Egzaminy Doniosłe w Oczach Badaczy Polskich[1]

Elżbieta Zawadowska-Kittel
Lingwistyczna Szkoła Wyższa, Warsaw

ABSTRAKT

Artykuł przedstawia problem doniosłych egzaminów zewnętrznych w szkolnictwie polskim. Jego podstawowy cel to ukazanie problematyki związanej z wprowadzeniem egzaminów doniosłych w Polsce i efekty (również niezamierzone) tej decyzji dla systemu edukacji. Główną część artykułu poprzedza wprowadzenie opisujące historię egzaminów doniosłych w Wielkiej Brytanii i Stanach Zjednoczonych oraz analiza definicji tego pojęcia autorstwa Bolesława Niemierki. W artykule poruszony jest również problem funkcji egzaminów doniosłych w kontekście założeń ewaluacji sumującej i kształtującej. Główną część publikacji stanowi problem złudzeń i mitów, jakie powstały w środowisku uczestników procesu edukacji w związku z wprowadzeniem testów zewnętrznych do szkolnictwa polskiego, oraz próba analizy przyczyn, dla których ten sposób testowania okazał się niedoskonały. Wiele miejsca poświęca się tu również opisowi zagrożeń, jakie pociąga za sobą standaryzacja, problemowi edukacyjnej wartości dodanej, a także analizie roli egzaminatora w doniosłych egzaminach zewnętrznych. Artykuł stawia również tezę o konieczności prowadzenia szerokich badań empirycznych w tym zakresie.

Słowa klucz: egzamin doniosły, edukacyjna wartość dodana, ewaluacja kształtująca, ewaluacja sumująca, efekt zwrotny

---

[1] Artykuł stanowi część większego opracowania, zatytułowanego „Efekt zwrotny egzaminów doniosłych na proces nauczania/uczenia się na przykładzie Nowej Matury z języka angielskiego" przygotowywanego przez Elżbietę Zawadowską-Kittel.

ABSTRACT
The article deals with the problem of high-stake centralized examinations in Polish education. Its main aim is to present the effects of introducing this kind of evaluation into the Polish system of education. In the first part of the article the history of high-stakes exams in Poland and in the United States is presented followed by the analysis of the term "high-stakes examinations". The article also deals with the functions of high-stake examinations in connection with the principles of summative and formative evaluation. Its main part is centered on illusions and myths around high-stake examinations and reasons why this form of evaluation has not proved to be perfect. This article explores the dangers resulting from the role of the examiner in teaching/learning and deals with the process of standardization as well as the meaning of the value added system for education. It then presents a rather critical view of high-stakes exams and concludes with a proposal for further research in the area.

Keywords: high-stakes exam, value-added assessment, formative evaluation, summative evaluation, washback effect

Celem artykułu jest omówienie zagadnień, jakie niesie ze sobą wprowadzenie egzaminów zewnętrznych do szkolnictwa polskiego. Proces ten, związany z przemianami ustrojowymi i reformą oświaty, rozpoczął się w 2005 roku, kiedy na szeroką skalę wprowadzono centralne egzaminy maturalne. Ponieważ temat jest w badaniach polskich stosunkowo nowy, literatura przedmiotu nie jest jeszcze tak bogata jak literatura anglojęzyczna – wynika to właśnie z tradycji edukacyjnych. Tym bardziej jednak warto zapoznać się z opiniami badaczy polskich i ich refleksjami na temat zreformowanego sposobu ewaluacji.

1. Czym są egzaminy doniosłe?

W badaniach anglojęzycznych tematyka egzaminów/testów wysokich stawek (ang. *high-stake tests*) nazywanych w Polsce egzaminami doniosłymi (Niemierko „Ocenianie szkolne" 252) zajmuje ważne miejsce już od dawna, ponieważ mają one długą tradycję edukacyjną zarówno w Wielkiej Brytanii, jak i w Stanach Zjednoczonych. Już w cytowanym przez Stephena Wisemana raporcie z 1911 roku odnajdujemy nawiązania do Komisji

Tauntona, która zajmowała się stanem edukacji brytyjskiej w latach 1864–1868. Badania przez nią prowadzone wykazały niewystarczający dostęp społeczeństwa do edukacji na poziomie średnim. Należy zauważyć, że jej udostępnienie stało się obowiązkowe dopiero po uchwaleniu Ustawy Edukacyjnej w 1918 roku. Komisja Tauntona rekomendowała wówczas stworzenie narodowego systemu oświaty bazującego na dotychczas istniejących szkołach dotowanych. W związku z tym członkowie komisji opowiedzieli się za wprowadzeniem do systemu edukacji egzaminów zewnętrznych, które odegrały niezwykle ważną rolę w systemie oświaty brytyjskiej. Spowodowały one wzrost rywalizacji wśród uczniów, wpływając zasadniczo na proces nauczania. Egzaminy takie istniały zresztą od połowy XIX wieku i organizowały je na przykład uniwersytety Oxford i Cambridge (Gillard). Należy wspomnieć, iż ten egzamin doniosły uczniowie brytyjscy zdawali już w wieku jedenastu lat, a jego wynik przesądzał o typie szkoły, w jakiej dany uczeń mógł kontynuować edukację. Raport kończy się podsumowaniem, w którym Komisja stwierdza, że egzaminy zewnętrzne niosą ze sobą potencjalne niebezpieczeństwa dla nauczyciela i ucznia. Mimo to Komisja jest absolutnie przekonana o konieczności ich wprowadzenia w szkołach średnich. Jej członkowie wyrażają jednocześnie pewność, że tego typu egzaminy powinny być poddawane surowym pryncypiom co do ich liczby, wieku zdających, a także ogólnego kształtu. Jednocześnie Komisja wskazuje na konieczność ograniczania niebezpieczeństw związanych z ich wprowadzeniem i maksymalizowania potencjalnych efektów pozytywnych.

Wnioski Komisji Tauntona można uznać za wciąż aktualne w odniesieniu do wszelkich egzaminów zewnętrznych mających wpływ na dalszą karierę ucznia. Nie wymyślono dotąd alternatywnej formuły, która znalazłaby zastosowanie w praktyce, nie budząc kontrowersji i zaspakajając ambicje twórców nowoczesnej edukacji dążących do możliwie obiektywnego testowania i oceniania, a zarazem miałaby jednak również funkcje formatywne. Można zatem chwilowo poprzestać na wniosku Wisemana, który twierdzi, iż „większości negatywnych efektów zwrotnych egzaminów można by uniknąć, jeśli testy zostałyby przygotowane ze świadomością celów edukacyjnych programu nauczania" (162).

Egzaminy doniosłe (ang. *high-stakes tests*[2]) zostały zdefiniowane przez George'a F. Madausa, który stwierdza, że są to egzaminy „postrzegane – zasadnie bądź mylnie

[2] Termin *high-stake tests* tłumaczony bywa czasem na język polski jako „egzaminy wysokich stawek". Wydaje się, że najlepszym ekwiwalentem jest termin „egzaminy kluczowe", lecz by zachować porządek terminologiczny, w artykule używamy konsekwentnie terminu B. Niemierki „egzaminy doniosłe".

– przez studentów, nauczycieli, zarządzających oświatą, rodziców lub opinię publiczną jako egzaminy wykorzystywane do podejmowania ważnych decyzji, które mają na nich natychmiastowy i bezpośredni wpływ" (7). Według Davida W. Chapmana i Conrada W. Snydera siła tego typu testów wynika z faktu, iż są one podstawą do oceny przyszłych możliwości edukacyjnych lub pracowniczych osób im poddawanych. Najważniejszą zatem cechą egzaminów doniosłych jest ich znaczenie dla zdających.

2. Funkcje doniosłych egzaminów zewnętrznych

Egzaminy doniosłe pełnią w różnych krajach różne funkcje i z natury rzeczy mają selektywny charakter, ponieważ służą z reguły do wytypowania najlepszych kandydatów do szkoły na kolejnym etapie edukacyjnym, ze studiami włącznie, lub wyboru najlepszych profesjonalistów (jeśli test ma charakter zawodowy). Mogą również stanowić jedynie zwieńczenie danego etapu edukacji. Ze względu na ich wagę egzaminy takie mają najczęściej charakter zewnętrzny – prowadzone są zwykle przez niezależnych egzaminatorów, z którymi testowany wcześniej się nie zetknął. Takie rozwiązanie jest powszechnie uważane za mniej korupcjogenne i bardziej obiektywne, a zatem wiarygodne. Mimo to doniosłe egzaminy zewnętrzne budzą poważne kontrowersje omawiane w dalszej części artykułu.

Egzaminy doniosłe w Polsce pełniące takie same funkcje jak w krajach anglojęzycznych zaczęły być stosowane dopiero na początku XXI wieku, toteż badania nad ich istotą nie mają zbyt długiej tradycji. Problemem testowania w nauczaniu języków obcych zajmuje się w Polsce na szeroką skalę głównie Hanna Komorowska (2000), testowaniem zewnętrznym w ujęciu pedagogicznym – Bolesław Niemierko, który wykorzystuje w swoich rozważaniach pojęcia ewaluacji sumującej i kształtującej.

3. Ewaluacja kształtująca i sumująca

Już w latach sześćdziesiątych Michael Scriven stworzył podwaliny podziału ewaluacji systemu edukacyjnego, przyjętego również obecnie, na ewaluację kształtującą (formatywną, konstruktywną, bieżącą) „jako wartościowanie systemu edukacyjnego w trakcie jego powstawania" (Niemierko „Ocenianie szkolne" 187) i stanowiącą jej przeciwieństwo ewaluację sumującą (sumatywną, konkluzywną), wartościującą system edukacyjny

w świetle wyników końcowych. Według Niemierki (*Ocenianie szkolne*) ewaluacja sumująca jest charakterystyczna dla systemów autorytarnych – „ocenianie osiągnięć uczniów na wyjściu procesu kształcenia jest ostatecznym, i bywa bezwzględnym rozliczeniem pracy ucznia i nauczyciela" (188), dyscyplinuje kształcenie w systemie. W systemie humanistycznym największe znaczenie przypisuje się natomiast ewaluacji diagnostycznej, która ma stanowić formę pomocy dla ucznia, egzaminy zewnętrzne są w tym ujęciu oceniane zdecydowanie negatywnie. System technologiczny z kolei największą rolę przypisuje ewaluacji kształtującej, której celem jest wspomaganie pracy uczniów w całym procesie kształcenia – dlatego w tego typu systemie ewaluacja sumująca traci sens, gdyż potwierdza tylko wcześniejsze wnioski. Ich wykorzystanie przez nauczycieli nie jest jednak możliwe, gdyż tracą kontakt z uczniem, kiedy dany etap edukacji dobiega końca.

