

Bożena ŻUK
konsultant Ośrodka Doskonalenia Nauczycieli
Słupsk

O POTRZEBIE I SPOSOBACH BADANIA TRAFNOŚCI TESTÓW STOSOWANYCH W EGZAMINACH ZEWNĘTRZNYCH

*Bez trafnych narzędzi pomiaru osiągnięć uczniów
standardy edukacyjne są niewiele więcej
niż dekoracją systemu oświatowego.*
Bolesław Niemięko

Wśród wielu wymogów stawianych egzaminom zewnętrznym najważniejszy dotyczył będzie ich jakości, w tym — jakości narzędzi pomiaru osiągnięć uczniów. Pełnej odpowiedzi na pytanie dotyczące jakości testów użytych na egzaminach zewnętrznym nie da się uzyskać na podstawie informacji otrzymanych w wyniku zastosowania nawet najbardziej skrupulatnych wyliczeń parametrów testu i przeprowadzenia wnikliwych analiz statystycznych. O jakości testów mierzących osiągnięcia uczniów, którzy kończą określony etap kształcenia, wiele powiedzą prowadzone przez dydaktyków merytoryczne i rzeczowe analizy wszystkich aspektów trafności stosowanych testów.

Samuel Messick, autor nowej koncepcji trafności, przekonuje, iż nie sposób mówić o trafności testu bez analiz kontekstu działań obejmującego wartości humanistyczne i skutki społeczne dokonywanych pomiarów osiągnięć uczniów. Szeroko rozumiane kryterium trafności testu staje się prymarne we współczesnych teoriach egzaminowania.

Zmiany zachodzące w pracy reformujących się szkół, a w szczególności w pracy nauczyciela z uczniem, ukierunkowane są w dużej mierze na realizację różnorodnych modeli pedagogik humanistycznych, co znajduje swój wyraz między innymi w realizowanych programach kształcenia, zapewniających wszechstronny rozwój uczniom, a także ich nauczycielom. Z chwilą wejścia w życie rozporządzenia MEN z 19 IV 1999 na dotychczasowy rytm pracy szkoły w sposób wyraźny wpływa planowany na 2002 r. sprawdzian-egzamin. Wprowadzone na mocy rozporządzenia powszechne sprawdzanie osiągnięć uczniów kończących poszczególne etapy kształcenia powoduje wzrost rangi egzaminu zewnętrznego, począwszy od sprawdzianu aż do matury, która jest najbardziej doniosłym egzaminem zewnętrznym. Jednak bez względu na charakter i doniosłość egzaminów — zarówno ci, którzy je wprowadzają, jak i ci, którzy doświadczą trudu rozwiązywania zadań testowych, zadają sobie szereg pytań dotyczących celowości i sposobu ich wprowadzania oraz skutków wynikających z podejmowanych działań. Wśród nich są wątpliwości rodzące potrzebę udzielenia odpowiedzi na następujące pytania:

