla ucznia, ale i dla nich samych.

Tabela 7.  
Kategorie skutków społecznych

Skutki społeczne	Wybrane przykłady komentarzy wskazujące, jakie konsekwencje społeczne zastosowania oceny uwzględniają nauczyciele (liczba/% odpowiedzi dotyczących przewidywanych skutków społecznych oceny)	Liczebność	
		Scenariusz I	Scenariusz II
A. Konsekwencje szkolne (31/81,6%)	1. Praca w innej klasie/szkole (13/34,2%) • Szkoła średnia umożliwi mu wyrównanie braków. • Moja ocena może zadecydować (lub nie) o dostaniu się uczennicy do odpowiedniej szkoły.	9	4
	2. Postawa wobec uczenia się (8/21,0%) • Może skłoni go do przemyśleń i zmiany postawy wobec uczenia się. • Pozostawienie na drugi rok niczego nie zmieni.	7	1
	3. Atmosfera w gronie pedagogicznym (5/13,2%) • Postawiłam dopuszczający dla tak zwanego „świętego spokoju”. • Kiedyś też prosiłam jako wychowawca o podniesienie oceny. To działa na zasadzie: on mi pomógł, więc ja też...	4	1
	4. Weryfikujący charakter egzaminu zewnętrznego (3/7,9%) • Braki ucznia i tak wyjdą na jaw na egzaminie. • Priorytetem jest przygotowanie ucznia do egzaminu.	3	0
	5. Przyszłe oceny (2/5,3%) • Jeśli sam nie uzupełni braków, będzie miał problemy z tym przedmiotem.	1	1
	1. Uczenie odpowiedzialności (4/10,5%) • Uczeń podjął decyzję i sam będzie ponosił konsekwencje wynikające ze swojego wyboru. • Jak ustąpię, to nauczy się, że wszystko można załatwić.	4	0
	2. Rozwijanie pozaszkolnych pasji (2/5,3%) • Może będzie ozdobą polskiej piłki nożnej?	2	0
B. Konsekwencje pozaszkolne (7/18,4%)	3. Poczucie własnej wartości u ucznia (1/2,6%) • Nie należy stwarzać sytuacji sugerującej uczennicy wyższe możliwości niż istniejące.	0	1
	Suma uzasadnień uwzględnionych w komentarzach respondentów:	30	8
% sumy wszystkich uzasadnień uwzględnionych w komentarzach respondentów:		20,3%	

## Interpretacja

Na podstawie analizy zadań otwartych można podjąć próbę odpowiedzi na pytanie, które aspekty trafności teoretycznej silniej oddziałują na decyzje nauczycieli związane z ocenianiem osiągnięć ucznia na koniec etapu kształcenia i jakie to może implikować konsekwencje. Ze względu jednak na przypadkowy dobór próbek i pilotażowy charakter badań wnioski te będą miały charakter wstępnych oszacowań.

Zestawienie procentowe wszystkich komentarzy zaklasyfikowanych jako reprezentatywne dla poszczególnych aspektów trafności przedstawia tabela 8.

**Tabela 8.**  
**Przyporządkowanie nauczycielskich komentarzy aspektom trafności według S. Messicka**

Uzasadnienie trafności	Funkcje oceny trafności	
	Interpretacja (58,5%)	Zastosowanie (41,5%)
Na podstawie dowodów (48,3%)	Trafność teoretyczna 27,1%	Stosowność/użyteczność 21,2%
Na podstawie konsekwencji (51,7%)	Kształtowanie wartości 31,4%	Skutki społeczne 20,3%

Wyniki w tabeli 8 wskazują na względną równowagę wszystkich aspektów trafności uwzględnianych przez nauczycieli w procesie oceniania sumującego. Pozostaje to w pewnej sprzeczności z wnioskami z badań S. Brookhart (1993), która twierdzi, że przy wystawianiu ocen nauczyciele skupiają się głównie na ich trafności konsekwencyjnej, czyniąc to czasem nawet kosztem ich interpretacji.

Należy jednak wziąć pod uwagę, że analizowane w niniejszej pracy scenariusze dobrane zostały tak, by ukazać sytuację oceniania podobnie uwarunkowaną, ale dotyczącą uczniów znacznie się różniących poziomem osiągnięć. W tym kontekście celowe wydaje się dokonanie zestawienia analogicznego jak w tabeli 8, lecz uwzględniającego odrębność scenariuszy. Pewnym dodatkowym uzasadnieniem w tym względzie jest zbliżona liczebność nauczycielskich komentarzy związanych z każdym ze scenariuszy.

Z porównania tabel 8a i 8b wynika jednoznacznie, że stopień uwzględniania przez nauczycieli w procesie oceniania poszczególnych aspektów trafności zależy nie tyle od sytuacji oceniania, ile od poziomu osiągnięć ucznia.