Ponieważ celem procesu kształcenia zawsze powinien być harmonijny rozwój ucznia, autorka podziela pogląd Niemierki (*Ocenianie szkolne*), który stwierdza, że „nie tylko nauczyciele, a może przede wszystkim rodzice uczniów oraz inni przedstawiciele opinii publicznej powinni być zainteresowani ewaluacją kształtującą osiągnięcia uczniów jako pomocą udzielaną młodzieży" (188).

W Polsce najważniejszą funkcję diagnostyczną pełni jednak edukacja sumująca, gdyż właśnie wyniki testów i egzaminów zewnętrznych aplikowanych kandydatom do szkół wyższego szczebla (zwłaszcza po ukończeniu liceum) mają decydujące znaczenie dla ich przyszłej kariery edukacyjnej. Takie rozwiązanie jest jednak dość ryzykowne, czego autorka stara się tu dowieść, omawiając różne aspekty egzaminów zewnętrznych.

4. Egzaminy powszednie i doniosłe

Uczący się w trakcie procesu nauczania zawsze są poddawani różnym formom ewaluacji. Są to, generalnie rzecz biorąc, różnego typu testy nauczycielskie oraz inne formy oceniania wewnętrznego, jak również wprowadzone w Polsce od niedawna egzaminy zewnętrzne przygotowywane przez Centralną Komisję Egzaminacyjną. Największe znaczenie mają dla uczniów egzaminy wieńczące ich dwunastoletnią edukację i będące zarazem egzaminami wstępnymi na studia wyższe, czyli egzaminy maturalne.

Nauczyciele pragną, by uczniowie zaprezentowali się na testach zewnętrznych jak najlepiej, i choć zjawisko takie ma wiele skutków negatywnych, można je z łatwością wytłumaczyć. Na fakt ten wskazuje m.in. Niemierko („Ocenianie szkolne" 252), który wprowadza rozróżnienie pomiędzy egzaminami powszednimi („niskich stawek", ang.

low-stake exams) i egzaminami doniosłymi („wysokich stawek", ang. *high-stakes exams*). I tak egzamin powszedni to egzamin, w którym znaczenie komentarza dydaktycznego jest większe niż znaczenie informacji o wyniku uczenia się, a egzamin doniosły jest egzaminem, w którym znaczenie informacji o wyniku uczenia się jest większe niż znaczenie komentarza dydaktycznego. Jest to egzamin stanowiący czasem jedyną podstawę decyzji o uczniu, a zwykle są to decyzje tak ważne, że związek wyniku z dalszym uczeniem się prawie się zaciera; uczniowi nie zależy na wskazówkach, chce jedynie zdać egzamin.

Dlatego egzaminy doniosłe powodują często negatywny efekt zwrotny, który rozciąga się na wszystkich uczestników procesu edukacji, w tym również na nauczycieli, niejednokrotnie ocenianych przez zwierzchników na podstawie wyników tych egzaminów. Stawiając ważkie pytanie: „Kto potrzebuje egzaminów zewnętrznych?", Niemierko definiuje jednocześnie potrzebę, która jest według niego „uzależnieniem od czynnika (osoby, przedmiotu, okoliczności), który może ją zaspokoić" („Ocenianie szkolne" 264).

Według Adama Brożka (84) wyniki egzaminów mogą być wykorzystywane do autentycznego podniesienia jakości pracy szkoły, zwłaszcza jeśli będą one postrzegane przez odbiorców jako wiarygodne i dydaktycznie użyteczne, a odbiorcy zostaną przygotowani do ich właściwego interpretowania. Ważne jest również, by odpowiednio zinterpretowane wyniki zostały wykorzystane w szkołach do planowania i realizacji działań pedagogicznych, a ich uwzględnienie w tych działaniach miało charakter procesu rzeczywistego, a nie pozorowanego.

Ponieważ spełnienie powyższych warunków zależy od wielu podmiotów systemu edukacyjnego, wymaga ono również zróżnicowanych działań, takich jak właściwie zaplanowane i wdrożone działania systemu doskonalenia nauczycieli, rzeczywiste zaangażowanie nauczycieli i pracowników nadzoru pedagogicznego, a przede wszystkim profesjonalizm i jakość działań Komisji Egzaminacyjnych – Centralnej i Okręgowych (Brożek 84).

Należy zauważyć, że w polskich szkołach wyniki egzaminów zewnętrznych traktowane są zdecydowanie zbyt często jako jedyne skutecznie narzędzie ewaluacji pracy szkoły. Przyczyniają się do tego m.in. nagłaśniane medialnie rankingi szkół (zarówno gimnazjalnych, jak i licealnych), które przybierają czasem charakter działań marketingowych. Dla rodziców zatroskanych o przyszłość dzieci wyniki takie stanowią jedyne miarodajne, rzetelne źródło informacji o szkole, choć są one wyłącznie pewną wypadkową różnych czynników i mogą się zmieniać. Dlatego nawet najwyższa pozycja danej szkoły w rankingu układanym na podstawie wyników egzaminu gimnazjalnego lub Nowej Matury nie daje gwarancji, że wyniki uzyskane przez uczniów w danym roku muszą się powtórzyć. Wprowadzenie egzaminów zewnętrznych skutkuje jednak zwykle przypi-

sywaniem im przez niektórych uczestników procesu edukacji roli jedynego skutecznego wyznacznika jakości. Inne ważne czynniki stanowiące o skuteczności procesu nauczania i uczenia się, takie jak funkcja wychowawcza szkoły, panująca w niej atmosfera czy oferta zajęć pozalekcyjnych, tracą w odbiorze społecznym kluczowe znaczenie, co prowadzi do zubożonego pojmowania roli edukacji w życiu młodego człowieka.

5. Egzaminy doniosłe w oczach uczestników procesu edukacji

Temat egzaminów doniosłych budzi we wszystkich krajach, które je wprowadzają, spore kontrowersje i burzliwe dyskusje. Taka sytuacja ma miejsce również w Polsce. Opinia na temat egzaminów doniosłych różni się w zależności od miejsca zajmowanego przez opiniodawcę w systemie edukacji (bądź poza nim).

Niemierko jest zdania, że uczestnicy procesu edukacji mogą postrzegać egzaminy doniosłe w różny sposób – jedni (nauczyciele i uczniowie) „jako barierę między obecną szkołą i przyszłym awansem" („Ocenianie szkolne" 264). Uczestniczą w procesie egzaminacyjnym motywowani czynnikami ekonomicznymi (dobra praca) i prestiżowymi (zdobycie wykształcenia lub zawodu), ale nie mieliby nic przeciwko temu, by bariera ta została zniesiona, gdyż wywołuje u nich ogromny stres. Według Niemierki „nauczyciele są bardziej skłonni uznać egzaminy doniosłe za niezbędny element systemu kształcenia" („Ocenianie szkolne" 265). Nie znaczy to jednak, że kierują nimi pobudki ekonomiczne. Wiążą jednak egzaminy zewnętrzne z ewentualnym podniesieniem prestiżu szkoły.

Społeczeństwo i władza oczekują natomiast od egzaminów zewnętrznych korzyści ekonomicznych, choć w dalszej perspektywie. Korzyści takie mogą wynikać z podnoszenia kwalifikacji i właściwego doboru pracowników. Daleko ważniejszy jest jednak dla nich zysk polityczny wyrażający się w większej sprawności administracyjnej i zachowaniu ciągłości władzy. Nie bez znaczenia tym procesie jest praca dziennikarzy informujących społeczeństwo o przebiegu i wynikach egzaminów. Jak pisze Niemierko, „rząd, samorządy i władze oświatowe mogą dość nerwowo reagować na ujawnianie skazy na obliczu edukacji narodowej, dając przy tym nadmierną wiarę liczbowym wynikom egzaminów i dziennikarskim komentarzom" („Ocenianie szkolne" 265).

Na podstawie wstępnych badań nad wpływem Nowej Matury z języka angielskiego na proces uczenia się i nauczania w Polsce prowadzonych przez Jo Lewkowicz i Elżbietę Zawadowską-Kittel wprowadzenie Nowej Matury z języka angielskiego miało duży wpływ na nauczanie języka angielskiego w Polsce. Egzamin zmienił sposób nauczania,

co nie znaczy, że spotkał się aprobatą uczących. Najwięcej badanych zgłasza zastrzeżenia do trzydziestoprocentowego progu zdawalności egzaminu oraz do egzaminu ustnego, któremu zarzucana jest sztuczność (mimo iż z założenia egzamin ten miał promować naturalną komunikację).

Z dalszych prac nad tym zagadnieniem podjętych przez piszącą te słowa wynika, że niektórzy nauczyciele zmieniają jednak zdanie na temat egzaminu, dostrzegając w nim również cechy pozytywne (Zawadowska-Kittel 2010). Badania pilotażowe potwierdzają założenie Williama Pophama, iż egzamin doniosły obowiązkowy dla wszystkich wpływa pozytywnie głównie na uczniów słabszych i może działać demotywująco na uczniów zdolniejszych. Egzaminy zewnętrzne postrzegane są zatem różnie przez różnych uczestników procesu edukacji.

Wydaje się, że największe niebezpieczeństwo egzaminów zewnętrznych tkwi w ich „sile rażenia". Jeden niezdany lub słabiej zdany egzamin może bowiem przekreślić wieloletnie wysiłki zarówno uczniów, jak i nauczycieli. Kolejnym problemem, który budzi niepokój, jest rodząca się w Polsce nowa tendencja nauczania „pod test", która stanowi zawsze pokłosie egzaminów zewnętrznych. Wyniki egzaminów są bowiem tak ważne, że dla nauczycieli i uczniów stanowią cel sam w sobie. Co gorsze, przekazywane są one w formie wyłącznie procentowej, nie opisowej, co już z założenia w zbyt wielkim stopniu ogranicza wszelkie idee promujące indywidualizację procesu kształcenia.

6. Złudzenia co do egzaminów doniosłych

Niemierko stwierdza, egzaminy zewnętrzne wiążą się z wieloma złudzeniami co do szans na rozwiązanie kwestii obiektywizmu oceniania osiągnięć uczniów i nauczycieli. Jednym z tych złudzeń jest przekonanie, jakoby standardy edukacyjne, a dokładniej standardy wymagań programowych dokładnie wyznaczały poziom osiągnięć uczniów („Ocenianie szkolne" 267).