1. Na ile informacje otrzymane w wyniku dokonanego pomiaru są prawdziwe?
oraz

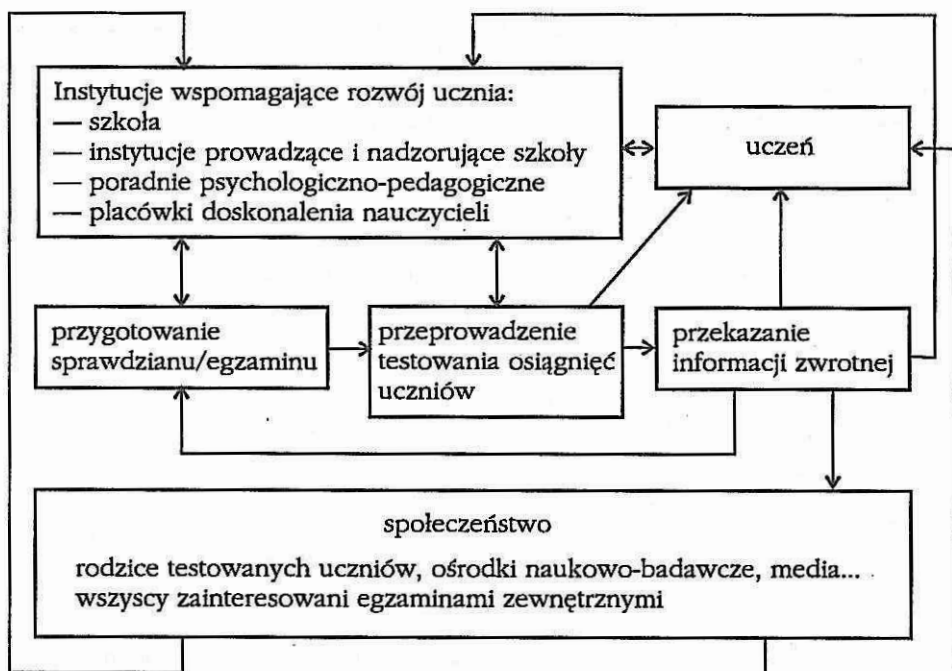
2. Jakie będą jego skutki? Mieszczą się one w pytaniu o trafność egzaminów zewnętrznych.

U progu lat dziewięćdziesiątych XX w., głównie pod wpływem prac Samuela Messicka, ukształtowało się nowe, rozszerzone i jednolite pojęcie trafności pomiaru dydaktycznego. „Trafność — według Samuela Messicka — jest zintegrowanym sądem wartościującym na temat stopnia, w jakim dane empiryczne i uzasadnienia teoretyczne stanowią o tym, że wnioski i działania oparte na wynikach testowania i innych sposobach oceniania są odpowiednie i właściwe”. Ten sąd wartościujący, wykorzystujący „niemal wszystkie formy dowodów trafności” i akcentujący harmonijne powiązanie operacji logicznych z empirycznymi, odpowiada wcześniejszemu pojęciu „trafności teoretycznej”. Tak więc trafność teoretyczna pomiaru znalazła się obecnie w centrum uwagi pedagogów i psychologów, wchłonęła bowiem pozostałe typy trafności pomiaru. Zintegrowane pojęcie trafności jest potrzebne m.in. po to, by objąć nim nowe, niezmiernie ważne dla współczesnej pedagogiki dziedziny analiz: wartości humanistyczne i skutki społeczne przeprowadzanych egzaminów. Akcentuje się pogląd, że wyniki testowania są nie tylko funkcją zadań i innych właściwości bodźca, lecz także — badanych osób i kontekstu oceniania.

W związku z tym w każdym przypadku ustalenia trafności dokonujemy wyniku empirycznej ewaluacji znaczenia i konsekwencji pomiaru. Ważne jest przy tym nie tylko to, co istnieje, lecz także to, co stanowi umysłową reprezentację tej rzeczywistości u działających osób. Stąd też w badaniach nad trafnością nie można ograniczać się do badań trafności zadań testowych czy też całych testów — należy je rozważać jako element trafności egzaminów zewnętrznych. Zastosowanie analiz jakościowych przy rozpatrywaniu poszczególnych aspektów szeroko rozumianej trafności pomiaru dydaktycznego w ujęciu Samuela Messicka pozwoli na zasadne wnioskowanie o osiągnięciach uczniów i podejmowanie skutecznych działań na podstawie wyników dokonywanych pomiarów. Wśród nich nie powinno zabraknąć badań nad praktyczną użytecznością pomiaru i długofalowymi skutkami jego dokonania, wśród których niezmiernie ważne będą badania nad wpływem testowania na postawy uczniów i ich nauczycieli. Z wypowiedzi dyrektorów szkół podstawowych uczestniczących w eksperymencie wałbrzyskim wynika, iż nauczyciele czuli się bardziej odpowiedzialni za osiągnięcia uczniów, a powszechne badanie osiągnięć uczniów klas ósmych było elementem aktywizującym uczniów i nauczycieli do pracy przed sprawdzianem. Równie istotnym elementem badań nad trafnością egzaminów zewnętrznych będzie analizowanie zmian zachodzących w pracy instytucji wspomagających rozwój ucznia i ustalenie ich wpływu na bieżące i odroczone skutki dokonywanych pomiarów osiągnięć uczniów.