**Tabela 8a.**  
**Przyporządkowanie nauczycielskich komentarzy związanych ze Scenariuszem I aspektom trafności według S. Messicka**

Uzasadnienie trafności	Funkcje oceny trafności	
	Interpretacja (45%)	Zastosowanie (55%)
Na podstawie dowodów (34,2%)	Trafność teoretyczna 8,7%	Stosowność/użyteczność 25,5%
Na podstawie konsekwencji (65,8%)	Kształtowanie wartości 36,3%	Skutki społeczne 29,5%



**Tabela 8b.**  
**Przyporządkowanie nauczycielskich komentarzy związanych**  
**ze Scenariuszem II aspektom trafności według S. Messicka**

Uzasadnienie trafności	Funkcje oceny trafności	
	Interpretacja (74,4%)	Zastosowanie (25,6%)
Na podstawie dowodów (65,1%)	Trafność teoretyczna 48,8%	Stosowność/użyteczność 16,3%
Na podstawie konsekwencji (34,9%)	Kształtowanie wartości 25,6%	Skutki społeczne 9,3%

W przypadku obydwu scenariuszy relacje procentowe uzasadnień trafności na podstawie dowodów i na podstawie konsekwencji są zbliżone do relacji między wyborami dokonywanymi przez respondentów w zadaniach zamkniętych (tab. 3). Dominacja trafności konsekwencyjnej idzie w parze z decyzją o podniesieniu do dopuszczającego oceny końcowej uczniowi, wśród którego ocen przeważają niedostateczne. Natomiast dowody trafności są dla nauczycieli ważniejsze w sytuacji, w której uczennica otrzymuje na koniec roku dobry, zgodnie z poziomem jej umiejętności określonym wynikami egzaminów powszednich.

Wynika z tego, że uczeń zagrożony brakiem promocji z jednego przedmiotu może otrzymać z niego na koniec roku ocenę dopuszczającą niejako mimo swoich osiągnięć szkolnych. Większość nauczycieli uwzględni bowiem przy podejmowaniu decyzji nie zawartość treściową oceny, ale czynniki kontekstowe niezależne od postawy ucznia wobec obowiązków szkolnych, takie jak: koniec etapu edukacyjnego, pozaszkolne zainteresowania ucznia (niezwiązane z przedmiotem), a także własną wyrozumiałość i przekonania oraz prośby osób trzecich.

Te czynniki kontekstowe, sprzyjające występowaniu w szkole drugiego układu wymagań programowych, nie wpływają w żaden znaczący sposób na decyzje większości nauczycieli dotyczące oceny końcowej odpowiadającej osiągnięciom ponadpodstawowym. By uzasadnić swoją niechęć do wystawienia oceny bardzo dobry, czyli wyższej niż wynika to z ocen bieżących uzyskanych przez uczennicę, sięgają do dowodów potwierdzających trafność dokonanego przez siebie wyboru. Wśród empirycznych uzasadnień przytaczanych przez nauczycieli dominują odwołania do rozumienia oceny jako wynagrodzenia za pracę wykonaną zgodnie z określonymi kryteriami. Pewną kategoryczność podjętych rozstrzygnięć nauczyciele interpretują, odwołując się przede wszystkim do poczucia sprawiedliwości.

Omawiane zagadnienia potwierdzają spostrzeżenie S. Brookhart (1993), że nauczyciel w procesie oceniania musi godzić dwie wewnętrznie sprzeczne postawy: jako „sędzia” powinien rozstrzygać obiektywnie, na jakim poziomie uczeń spełnił wymagania programowe, zaś działając jako „advokat”, bierze zawsze pod uwagę konsekwencje zastosowania ocen. Rozróżnienie to ma to kluczowe znaczenie dla szacowania zawartości treściowej ocen końcowych z przedmiotu.



Jak wynika z dotychczasowych rozważań, rola „adwokata” zyskuje wyraźną przewagę nad rolą „sędziego” w sytuacji, gdy stopień ma decydować o uzyskaniu bądź nieuzyskaniu promocji przez ucznia, co sprawia, że trafność teoretyczna takiej oceny radykalnie spada.

Charakterystyczne jednak jest, że część nauczycieli, przede wszystkim tych, którzy w swych decyzjach o podniesieniu oceny końcowej kierowali się bardziej konformizmem niż swoimi przekonaniem (por. Szyling, 2002: 80), odczuwa pewien dyskomfort z tego powodu. Pociąga to za sobą niezbyt przyjazne dla ucznia oczekiwanie, że przysze doświadczenia, zwłaszcza w szkole średniej, zweryfikują poziom jego umiejętności zgodnie z obiektywnymi kryteriami i „zmuszą go do pracy nad uzupełnianiem braków”. Tak więc część nauczycieli uwzględnia w swoich ocenach konsekwencje ich zastosowania pozostające w oczywistej sprzeczności z wyrozumiałością roli „adwokata”.

Z drugiej zaś strony część z tych nauczycieli, którym bliższa jest rola „sędziego”, wskazuje na kategorie nieosiągnięte przez ucznia, jako dowody uzasadniające zastosowanie oceny do opisanego osiągnięć. Implikuje to rozumienie oceny jako swoistego odzwierciedlenia braków, zwłaszcza związanych ze zbyt małym zaangażowaniem ucznia w osiąganie przez niego celów szkolnych.