Zdaniem Niemierki znajomość standardów edukacyjnych, a zwłaszcza próby ich zastosowania, przekonują jednak o czymś wręcz przeciwnym, a mianowicie o ich chwiejności: „Nawet najściślejsze opisy czynności uczniów nie są w stanie ustanowić normy pozwalającej na bezbłędne rozstrzygnięcie, czy uczeń opanował te czynności" („Znaczenie edukacyjne" 5).

Standardy egzaminacyjne kwestionuje również Klemens Stróżyński („Przymiarka"). Jego zdaniem spełnienie standardów egzaminacyjnych można bowiem stwierdzić

tylko przez pomiar sprawdzający, nigdy różnicujący, a egzaminy zewnętrzne to w istocie pomiar różnicujący, dzięki któremu można określić, który wynik jest niższy, a który wyższy. Egzaminy te nie pozwalają jednak na ustalenie, który wynik jest zadowalający, czyli oznaczający spełnienie jakiejś normy[3].

Słowo „standard" według Stróżyńskiego stało się zaklęciem, składnikiem magii oświatowej, która ma zastąpić rzeczywiste określenie jakości pracy szkół. „Uczymy pod standardy, które nie są standardami, robimy egzaminy według standardów, które nie są standardami, otrzymujemy wyniki, które niewiele nam mówią o przygotowaniach i osiągnięciach uczniów" („Przymiarka" 7).

Nawet najdokładniejsze opisy osiągnięć uczniów nie są w stanie ustanowić normy pozwalającej na bezbłędne rozstrzygnięcie, czy uczeń opanował te czynności. Jeśli zaś punktowanie zadań waha się w ujęciu różnych egzaminatorów, mamy do czynienia z efektem egzaminatora (Niemierko „Znaczenie edukacyjne").

Niemierko („Ocenianie szkolne") podaje przykłady zagadnień, które obejmują standardy testowania doniosłego, czyli – innymi słowy – przepisy regulujące przygotowywanie i stosowanie testów standaryzowanych. Standardy te zostały opracowane na podstawie dokumentu sporządzonego w Stanach Zjednoczonych (1974) i wydane w Polsce pod redakcją Jerzego Brzezińskiego. Autorzy standardów podkreślają m.in. fakt, iż każdy zakres zastosowania testu standaryzowanego powinien być uzasadniony teoretycznie i poparty przykładami w podręczniku, a procedura ustalania oraz interpretacja ewentualnej normy ilościowej wymagań egzaminacyjnych powinny być starannie objaśnione. Szczególną uwagę zwraca grupa wskazówek dotyczących interpretacji wyników testu na skalę kraju. Autorzy standardów dostrzegają bowiem konieczność przygotowania dla uczniów, rodziców, nauczycieli, administracji szkolnej i środków masowego przekazu odpowiedniej informacji o teście, o jego wynikach oraz o najczęstszych błędach interpretacji tych wyników. W praktyce polskiej wyniki egzaminów zewnętrznych są bowiem interpretowane bardzo powierzchownie, jedynie w poważniejszych dyskusjach naukowych poruszany jest temat możliwości błędnej interpretacji wyników testu.

Nie mówi się również powszechnie o kolejnym ważnym problemie uznawanym przez grupę J. Brzezińskiego za jeden z najważniejszych standardów testowania, a mianowicie o ewentualnym sprawdzeniu, czy jakieś lokalne, etniczne, inaczej wyposażone genetycznie lub inaczej kształcone grupy uczniów nie są dyskryminowane przez daną treść

[3] [2]W artykule autor koncentruje się na egzaminach po szóstej klasie szkoły podstawowej, które nie wprowadzają progów.

i formę testu. Takie badania nie są w Polsce prowadzone, choć istnienia takiego zjawiska wykluczyć nie można. Nie jest również stosowana zasada mówiąca, że informacja o błędzie pomiaru powinna być przedstawiana każdemu użytkownikowi jego wyników w takiej formie, jaka odpowiada jego kompetencji statystycznej. Całkowicie ignoruje się zasadę, że informacja o różnicach między wynikami szkół powinna być prezentowana na tle kontekstowych danych o warunkach pracy tych szkół. Naszym zdaniem takich informacji dostarczyć mogą niepodejmowane dotąd badania empiryczne. Jest to problem ważki, bowiem egzaminy zewnętrzne mają być w założeniu obiektywne. Nie należy jednak zapominać, że ich wyniki nie są jedynym elementem diagnostycznym, jaki można wykorzystywać w edukacji. Trudno bowiem zgadywać, jaki wynik mógłby osiągnąć dany kandydat, gdyby uczył się w warunkach innych niż te, w których przyszło mu się kształcić, a zatem nie można oceniać potencjału intelektualnego zdającego wyłącznie na podstawie testu. Dlatego według standardów ważne decyzje o uczniu nie powinny być podejmowane na wyłącznej podstawie egzaminu doniosłego, bez wykorzystania innych źródeł informacji o jego osiągnięciach. Ta zasada jest w Polsce również niemal całkowicie pomijana. Uczelnie muszą przyjmować kandydatów wyłącznie na podstawie pozornie obiektywnego egzaminu doniosłego (Nowej Matury),ęwyjątek stanowią tu jedynie uczelnie artystyczne, którym pozostawiono możliwość sprawdzania tzw. przydatności do zawodu. Przywilej przeprowadzania selekcji wewnętrznej mają również niektóre licea, które poza wynikami testów gimnazjalnych biorą pod uwagę także oceny, jakie uczeń uzyskał na świadectwie. Jakie są jednak kryteria wystawiania tych ocen? Czy stanowią one wypadkową ocen bieżących? Jaką rolę odgrywają w nich wyniki testów sumujących? Odpowiedzi na te pytania będzie można znaleźć jedynie we wnioskach badań empirycznych, jeśli takie zostaną podjęte.

Organizator egzaminów obowiązkowych i badań osiągnięć szkolnych powinien dołożyć starań, by zminimalizować szkody, jakie mogą wyniknąć z niezrozumienia i nadużycia wyników testowania. Dlatego tak ważne wydają się kolejne standardy, również nieprzestrzegane w Polsce, a mianowicie zasady, że udostępniając lub ogłaszając wyniki testowania, jego organizator powinien przedstawić wszelkie dane, które mogą zapobiec błędom w interpretacji tych wyników. Szczególnej uwagi wymagają różnice wyników testowania między szkołami i grupami uczniów, bowiem kontekst kształcenia bywa niedoceniany przez niespecjalistów. W materiałach CKE istnieje wprawdzie podział demograficzny, według którego podawane są wyniki egzaminu w zależności od wielkości miejsca przeprowadzanego egzaminu, nie są jednak prowadzone dogłębne analizy porównawcze między szkołami, grupami, uczniów czy poszczególnymi uczniami, a tylko

taka polityka edukacyjna mogłaby się przyczynić do prawdziwej reformy edukacji opartej na solidnych podstawach.

Kolejnym złudzeniem omawianym przez Niemierkę jest opinia, jakoby krajowi specjaliści wytwarzali zadania najlepiej mierzące osiągnięcia uczniów. Zdaniem Niemierki nie potrafimy jednak stworzyć w pełni trafnych zadań, nie znając autorskich planów wynikowych i sposobu ich realizacji w danej szkole. Według Henryka Szaleńca (2007) pewne oczekiwania wobec narzędzi pomiaru są sprzeczne między grupami użytkowników, których oczekiwania zgodnie z ogólną teorią jakości cytowanego przez niego Williama E. Deminga powinniśmy spełnić.

Dlatego – według Szaleńca – aby dopracować się satysfakcjonującego arkusza, należy zaplanować terminarz kolejno po sobie występujących prac i systematycznie przestrzegać ich realizacji, wypracować koncepcję tworzenia arkusza zgodnie z obowiązującymi standardami egzaminacyjnymi, zadbać o zgodność wymagań egzaminacyjnych z treścią wymagań opublikowanych w informatorach egzaminacyjnych oraz o różnicowanie badanych czynności ucznia i treści kształcenia. Ważna jest również dbałość o trafność zadań oraz o to, by zadania łączyć w związki tematyczne i fabularne, by polecenia były jasne, używane terminy znane uczniom. Należy pytać o rzeczy ważne, pobudzać do myślenia, zadbać, by zadania występujące w sąsiedztwie znacznie nie różniły się od siebie formą. Instrukcja dla ucznia powinna być jasna i wyjaśniać oczekiwany sposób rozwiązywania zadań. Trzeba zwrócić uwagę na wszystkie elementy, które będą podlegać ocenie, a więc np. na poprawność merytoryczną, rozwinięcie, układ, styl, precyzję wypowiadania się, staranność, czytelność itp., a także wypróbować każde zadanie.

W związku z występującymi sprzecznościami opinia o zadaniach najlepiej mierzących osiągnięcia uczniów podnosi wprawdzie autorytet centralnych władz oświatowych i zatrudnionych przez nie ekspertów, ale nie da się utrzymać w świetle teorii pomiaru dydaktycznego. Założenie, jakoby ujednolicenie narzędzi zapewniało nam pełną porównywalność wyniku egzaminu, Niemierko również nazywa złudzeniem i przypomina, że niespecjaliści pomiaru (urzędnicy, dziennikarze, rodzice) łatwo zapominają, że egzamin jest tylko niewielką próbą czynności reprezentującą wielki obszar opanowywanej treści kształcenia („Znaczenie edukacyjne").

Szkodliwym złudzeniem jest również wiara, iż egzaminy doniosłe pozwalają na obiektywną ocenę pracy nauczycieli i szkół. Niemierko („Znaczenie edukacyjne") uważa je za najtrudniejsze do wykorzenienia ze świadomości administracji szkolnej, rodziców i opinii publicznej, ponieważ wyrządza szkodę szkołom pracującym w trudniejszych

warunkach. Nie jest bowiem tak, że egzamin demaskuje braki szkoły bez względu na różnice wstępnego przygotowywania uczniów i kulturę ich środowiska.

Pogląd ten podziela Stróżyński („Przymiarka"), dowodząc, że jeśli dana szkoła ma relatywnie wysokie wyniki, np. w zakresie sprawności czytania, nie musi to koniecznie oznaczać, że w tym zakresie dobrze kształci. Może się bowiem zdarzyć tak, że w tym (lub innym) zakresie niepoślednią rolę odgrywa środowisko domowe uczniów.

Dlatego – zdaniem Niemierki „gdy szkoły są traktowane jako 'dobre' i 'słabe' na podstawie jednej tylko miary, wyniku egzaminu zewnętrznego, popełniany jest podstawowy błąd atrybucji osiągnięć uczniów, polegający na przecenianiu roli szkoły w uzyskiwaniu tych osiągnięć przez uczniów, a niedocenianiu znaczenia sytuacji na wejściu oraz kontekstu w systemie kształcenia. Podobnie jak w przypadku atrybucji dokonywanych przez uczniów przypisujących sukcesy raczej sobie, a porażki raczej czynnikom zewnętrznym, społeczeństwo przypisuje nauczycielom sukcesy i porażki uczniów w egzaminach zewnętrznych bez względu na selekcję kandydatów do tych szkół i warunki, w jakich pracują" („Znaczenie edukacyjne" 8).

Niemierko wprowadza również pojęcie oceniania wąskodydaktycznego („Znaczenie edukacyjne" 5), typowego dla egzaminów zewnętrznych, tj. oceniania osiągnięć i umiejętności z pominięciem sfery społeczno-wychowawczej. Jeszcze bardziej problematyczne jest znaczenie treściowe wyników egzaminu zewnętrznego ograniczone do obowiązkowej części wspólnej programów kształcenia, a zatem do podstawy programowej, z pominięciem elementów autorskich, gdyż z założenia ma być to ocenianie zobiektywizowane, wyzute z kryteriów programowych.

Według Niemierki można mieć bowiem wątpliwości co do obiektywizmu matury pojmowanego jako dokładność, z jaką wyniki tego pomiaru są wyznaczone wymaganiami programowymi. Przy normie ilościowej (próg zdawalności) wynoszącej 30% można mieć poważne wątpliwości co do znaczenia treściowego wyniku egzaminu maturalnego, ponieważ abiturient może nie podjąć znacznej większości zadań lub źle je rozwiązać, a mimo to uzyska świadectwo dojrzałości. Niemierko („Znaczenie edukacyjne") przypisuje tę normę ostrożności, a także chęci zachowania trafności fasadowej matury, stąd jego zdaniem przyszłość wymusi dalsze ustopniowanie jej poziomów. Dlatego – według H. Szaleńca (2007) – należy tym bardziej zadbać o to, by arkusze egzaminacyjne umożliwiały sprawdzenie reprezentatywnej próby całego uniwersum nauczanych treści.

Wydaje się, że Niemierko („Znaczenie edukacyjne") bardzo trafnie podsumowuje powszechne opinie i przekonania na temat egzaminów zewnętrznych, nazywając je złudzeniami. Można dodać, iż są one wysoce szkodliwe, zwłaszcza wiara w omawianą wyżej

moc standardów egzaminacyjnych, a także złudzenie co do pełnej porównywalności egzaminów, co z kolei wynika właśnie z nieprzestrzegania zasad koniecznych do stworzenia warunków choćby pewnej obiektywizacji wyników. Pesymizm Niemierki (ibid.) co do możliwości trafnego skonstruowania zadań egzaminacyjnych wynika również z obaw o to, że jakość nauczania w Polsce nie może być – mimo ciągłych obietnic wyrównywania szans edukacyjnych między ośrodkami – porównywana.

7. Zewnętrzne egzaminy doniosłe a dehumanizacja szkoły

Heliodor Muszyński jest zdania, że nadanie indywidualnej diagnozie edukacyjnej charakteru zewnętrznego „niesie ze sobą cały szereg niebezpieczeństw, które można objąć wspólnym mianem dehumanizacji szkoły" (34). Interesujące jest, że Muszyński ma tu na myśli potencjalne zagrożenia, jakie niesie ze sobą diagnoza nauczycielska przybierająca zewnętrzny charakter. Zdaniem autorki prezentowane przez Muszyńskiego wnioski tym bardziej można odnieść do egzaminów z założenia zewnętrznych, zwłaszcza egzaminów doniosłych. A zatem, wskutek przywiązywania nadmiernej wagi do diagnozy zewnętrznej, przede wszystkim ucierpieć może diagnoza wewnętrzna, z której nauczyciel często rezygnuje na rzecz tej pierwszej. Może też zdarzyć się tak, że nauczyciel ograniczy swoją diagnozę do „tych aspektów rozwoju ucznia, które mają być składnikiem diagnozy zewnętrznej w stosunku do jego relacji z uczniem" (ibid. 34). Znaczy to w praktyce tyle, że nauczyciel będzie się interesował wyłącznie postępami ucznia w zakresie danego przedmiotu. Co więcej, diagnoza zewnętrzna musi z definicji podlegać pewnej standaryzacji, co oznacza, że każdego ucznia rozpatruje się według tych samych kryteriów – tym samym szuka się w nim tego, co ma wspólnego z innymi, a nie tego, co stanowi o jego odrębności. Wreszcie, diagnozy cząstkowe podlegają algorytmizacji, co skutkuje sprowadzaniem ich do wspólnej miary w postaci średniej.

Muszyński dowodzi, że takie tendencje mogą stanowić poważny przyczynek do dehumanizacji szkoły, choć zewnętrzna diagnoza edukacyjna jest elementem naszej rzeczywistości społecznej i oświatowej. Wymienia trzy podstawowe funkcje tej diagnozy: kontrolną (kontroluje funkcjonowanie szkoły), unifikującą (umożliwia każdemu minimum wykształcenia gwarantującego uczestnictwo w kulturze i rozwój), selekcyjną (stanowi do pewnego stopnia podstawę procesów selekcyjnych występujących w społeczeństwie).

Współczesne systemy edukacyjne będą musiały funkcjonować w warunkach zewnętrznej kontroli, a diagnostyka edukacyjna będzie podporządkowana tej funkcji, co

niesie ze sobą poważne niebezpieczeństwa, m.in. tendencje do standaryzacji diagnoz, a zatem ich jednostronności. Generalnie tego typu przemiany prowadzą nieuchronnie do biurokratyzacji szkoły, co z kolei zagraża procesowi jej humanizacji (ibid. 35).

Takie zagrożenie dla humanizacji dostrzega również Niemierko („Ocenianie szkolne" 265), piszący o uczniu jako „numerze kodowym" na egzaminie zewnętrznym. „Uczeń w szkole jest osobą, odróżnianą pod względem (nie zawsze pochwalanych) cech i obyczajów, mającą pewną pozycję i pewien dorobek we współpracy z nauczycielem. To wszystko znika na egzaminie zewnętrznym" („Egzaminy pożądane" 17). Niemierko dostrzega jednak również dobrą stronę tej sytuacji – gdy uczeń występuje jako numer kodowy, pokusy nagradzania ulubieńców i zwalczania oponentów nie mają do egzaminatora dostępu.

W celu humanizacji egzaminów Niemierko („Chłodne oblicze" 18–19) proponuje sześć kierunków działania: rozstanie się z mitem o bezbłędności egzaminów powszednich i doniosłych, a w zamian realistyczną ich ocenę, rozróżnianie pomiędzy egzaminami powszednimi i doniosłymi zamiast ich wartościowania, ulepszanie egzaminów powszechnych i – odwrotnie – ulepszanie egzaminów doniosłych poprzez upodobnianie ich do egzaminów powszechnych oraz wspominane wyżej opieranie prognoz edukacyjnych na obu rodzajach egzaminów. Niemierko (ibid.) podkreśla również, iż zmiany systemowe w zakresie egzaminowania szkolnego wymagają intensywnych badań naukowych, których nie zastąpią konsultacje środowiskowe.

Wydaje się, że wszelka standaryzacja stanowi poważne zagrożenie dla humanistycznego charakteru szkoły. Przykłady można mnożyć. Jednym z jaskrawszych wydaje się choćby narzucanie wszystkim uczniom takiej samej formy opowiadania na maturze z języka angielskiego na poziome rozszerzonym. Opowiadanie jest formą literacką, która powinna wyzwalać w piszącym inwencję twórczą, skłaniać do poszukiwania oryginalnych rozwiązań, również formalnych. Standaryzacja przekreśliłaby szanse takich mistrzów krótkiej formy, jak Ernest Hemingway czy Franz Kafka, na jakąkolwiek karierę literacką. Należy zatem poszukiwać rozwiązań, które oprócz złudnej obiektywizacji wyników umożliwiłyby uczniom pełen rozwój osobowości i nie tłumiły ich talentów, lecz je wyzwalały.

8. Pomiarowe zalety i wady dwu rodzajów egzaminu

Niemierko uważa, że „rozróżnienie egzaminu wewnętrznego, regulowanego przez własnego nauczyciela danych uczniów, i egzaminu zewnętrznego, regulowanego spoza nauczycielskiego systemu kształcenia, nie jest ostre (...), [ponieważ] o typie egzaminu

decyduje nie oficjalny egzaminator, lecz lokacja normowania wyników" („Egzaminy pożądane" 16). Niemierko dopuszcza zatem istnienie egzaminów pośrednich, zewnętrzno-zewnętrznych, takich na przykład jak stara matura, podatna z jednej strony na efekt szarej strefy wymagań programowych, z drugiej strony podlegająca normom staninowym populacji ogólnopolskiej w zakresie układu odniesienia wyniku, o którym jednak decyduje szkoła („Egzaminy pożądane" 16). Należy zatem pamiętać, że egzaminy doniosłe nie muszą być egzaminami zewnętrznymi, choć przedmiotem naszego zainteresowania jest zewnętrzny egzamin doniosły.

Niemierko porównuje egzaminy zewnętrzne i wewnętrzne pod kątem pięciu właściwości sprawdzania osiągnięć („Ocenianie szkolne" 170), wyróżniając: 1) bezstronność sytuacji sprawdzania, 2) dokładność punktowania wyników sprawdzania, 3) rzetelność metody sprawdzania, 4) trafność metody sprawdzania, 5) obiektywizm sprawdzania. Dowodzi, że egzaminy zewnętrzne stoją znacznie wyżej od wewnętrznych pod względem bezstronności rozumianej jako zapewnienie wszystkim uczniom jednakowo sprzyjających warunków do wykazania się opanowaniem przewidzianej treści kształcenia. Zwraca jednak uwagę na fakt, że egzaminom zewnętrznym wciąż daleko do ideału, ponieważ „głębiej rozumiana bezstronność obejmuje nie tylko ujednolicenie (lub wyrównywanie) zadań i organizacji sprawdzianu, lecz także przygotowywanie uczniów, nie wyłączając kulturowych wpływów środowiskowych (Hornowska), oraz pewnego rodzaju indywidualizację, niezbędną we wszystkich przypadkach specjalnych potrzeb edukacyjnych" („Egzaminy pożądane" 17).

Analizując oba typy egzaminów, Niemierko wskazuje na wyższość dokładności punktowania rozumianej jako zgodność z właściwym kluczem również na egzaminie zewnętrznym. Jego zdaniem o tej wyższości decyduje wkład pracy w przygotowanie egzaminu. Podkreśla jednak, że nigdy się nie da wyeliminować nieporadności i pomyłek w kluczu, a zwłaszcza nie można zignorować zjawiska „gubienia diamentów w postaci twórczych interpretacji zadań oryginalnego zapisu działań" („Egzaminy pożądane" 17).