## Robocze wnioski

Przedstawiony w niniejszym studium wykaz kryteriów trafności oceny, które uwzględniają nauczyciele podejmujący decyzje o końcowej ocenie z przedmiotu, ukazuje, jak różnie ta ocena jest przez nich rozumiana i jak skomplikowane zależności decydują o jej ostatecznym kształcie. Natomiast dosyć powszechnym oczekiwaniem jest, by ocena końcowa odzwierciedlała przede wszystkim poziom umiejętności ucznia adekwatny do osiągniętych przez niego umiejętności mierzonych w odniesieniu do wymagań programowych.

W tym kontekście nieuchronny wydaje się konflikt między oczekiwaniami odwołującymi się do uzasadnień trafności oceny na podstawie dowodów a postawami nauczycieli, którzy przy podejmowaniu decyzji doniosłych dla uczniów, biorą pod uwagę przede wszystkim trafność konsekwencyjną. Zjawisko to nabiera szczególnej wyrazistości w przypadku ocen wystawianych w ramach drugiego układu wymagań programowych, czyli tam, gdzie warunki pracy szkoły wymuszają nieformalne obniżanie poziomu wymagań podstawowych.

Na konsekwencje tego zjawiska zwraca uwagę P. Knight (2000: 50):

Tam, gdzie ocenianie sprawdzające utożsamia się z faworyzowaniem trafności nad rzetelność, pojawiają się problemy przy próbach jego zastosowania dla doniosłych celów.

Biorąc zatem pod uwagę społeczne zainteresowanie wykorzystaniem ocen uzyskanych przez ucznia na świadectwie ukończenia gimnazjum, należy stwierdzić, że

podstawowym warunkiem jakości takich działań jest prowadzenie intensywnych badań nad trafnością zarówno ocen, jak i egzaminów. Ich wyniki na pewno nie rozwieją wszystkich wątpliwości (por. Skorupiński, 2002), ale pozwolą z dużo większą pewnością podejmować decyzje doniosłe dla ucznia i jego przyszłości.

## Bibliografia

- Brookhart S. (1993), *Teachers' Grading Practices: Meaning and Values*, „Journal Educational Measurement”, Summer, vol. 30, nr 2, s. 123–142.
- Dubiecka A. (2002), *Wewnątrzszkolna ocena ucznia na zakończenie drugiego etapu edukacji a wynik ze sprawdzianu*, [w:] B. Niemiecko, J. Brzdąk J. (red.), *Dwa rodzaje oceniania szkolnego. Ocenianie wewnątrzszkolne i zewnętrzne a jakość pracy szkoły*, Katowice, s. 319–327.
- Knight P. (2000), *Wartość ogólnoprogramowego podejścia do oceniania*, [w:] B. Niemiecko (red.), *Pedagogika, pomiar dydaktyczny i ocenianie szkolne*, skrypt dla uczestników II PSED UG, Gdańsk, s. 43–56.
- Konarzewski K. (2000), *Jak uprawiać badania oświatowe. Metodologia praktyczna*, WSiP, Warszawa.
- Messick S. (1989a), *Meaning and Values in Test Validation: The Science and Ethics of Assessment*, „Educational Researcher”, vol. 18, nr 2, s. 5–17.
- Messick S. (1989b), *Validity*, [w:] B. Niemiecko (red.), *Pedagogika, pomiar dydaktyczny i ocenianie szkolne*, skrypt dla uczestników II PSED UG, Gdańsk, s. 38–42.
- Niemierko B. (1999), *Pomiar wyników kształcenia*, WSiP, Warszawa.
- Niemierko B. (2002), *Ocenianie szkolne bez tajemnic*, WSiP, Warszawa.
- Skorupiński P. (2002), *Trafność interpretacji wyników egzaminu gimnazjalnego w rekrutacji kandydatów do publicznej szkoły ponadgimnazjalnej*, [w:] B. Niemiecko, J. Brzdąk (red.), *Dwa rodzaje oceniania szkolnego. Ocenianie wewnątrzszkolne i zewnętrzne a jakość pracy szkoły*, Katowice, s. 43–52.
- Szyling G. (2002), *„Idź w świat i radź sobie, jak potrafisz”*. *Drugi układ wymagań programowych w oczach uczniów i nauczycieli*, [w:] B. Niemiecko, J. Brzdąk (red.), *Dwa rodzaje oceniania szkolnego. Ocenianie wewnątrzszkolne i zewnętrzne a jakość pracy szkoły*, Katowice, s. 73–86.
- Walczak G. (2002), *Relacje między ocenianiem wewnątrzszkolnym i zewnętrznym*, [w:] B. Niemiecko, J. Brzdąk (red.), *Dwa rodzaje oceniania szkolnego. Ocenianie wewnątrzszkolne i zewnętrzne a jakość pracy szkoły*, Katowice, s. 298–313.