Jako trzecie kryterium porównywalności (jak również przewagi egzaminu zewnętrznego nad wewnętrznym) Niemierko (ibid.) wymienia jego rzetelność rozumianą jako powtarzalność wyniku egzaminu w podobnych warunkach. Niemierko (ibid.) uważa, że w większości przypadków rzetelność testu uzyskuje się dzięki staranniejszej standaryzacji narzędzia pomiaru, wydłużeniu czasu egzaminowania, bardziej analitycznemu podejściu polegającemu na zwiększeniu liczby zadań i kryteriów, zwiększeniu rozrzutu osiągnięć uczniów w próbie przez różne poziomy szkół. Mimo niemożności wyeliminowania

losowych błędów pomiaru, takich jak niepełna dokładność punktowania zadań otwartych i zgadywanie odpowiedzi na zadania zamknięte, i tak udział głównych źródeł błędu jest lepiej kontrolowany w testach standaryzowanych.

Kolejnym kryterium porównywania egzaminów wewnętrznych i zewnętrznych jest trafność pomiaru rozumiana przez Niemierkę (ibid.) jako użyteczność wniosków opartych na jego wynikach do określonego celu, np. przewidywania wyników dalszego uczenia się w szkole wyższego szczebla. W tym przypadku jednak Niemierko (ibid.) dostrzega przewagę egzaminów wewnętrznych, a przyczynę takiego stanu rzeczy upatruje w dwóch cechach egzaminów doniosłych – zawężeniu treści kształcenia i neutralności wychowawczej. Niemierko jest zdania, że należy brać pod uwagę wyniki obu rodzajów egzaminu („Dawne i nowe formy" 19). Przypomina, że dokonano wielu analiz pokazujących, że ani konkurs świadectw, ani egzamin wstępny do szkół wyższego szczebla z osobna nie gwarantują wysokiej trafności prognostycznej. Niemniej jednak połączenie informacji z tych dwu źródeł prawie zawsze przynosi wzrost trafności. Uważa wręcz za bardzo ryzykowne „zachęcanie szkół wyższego szczebla do rezygnacji z innych źródeł informacji o osiągnięciach uczniów niż egzamin doniosły – organizowanych centralnie lub przez nią samą – bez uzyskania dowodów jego trafności, a dokładniej bez dokonania rachunku strat, jakie wskutek tej rezygnacji poniesiemy. Stratą dla szkoły jest przyjęcie niewłaściwych kandydatów, a dla kandydata nieprzyjęcie do szkoły, w której mógłby osiągnąć sukces dzięki cechom emocjonalnym, wykazanym w roku ucznia, a nie jednokryterialnym egzaminom doniosłym" („Dawne i nowe formy" 20). B. Niemierko stwierdza również, że przy porównywaniu cech testu jego trafność stoi w hierarchii wyżej niż rzetelność pomiaru. Dlatego jej niedostatek budzi tak wielką troskę specjalistów pomiaru dydaktycznego („Egzaminy pożądane" 18).

Ilustruje problem, przytaczając model Arlety Poręby-Konopczyńskiej.

Tabela 1. Rzetelność i trafność egzaminów szkolnych według Arlety Poręby-Konopczyńskiej

Właściwości egzaminu	Wysoka rzetelność	Umiarkowana rzetelność	Niska rzetelność
Wysoka trafność	Perły profesjonalizmu	Nauczyciel – światły profesjonalista	Nauczyciel obdarzony intuicją pedagogiczną
Umiarkowana trafność	Egzaminy zewnętrzne	Nauczyciel – dobry rzemieślnik	Codzienna praktyka szkolna
Niska trafność	**Pole minowe egzaminowania**		

Pola minowe zagrażają procesowi kształcenia, a trafność egzaminu zewnętrznego jest ograniczona wyobcowaniem nauczycielskiego systemu kształcenia, z kolei wysoka rzetelność pojawia się tylko na egzaminach zewnętrznych i u najbardziej zaawansowanych pomiarowców. Dlatego egzaminy wewnętrzne cechuje niska lub umiarkowana rzetelność, ale za to wysoka trafność.

Kolejnym kryterium różnicującym egzaminy zewnętrzne od wewnętrznych jest obiektywizm pomiaru, czyli reprezentowanie ustalonych wymagań programowych. Trudność polega tutaj na wskaźnikowaniu jakości osiągnięć, co w przypadku egzaminów wewnętrznych jest incydentalnie rozwiązywane poprzez często niejawne obniżanie wymagań w postaci drugiego układu wymagań programowych i minimalizowanie normy ilościowej.

W przypadku egzaminów zewnętrznych procedury są jawne, toteż w sytuacji, gdy politycy oświatowi odrzucają wielopoziomową strukturę egzaminu, pozostaje stosowanie norm ilościowych, a te z kolei – bardzo przecież niskie – „osłabiają znaczenie treściowe egzaminów, to jest możliwość ustalenia, co umie uczeń uzyskujący te wyniki". Normy 20–30% nazbyt zbliżają się do wartości oczekiwanej (10% i 13%) czystego zgadywania odpowiedzi na zadania wyboru wielokrotnego w obu testach, dopuszczają fragmentaryczną i mało funkcjonalną wiedzę ucznia" (Niemierko „Egzaminy pożądane" 20).

Należy tu zauważyć, że ze zjawiskiem obniżania norm ilościowych mieliśmy jednak do czynienia w Polsce w 2007 roku, kiedy to ówczesny minister edukacji uśrednił wyniki matur ze wszystkich przedmiotów, ustalając średnią na 30%, co umożliwiło osobom, które nie uzyskały z poszczególnych przedmiotów wyniku trzydziestoprocentowego, „nadrobienie" deficytów wynikiem lepszym z innego przedmiotu. Zaistniała wówczas sytuacja (matury nie zdałoby w 2005 i 2006 roku łącznie 53 000 maturzystów) świadczy o niezwykle słabej kondycji polskiej edukacji (www.egzaminy.edu.pl).

Grażyna Szyling jest zdania, że wprowadzenie egzaminów zewnętrznych sprzyja podporządkowywaniu kultury szkoły i oceniania szkolnego wąsko rozumianym wartościom utylitarnym i wiąże się z „traktowaniem kompetencji uczniów jako towaru sprzedawanego na rynku edukacyjnym" (377). Łączy się to ściśle z zainteresowaniem, jakim opinia publiczna darzy wyniki egzaminów zewnętrznych traktowanych jako jedyny miernik realizacji przez szkołę oczekiwań społecznych narzucanych przez dyskurs ekonomiczny zastępujący wartości etyczne hasłami skuteczności. Wartość oceniania zewnętrznego jest zatem zredukowana do jego funkcji stratyfikacyjnej i selekcyjnej, co powoduje zmianę kultury oceniania.

Według Szyling kulturę oceniania wewnętrznego otwartą na możliwości uczniów słabszych i mniej zdolnych przeciwstawia się jednowartościowej kulturze oceniania zewnętrznego, proponującej zwarty system norm postrzeganych pozytywnie przez rynek edukacyjny nastawiony na efektywność kształcenia, co z kolei skutkuje polityką oświatową wykorzystującą pomiar dydaktyczny jako narzędzie sterowania edukacją, natomiast cechą dystynktywną egzaminów zewnętrznych jest troska o rzetelność pomiaru osiągnięć i porównywalność wyników. Dla porównania cechą dystynktywną oceniania wewnętrznego jest komentarz dydaktyczny wspierający rozwój ucznia. Dlatego sposób komunikowania rezultatów egzaminów zewnętrznych znacznie szerzej aniżeli tylko poprzez suchą informację o osiągniętym wyniku procentowym ma duże znaczenie. W Polsce forma opisowa zachowana jest tylko w przypadku egzaminów po szóstej klasie szkoły podstawowej.

Jan Potworowski zwraca natomiast uwagę, iż za kamień węgielny Nowej Matury uznaje się porównywalność, co jest według niego dość zaskakujące, gdyż ważniejsza jest przecież wysoka trafność egzaminu, która „nie kierowałaby całej uwagi na kłopotliwy problem niezależności komisji egzaminacyjnych" (58). Potworowski jest zdania, że ujednolicenie oceniania w formie idealnej polegałoby na tym, by „sklonowani egzaminatorzy" reagowali w identyczny sposób na prace maturzystów.

Pisząca te słowa jest egzaminatorem maturalnym OKE i miała okazję uczestniczyć w wielu dyskusjach prowadzonych podczas oceniania prac pisemnych. Za największy problem uważa zdecydowanie zbyt mało precyzyjny charakter kryteriów oceny. Skoro już bowiem decydujemy się na niepozbawione tylu wad ocenianie zewnętrzne, należałoby je uchronić od dowolności stojącej w rażącej sprzeczności z jakąkolwiek standaryzacją. Taki brak doprecyzowania pewnych kluczowych pojęć, jak na przykład bogactwo językowe, musi skutkować trudnościami w ocenie. Wydaje się, że problem wynika częściowo z nie dość wyraźnego usytuowania egzaminu maturalnego na mapie Europejskiego Systemu Opisu Kształcenia Językowego.

9. Rola egzaminatora w doniosłych egzaminach zewnętrznych

W sytuacji gdy zwolennicy egzaminów zewnętrznych uważają za jedną z ich ważniejszych zalet obiektywizm i porównywalność wyników, należy zdać sobie sprawę z kluczowej roli, jaką w tego typu egzaminach odgrywa egzaminator.

Zasadnicza rola egzaminatora w przeprowadzaniu egzaminów zewnętrznych powinna znaleźć zdaniem Krzysztofa Bednarka i Cezarego Lempy odzwierciedlenie w postrze-

ganiu jego roli jako osoby zaufania publicznego. Aby egzaminator mógł taką rolę pełnić, powinien przestrzegać zasad tajemnicy zawodowej, rozumieć, na czym polega etyka wykonywanej przez niego pracy, musi posiadać wysokie kwalifikacje i wciąż je doskonalić, a wreszcie wyznawać „etos dobrej roboty", co znaczy, że ma mu towarzyszyć poczucie odpowiedzialności za własną pracę, którą przecież wykonuje w żywotnym interesie zainteresowanych. Istotnym mechanizmem, jaki powinien działać w tej grupie zawodowej, jest potencjalna zdolność do autoregulacji.

W niektórych komisjach egzaminacyjnych oczekiwania wobec egzaminatora zostały zapisane w kodeksie etyki egzaminatora opracowanym na podstawie ankiety przeprowadzonej wśród nauczycieli w 2003 roku (Bartmańska). Kodeks ten funkcjonuje czasem jako suplement materiałów dydaktycznych dla egzaminatorów (25). Według kodeksu egzaminator powinien wykonywać swoje obowiązki, przestrzegając zasady szacunku dla prawdy, systematycznie rozwijać swoje umiejętności zawodowe, oceniać zdających uczciwie i z zaangażowaniem, nie oceniać poglądów zdającego, zobowiązany jest do zachowania w tajemnicy wszystkich spraw związanych z sesją egzaminacyjną, ma się starać o osiągnięcie jak najwyższych standardów merytorycznych i organizacyjnych pracy, nigdy nie wykorzystywać swojej funkcji dla korzyści materialnych, a jego powinnością jest stosowanie przyjętych kryteriów oceniania w sposób obiektywny.

Wprowadzenie egzaminów zewnętrznych stworzyło w świecie edukacji nową sytuację. Dotychczasową dwoistość ról (nauczyciel – egzaminator) zastąpiono ocenianiem przez egzaminatorów zewnętrznych, by zobiektywizować ocenianie i uniezależnić je od kryteriów pozamerytorycznych.

Prześledzono dokumentację egzaminu maturalnego (Bartmańska 27), z której wyłania się jednak obraz daleki od ideału. Egzaminatorzy oceniali pozakryterialnie, „na pamięć", intuicyjnie, co skutkowało nierzetelnością. Mimo że do komisji wprowadzono nauczycieli spoza szkoły abiturienta[4], ich rola okazała się marginalna lub żadna. Niejednokrotnie zachowywali się oni pasywnie, by nie prowokować nauczycieli z danej szkoły do „zemsty" na własnych uczniach.

Szaleniec przypomina, że w trakcie egzaminów pisemnych w maju 2005 roku zdarzyły się przypadki ściągania, pewne niedozwolone zachowania uczniów ignorowano z obawy przed odwołaniami. Obserwatorzy nie stwierdzali ewidentnych nieprawidłowości.

[4] W trzecim roku funkcjonowania egzaminu wprowadzono zasadę, że nauczyciel uczący ucznia w klasie maturalnej nie może uczestniczyć w egzaminie, powinien w nim natomiast uczestniczyć nauczyciel z innej szkoły na mocy umowy zawartej między dyrektorami szkół.

Zawiodły zatem zespoły nadzorujące, osoby spoza szkoły, nauczyciele z innych szkół, egzaminatorzy. Analiza prac wskazuje na to, że maturzyści przepisywali prace od kolegów, korzystali z tzw. gotowców, mieli ze sobą wydruki z internetu. Kiedy na maturze z języków obcych zastosowano różne wersje arkuszy, nastąpiło obniżenie wyników, ponieważ uczniowie, wypełniając testy zamknięte, odpisywali bezmyślnie odpowiedzi z innych grup[5].

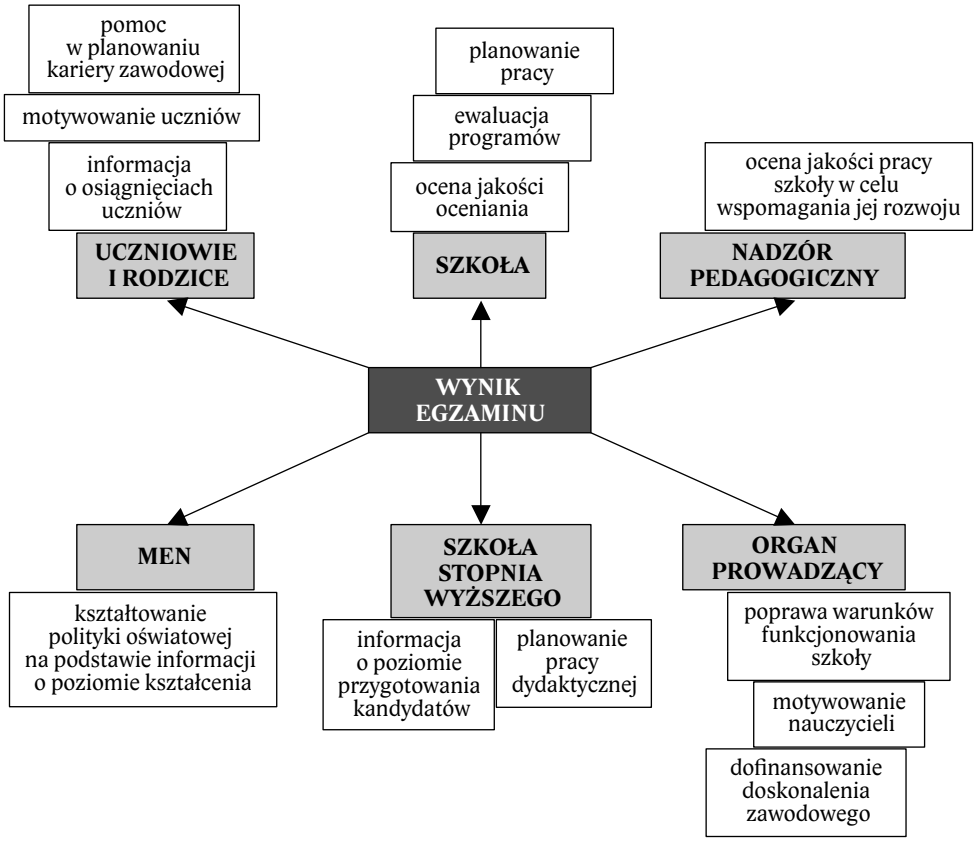
Zmiana sposobu oceniania miała zatem skutki zarówno pozytywne, jak i negatywne. Nauczyciele występujący w roli egzaminatorów zewnętrznych nie stawali przed dylematami moralnymi, jakie towarzyszą im często w sytuacji, gdy muszą oceniać własnych uczniów, występując w schizofrenicznej roli trenera i sędziego zarazem. Niemniej jednak okazuje się, że zakorzenione postawy moralne mogą być silniejsze niż intencje autorów nowego systemu. Zaistniała sytuacja dowodzi zdaniem Anny Bartmańskiej (2007: 28), że dotychczasowa relacja „uczeń kontra szkoła" zmieniła się w relację „szkoła kontra OKE". Przyczyną tego stanu rzeczy jest zmiana roli egzaminu maturalnego, którego wyniki są postrzegane jako narzędzie weryfikacji pracy nauczycieli i uczniów, co prowadzi do bardzo negatywnych konsekwencji.

Egzaminatorzy mają zatem do odegrania jeszcze jedną ważną rolę (Brzdąk). Powinni bowiem „szerzyć kulturę oceniania w szkołach, w których pracują, wpływać na wewnątrzszkolne systemy oceniania, przekazywać swoim uczniom rzetelną informację zwrotną o osiągnięciach i postępach, (...) zachęcać uczniów do dokonywania samooceny i to może mieć znaczny wpływ na motywacje uczniów" (41).

Ta informacja zwrotna o egzaminie powinna trafić do różnych grup i osób związanych bezpośrednio lub pośrednio ze szkołą. Ilustruje to schemat wg pomysłu Romana Dziedzica (2001: 41) podawany przez Jadwige Brzdąke.

Podział uczestników procesu edukacyjnego proponowany w schemacie wydaje się trafny, choć rola, jaką egzamin ten mógłby pełnić w ich życiu, jest chyba zbyt mocno wyidealizowana. By móc ją realizować, egzamin musiałby spełniać wszystkie standardy wymienione na początku rozdziału, a wszyscy uczestnicy procesu edukacji musieliby właściwie pojmować swoje funkcje i należycie je wypełniać. Budzi jednak wątpliwości wiara w możliwość oceny przez szkołę jakości nauczania na podstawie tylko tego jednego kryterium, warto zatem postulować tu rozwiązania wieloaspektowe.

[5] Problem oszustwa egzaminacyjnego omawiany jest szerzej w punkcie 10.



Rysunek 1. Schemat Romana Dziedzica

10. Oszustwo egzaminacyjne

Według Niemierki („Oszustwo egzaminacyjne") oszustwo egzaminacyjne przybiera wielorakie formy, i popełniane jest zarówno przez uczniów, jak i nauczycieli. W przypadku ucznia najczęściej spotykane jego przejawy to: wykradanie zadań, wspólne rozwiązywanie zadań, gdy to jest zakazane, korzystanie z ustnych podpowiedzi, „odglądanie" rozwiązań przez ramię sąsiada, wspólne rozwiązywanie zadań, przepisywanie notatek i innych materiałów, własnych lub użyczonych, przepisywanie cudzej pracy jako własnej (plagiat), zmyślanie wyników własnych „badań", fałszowanie cytatów, wysyłanie zastępcy („Murzyna") na egzamin, ułatwianie oszukiwania innym uczniom (pomaganie im i „krycie" ich), podawanie fałszywej informacji o uzyskanym wyniku („Oszustwo egzaminacyjne" 7). Można zatem określić oszustwo egzaminacyjne jako „ukryte posługiwanie

się informacją uzyskaną z innych źródeł niż własna wiedza i praca, dokonywane dla zdobycia pożądanego wyniku egzaminu" („Oszustwo egzaminacyjne" 8).

W przypadku nauczyciela oszustwo przybiera inne, bardzo wyrafinowane formy. Jeśli chodzi o egzaminy zewnętrzne, jest to podpowiadanie lub wskazywanie uczniom prawidłowych odpowiedzi[6], „przymykanie oka" na oszukańcze zachowania uczniów w trakcie egzaminu, dokonywanie ukrytych poprawek na karcie odpowiedzi ucznia, celowe „mylenie się" w obliczaniu punktów, naginanie norm do wyników i fałszowanie sprawozdań z egzaminu. Jest to zatem celowe naruszanie reprezentatywności treści i wyników egzaminu, wspomaganie i tolerowanie oszustwa egzaminacyjnego uczniów.

Według Kazimierza Denka nawet pojedyncze przypadki ściągania rzutują na opinię o całości. Problem ściągania przybiera charakter coraz bardziej powszechny, zwłaszcza, że wg CBOS-u tylko 28% Polaków nie akceptuje ściągania, 52% nie ma nic przeciwko, a 19% nie ma zdania na ten temat. Dlatego w regulaminach egzaminów zewnętrznych pojawiły się zapisy ostro występujące przeciwko tej popularnej nieuczciwości. Chodzi jednak o to, by polscy uczniowie sami cenili uczciwość, a odpowiedzialność za to spada na szkołę, w której ktoś, kto publicznie mówi, że nie ściąga i nie podpowiada, jest traktowany jako osoba, która nie przestrzega norm współżycia społecznego.

Naszym zdaniem zasługą Nowej Matury jest baczne (choćby teoretycznie) zwrócenie uwagi na problem ściągania. W Informatorze maturalnym czytamy: „Egzamin uważa się za niezdany, jeżeli: a) zdający z któregokolwiek egzaminu obowiązkowego lub jego części ustnej lub pisemnej otrzymał mniej niż 30% punktów możliwych do uzyskania, b) w trakcie egzaminu stwierdzono, że zdający pracuje niesamodzielnie i jego egzamin został przerwany, c) w trakcie sprawdzania egzaminator stwierdził niesamodzielność rozwiązywania zadań egzaminacyjnych (Informator maturalny 2003). Takie wyraźne określenie warunków wynika zapewne po części z przejęcia przez Polskę standardów angielskich. Angielskie *cheat* tłumaczone na język polski jako dobrotliwe „ściągać" znaczy po prostu oszukiwać. Z drugiej jednak strony fakt, iż egzamin ma tak doniosły charakter, do ściągania prowokuje. Dlatego jedynym wyjściem jest bezwzględny brak tolerancji i przyzwolenia społecznego dla ściągania, nie tylko na etapie matury, ale w całym procesie edukacyjnym.

[6] Stąd przepis, że specjalista z danego przedmiotu nie może wchodzić w skład komisji przedmiotowej na egzaminie pisemnym, np. anglista nie może uczestniczyć w egzaminie pisemnym z języka angielskiego. Przepis ten wprowadzono w trzecim roku funkcjonowania NM.

11. Porównanie opinii badaczy anglojęzycznych i polskich na temat egzaminów doniosłych

Badacze anglojęzyczni zajmują się zagadnieniem egzaminów doniosłych znacznie dłużej z racji odmiennych tradycji edukacyjnych, jakie w tych krajach występują. Ponieważ jednak egzaminy doniosłe zaczynają odgrywać dominującą rolę również w Polsce, gdzie aż do początku XXI wieku stosowano ewaluację formatywną, a egzaminy doniosłe występowały jedynie w postaci wewnętrznych egzaminów maturalnych – które jednak zwieńczały jedynie osiągnięcia uczniów szkół średnich, nie pełniąc przy tym roli egzaminów wstępnych – interesujące wydaje się zestawienie opinii badaczy z tych dwóch jakże różnych obszarów edukacyjnych.

Oto zestawienie opinii Angeli Engel i B. Niemierki („Ocenianie szkolne") na temat mitów i złudzeń związanych z egzaminami zewnętrznymi.

Tabela 2. Mity i złudzenia na temat egzaminów doniosłych według A. Engel i B. Niemierki

Angela Engel	Bolesław Niemierko
Egzaminy centralne standaryzowane są postrzegane jako środek podniesienia jakości pracy szkoły.	Standardy dokładnie wyznaczają poziom osiągnięć uczniów.
Dzięki egzaminom centralnym rosną oczekiwania wobec uczniów.	Krajowi specjaliści wytwarzają zadania najlepiej mierzące osiągnięcia uczniów.
Dzięki egzaminom centralnym uczniowie opanowują podstawę programową.	Ujednolicenie narzędzi zapewnia pełną porównywalność egzaminów.
Standaryzacja egzaminów umożliwia państwu, które je wprowadza, rywalizację z innymi państwami.	Egzaminy doniosłe pozwalają na obiektywną ocenę pracy nauczycieli i szkół.
Standaryzacja to najbardziej skuteczny sposób zwiększenia efektów pracy uczniów.	
Opinia publiczna wspiera testy zewnętrzne.	

Z powyższego zestawienia wynika, że zarówno Engel, jak i Niemierko („Ocenianie szkolne") widzą wiele negatywnych stron standaryzacji postrzeganej jako panaceum na podniesienie jakości pracy szkół (Engel) i obiektywną ocenę pracy nauczycieli i szkół (Niemierko „Ocenianie szkolne"). Standaryzacja oceniania nie mieści jednak w sobie pojęcia EWD, czyli edukacyjnej wartości dodanej (Roman Dolata) uważanej za najważniejsze

narzędzie edukacyjne ostatnich dwudziestu lat (Ian Schagen, Dougal Hutchinson). Najprościej rzecz ujmując, wartość dodana to przyrost wiedzy uczniów w wyniku danego procesu edukacyjnego; przyrost, którego nie można określić bez znajomości stanu początkowego w ujęciu prostym oraz kontroli wykorzystania zasobów. Istotne jest również, by pomiaru osiągnięć dokonywano każdorazowo przy pomocy tych skal pomiarowych (www.codn.edu.pl).

W Polsce powstały szacowania EWD dla gimnazjów oraz zaproponowano alternatywne metody szacowania EWD dla szkół maturalnych. Prace nad rozwojem tej metody mają trwać do roku 2013, dlatego trudno w tej chwili określić jej znaczenie dla poprawy sytemu ewaluacji. Nie znaczy to jednak, że system EWD jest niezawodnym sposobem na mankamenty związane ze standaryzacją. W Stanach Zjednoczonych, gdzie system wartości dodanej TVAAS (*Tennessee Value-Added Assessment System*) jest najbardziej rozbudowany, rozgorzała ożywiona dyskusja. Richard Rothstein uważa, że pojęcie efektywności oparte na systemie wartości dodanej nie dostarcza informacji, jakie praktyki edukacyjne za tę efektywność odpowiadają, a co więcej, TVAAS zakłada, jakoby nie istniała interakcja między kapitałem kulturowym ucznia a skutecznością różnych sposobów nauczania, ponieważ wyniki badań (Ross Vasta, Marshall Haith, Scott Miller) dowodzą, że nauczanie szkolne jest najbardziej efektywne wtedy, gdy metody oddziaływania stosowane przez nauczyciela są zgodne z wzorcami znanymi dziecku z doświadczeń kulturowych.

Niemniej jednak EWD mimo tych słabości zapobiega w dużym stopniu frustracjom nauczycielskim powodowanym przez mity o umożliwieniu rywalizacji między szkołami (Engel) i porównywalności egzaminów (Niemierko „Ocenianie szkolne"). Dzięki EWD nauczyciele pracujący wskutek silnej segregacji ekonomicznej w trudnych środowiskach społecznych mogą zyskać świadomość, że inni – pracujący w warunkach podobnych – uzyskują wyniki porównywalne. I odwrotnie, nauczyciele osiągający wyniki słabe mogą porównać je z osiągnięciami kolegów pracujących w podobnych środowiskach i przeanalizować przyczyny swoich porażek, nie przypisując ich li tylko trudnym warunkom pracy i składowi społecznemu szkół (www.ewd.edu.pl).

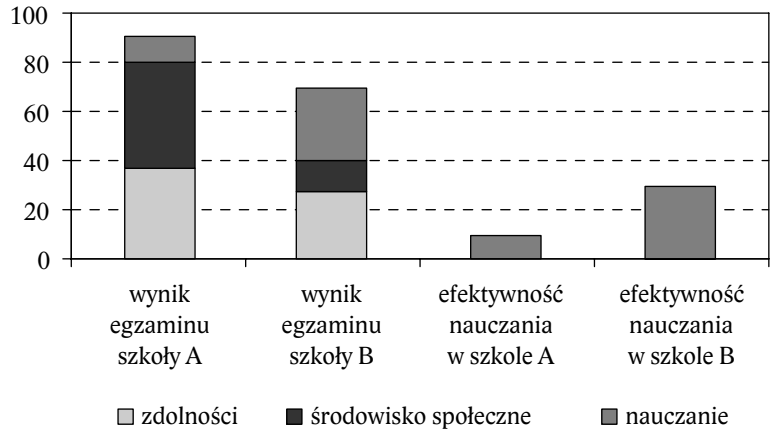
Można zatem wysnuć wniosek, że EWD stanowi – mimo wielu niedociągnięć – istotne uzupełnienie informacji o wynikach egzaminów zewnętrznych, mając jednocześnie świadomość, że EWD jest miarą jakości nauczania w zakresie sprawdzanym przez egzaminy zewnętrzne, dlatego pojawiła się propozycja, by nie mówić o Edukacyjnej Wartości Dodanej, a jedynie o Egzaminacyjnej Wartości Dodanej.

Wykorzystanie badań OKE na reprezentatywnej próbie gimnazjalistów w Krakowie pozwoliło na wyciągnięcie kilku istotnych wniosków korespondujących z zastrze-

żeniami Engel i Niemierki („Ocenianie szkolne") wobec złudzeń i mitów narastających wokół scentralizowanych egzaminów zewnętrznych. Najważniejszym z nich jest stwierdzenie, że EWD jest znacznie słabiej powiązana z charakterystykami społecznymi, kulturowymi i ekonomicznymi rodziny ucznia niż nieprzetworzone (podane w punktach) wyniki egzaminów zewnętrznych (Pokropek). Badanie nie wykazało również związku z nauczaniem „pod testy". W drugim badaniu porównywano gimnazja, grupując je według wyników surowych i wartości EWD (gimnazja o wysokich wynikach surowych i wysokiej wartości EWD, gimnazja o wysokich wynikach surowych i niskiej wartości EWD, gimnazja o niskich wynikach surowych i wysokiej wartości EWD oraz gimnazja o niskich wynikach surowych i niskiej wartości EWD). Tu wyniki badania budzą jednak niepokój, ponieważ w gimnazjach o wysokich wynikach egzaminacyjnych EWD wiąże się jednak pozytywnie z nauczaniem „pod testy", a negatywnie z uczniowską oceną nauczania do autonomii.

Zatem nawet rozbudowanie systemu ewaluacji o dalsze kryteria nie pozwala na dokonywanie uogólnień, można jednak przyjąć, że EWD jest miarą szerzej rozumianej trafności nauczania. Tym bardziej wyniki surowe, na podstawie których dokonywane są między innymi rankingi szkół postrzegane pozytywnie przez opinię publiczną wspierającą testy zewnętrzne (Engel) skonstruowane za pomocą narzędzi, które zapewniają ich porównywalność (co oprotestowuje Niemierko „Ocenianie szkolne"), należy traktować z daleko większym dystansem, aniżeli ma to obecnie miejsce w praktyce polskiej.

Brak korelacji pomiędzy wynikami egzaminów a efektywnością nauczania prezentuje poniższy wykres (www.ewd.edu.pl):



Wykres 1. Czynniki wpływające na wyniki egzaminu a ocena efektywności nauczania

W dyskusji dotyczącej złudzeń na temat egzaminów doniosłych ciekawy jest również sprzeciw A. Engel wobec obiegowej opinii, jakoby dzięki egzaminom centralnym rosły oczekiwania wobec uczniów, którzy dzięki nim opanowują podstawę programową, co w pewnym stopniu koresponduje ze sceptycyzmem Niemierki („Ocenianie szkolne") wobec przekonania o wyznaczaniu przez standardy poziomu osiągnięć uczniów. Jego obawy charakteryzuje najlepiej przenośnia o „sicie gubiącym diamenty" (268).

Choć trudno jest porównywać dokonania badaczy anglojęzycznych i polskich na polu ewaluacji głównie z racji innych ram czasowych, w jakich badania te były prowadzone, analizując poglądy Niemierki („Ocenianie szkolne") i Engel, można dojść do wniosku, że zarówno władze oświatowe wszystkich szczebli, jak i szeroko rozumiana opinia publiczna zbyt łatwo przyjmują te testy za jedyny skuteczny miernik zarówno osiągnięć uczniowskich, jak i jakości pracy szkoły. Taka postawa wynika z pokusy, jaką stwarza wymierność wyniku mierzona w procentach lub ocenach, pewnego wygodnictwa i zapewne braku lepszego rozwiązania na tym etapie badań na temat ewaluacji. Niestety, jak wspominano wcześniej, tak pojęta ewaluacja może doprowadzić do dehumanizacji szkoły i stworzyć cały szereg zagrożeń dla wszystkich uczestników procesu edukacji.

Zapewne lepszym pomysłem byłoby łączenie ewaluacji formatywnej i sumatywnej, to jednak wiązałoby się z pewnością z ogromnymi problemami logistycznymi, a co za tym idzie – finansowymi. Dlatego na tym etapie należy raczej skupić się na wszechstronnej ocenie istniejącego systemu oceniania i testowania i jego ustawicznym udoskonalaniu.

Ciekawy jest również fakt, iż w badaniach polskich jednym z najczęściej omawianych zagadnień jest rola egzaminatora w ewaluacji. Można by odnieść wrażenie, że badacze zjawiska nie do końca wierzą w skuteczność i jakość szkoleń przygotowujących nauczycieli do wypełniania tej funkcji. W jakiejś mierze niepokój i nieufność wynikają zapewne również z dotychczasowych doświadczeń.

Egzaminy doniosłe z uwagi na swoją rangę i znaczenie spotykają się ze zrozumiałym zainteresowaniem całego społeczeństwa: uczestników procesu edukacji, a zatem nauczycieli, uczniów i ich rodziców, władz oświatowych, władz państwowych (stanowiąc element polityki oświatowej, wpisują się w politykę państwa jako całości), są „gorącym tematem" dla dziennikarzy, budzą wiele emocji.

Zdaniem piszącej te słowa egzaminy doniosłe w Polsce powinny stać się tematem szerokich badań empirycznych prowadzonych na wszystkich szczeblach edukacyjnych w ramach wszystkich przedmiotów. Takie projekty, jak badania Xi Luxii nad maturą z języka angielskiego w Chinach, Liying Cheng ("The washback effect") na temat egzaminu maturalnego z języka angielskiego w Hong Kongu, rzuciły nowe światło na postrzeganie

tych egzaminów i metody nauczania. Cheng ("How does washback") stwierdziła jednoznacznie, że egzamin maturalny wywarł efekt zwrotny na system edukacyjny zarówno w skali makro, jak i w skali mikro, jak również na treści nauczania i wykorzystywane materiały, nie zmienił jednak niemal wcale lub zmienił w niewielkim stopniu metody nauczania. Tylko takie badania bowiem mogą się przyczynić do zrozumienia wpływu egzaminów doniosłych na procesy edukacyjne zachodzące w Polsce, zwłaszcza że doniosłe egzaminy zewnętrzne w naszym kraju należy w dalszym ciągu traktować jako innowację, której zasady przeszczepiono z Wielkiej Brytanii, kraju o zupełnie innych tradycjach oświatowych opisanych we wstępie.

Artykuł stanowi zaledwie przyczynek do szerszych dociekań i badań, sygnalizuje złożoność problemów, jakie wynikają z egzaminów zewnętrznych, oraz wieloaspektowy charakter zjawiska. Bardziej wnikliwą analizę zjawiska będzie można odnaleźć w obszerniejszej publikacji przygotowywanej przez autorkę.

Bibliografia

Bartmańska, Anna. „Wpływ egzaminów zewnętrznych na kształtowanie postaw moralnych uczniów, nauczycieli i egzaminatorów." *Edukacja*, *1*. (2007): 22 30.

Brożek, Adam. „Wyniki sprawdzianów i egzaminów zewnętrznych a podnoszenie jakości pracy szkoły." *Język Polski w Szkole IV-VI*. (2001/2002): 79–85.

Brzdąk, Jadwiga. „Rola egzaminów zewnętrznych w zapewnianiu jakości pracy szkoły." *Dyrektor Szkoły 7-8*. (2001): 40–42.

Brzeziński, Jerzy. *Elementy Metodologii Badań Psychologicznych*. Warszawa: PWN. 1984.

Bednarek, Krzysztof, Cezary Lempa. „Egzaminator – osoba zaufania publicznego." *Edukacja*, *1*. (2007): 31–37.

Chapman, David. W. i Conrad W. Snyder. "Can high-stakes national testing improve instruction: Reexamining conventional wisdom?" *International Journal of Educational Development 20*. (2000): 455–474.

Cheng, Liying. *The Washback Effect of Public Examination Change on Classroom Teaching: An Impact Study of the 1996 Hong Kong Certificate of Education on the Classroom Teaching of English in Hong Kong Secondary Schools*. Unpublished doctoral dissertation. University of Hong Kong: Hong Kong, 1988.

–-. "How does washback influence teaching? Implications for Hong Kong." *Language and Education 11*. (1997): 38–54.

Denek, Kazimierz. *Ku Dobrej Edukacji*. Kraków: Akapit, 2005.

Dolata, Roman. „Wykorzystywanie wyników egzaminów zewnętrznych do oceny jakości kształcenia". *Biuletyn Badawczy CKE 3*, (2005): 63–74.

Engel, Angela. "Exposing the myths of high-stakes testing." *Fairtest*. The National Center for Fair and Open Testing 2007.

Gillard, Derek. *Education in England: a brief history*

www.educationenland.org uk/history, 2011.

Hornowska, Elżbieta, *Stronniczość Testów Psychologicznych*. *Problemy – Kierunki – Kontrowersje*. Poznań: Wydawnictwo Fundacji Humaniora, 1999.

Informator Maturalny. Warszawa CKE.2003

Komorowska, Hanna. *Sprawdzanie Umiejętności w NauceJjęzyka Obcego. Kontrola – Ocena – Testowanie*. Warszawa: Fraszka Edukacyjna, 2002.

Lewkowicz, Jo, Zawadowska-Kittel Elżbieta. "Impact of the new school leaving examination in Poland". *Research Notes:* Issue 34, Cambridge ESOL. (2008): 27–31.

Madaus, George F. (1988). "The influence of testing on the curriculum." *Critical Issues in Curriculum: Eighty-seventh Yearbook of the National Society for the Study of Education*. Ed: Laural. N. Tanner, Chicago: NSSE, University of Chicago Press. 83–121.

Muszyński, Heliodor. „Diagnostyka edukacyjna a humanizacja szkoły". Red. Bolesław Niemierko i Wojciech Małecki. *Dawne i Nowe Formy Egzaminowania*? Wrocław: DSWE. 2001.

Niemierko, Bolesław. „Chłodne oblicze egzaminu zewnętrznego." *Edukacja: Studia, Badania, Innowacje*, *3* /75. (2001): 11–26.

–-. *Ocenianie Szkolne bez Tajemnic*. Warszawa: WSIP, 2002.

–-. „Egzaminy pożądane i niechciane." *Edukacja: Studia, Badania, Innowacje*, *4/*88. (2004): 7–26.

–-. „Znaczenie edukacyjne egzaminu doniosłego." *Meritum 3/4* (3). (2006/2007): 4–11.

–-. „Oszustwo egzaminacyjne." *Edukacja: Studia, Badania,Iinnowacje*. (2006): 7–22.

Pokropek, Artur. „Analiza trafności wskaźnika edukacyjnej wartości dodanej." *Biuletyn Badawczy CKE 8*. (2006): 9–20.

Popham, William J. "The Merits of Measurement-driven Instruction." *Phi Delta Kappan*, *May*, (2006): 679–682.

Potworowski, Jan. „Pierwiastek angielski w transformacji polskiej oświaty". *Pedagogika w Pokoju Nauczycielskim*. Red. Krzysztof. Kruszewski. Warszawa: WSiP. 2000.

Rothstein, Richard. *Class and Schools. Using Social, Economic, and Educational Reform to Close the Black-White Achievement Gap.* Washington: Teachers College, University Columbia, 2004.

Schagen, Ian, Hutchninson, Douglas. "Adding value in educational reserach – the marriage of data and analytical power." *British Educational Research Journal, 29/ 5.* (2003): 49–76.

Scriven, Michael. (1967). "The methodology of evaluation". Red. Robert. E. Stake, *Perspectives of curriculum evaluation* 1, Chicago: Rand McNally.(1967): 39–55.

Stróżyński, Klemens. „Przymiarka do mierzenia". *Nowe w szkole 4,* (2002): 2–7.

Szaleniec, Henryk „Oszukiwanie na egzaminie istotnym źródłem majowej porażki." *Edukacja: Studia, Badania, Innowacje 1/9.* (2007): 38-53.

Szyling, Grażyna. „Holistyczne i analityczne metody diagnostyki edukacyjnej – perspektywy informatyczne egzaminów szkolnych*". XI Konferencja Diagnostyki Edukacyjnej.* Gdańsk: Fundacja Rozwoju Uniwersytetu Gdańskiego, (2007): 26–28.

Vasta, Ross., Haith, Marshall. M. i Miller, Scott. *Psychologia Dziecka*. Warszawa: WSiP, 1995.

Wiseman, S. Red.. *Examinations and English Education*. Manchester, England: Manchester University Press. 1961.

Xi, Luxia. *The Intended Washback Effect of the National Matriculation English Test in China: Intention and Reality.* Foreign Language Teaching and Research Press: Beijing, 2004.

Zawadowska-Kittel, Elżbieta. „Efekt zwrotny egzaminu maturalnego z języka angielskiego na proces nauczania i uczenia się (badania pilotażowe)". *Języki Obce w Szkole 4.* (2010): 50–67.

WEBLIOGRAFIA

www.codn.edu.pl

www.ewd.edu.pl