Przyjmując, że proces przeprowadzania egzaminów zewnętrznych ilustruje ryc. 1, będziemy mogli prześledzić zależności zachodzące pomiędzy uczestnikami systemu egzaminowego, który stanowi jeden z elementów zmian społecznych. Egzamin odbywają się w kontekście, który stanowią instytucje wspomagające rozwój ucznia, instytucje i osoby zainteresowane egzaminami.

Egzaminy zewnętrzne funkcjonują na zasadzie naczyń połączonych. Powszechne testowanie osiągnięć uczniów oraz przekazywanie i wykorzystywanie informacji zwrotnej po odbytych egzaminach rozpatrywać należy jako stały proces, stąd zamknięta struktura proponowanego schematu przebiegu egzaminów. Stąd też potrzeba badania szeroko rozumianej trafności na każdym ich etapie.



Ryc. 1. Przebieg i kontekst egzaminów zewnętrznych (strzałki oznaczają wzajemny wpływ uczestników egzaminów zewnętrznych)

ETAP I. PRZYGOTOWANIE SPRAWDZIANU/EGZAMINU

Przygotowanie sprawdzianu/egzaminu opiera się na ściślejszej współpracy z instytucjami wspomagającymi rozwój ucznia. O trafności podejmowanych działań mówimy już w momencie podejmowania decyzji o wprowadzeniu określonych rozwiązań oraz w czasie projektowania działań przygotowujących egzaminy zewnętrzne. W przygotowaniach egzaminów zewnętrznych istotne są zarówno procedury organizacyjne, jak i prace związane z opracowaniem arkuszy egzaminacyjnych. Działania przeprowadzone w tej fazie wpłyną z pewnością na przebieg testowania osiągnięć uczniów oraz na trafność informacji zwrotnej przekazywanej po egzaminie. Doświadczenia eksperymentu wałbrzyskiego pokazują, że właściwa organizacja polega nie tylko na terminowym i poprawnym przygotowaniu regulaminu i instrukcji, ale przede wszystkim na pracy — z nauczycielami, dyrektorami szkół, pracownikami samorządów lokalnych — polegającej na przybliżaniu im formuły wprowadzającej zmiany w zakresie egzaminowania oraz w przechodzeniu z jednego do drugiego typu szkoły. Niezbędne są rozmowy, dyskusje, wyjaśnianie — bezpośredniego kontaktu nie zastąpią teksty pisane ani tym bardziej internet. Organizacja to również jak najlepsze opracowanie sposobu działania, podziału pracy. Trzeba przewidzieć najbardziej nietypowe sytuacje — jeśli myślimy, iż coś nie może, nie powinno się wydarzyć, z pewnością się wydarzy. Nie zdołamy opracować procedur rozwiązujących wszystkie problemy natury organizacyjnej, stąd wskazane by było

organizację egzaminów zewnętrznych powierzyć osobom myślącym koncepcyjnie oraz wyjątkowo odpornym na stres. O jakości organizacji dowiemy się analizując przebieg testowania uczniów od momentu przekazania arkuszy z zadaniami testowymi do drukarni do czasu ich powrotu do komisji egzaminacyjnej. Poza informacjami zawartymi w dokumentach towarzyszących testowaniu można śledzić odczucia społeczne, w tym uczestników, poprzez badania ankietowe lub analizę artykułów prasowych.

Opierając się na czteroletnich doświadczeniach związanych z wdrażaniem eksperymentu walbrzyskiego, doskonalenia organizacji, dyrektorzy szkół podstawowych uznali, iż:

- przeprowadzany egzamin obciąża teraz przede wszystkim nauczycieli szkół podstawowych, dezorganizuje w tym momencie pracę szkoły;
- praca komisji nadzorujących przebieg badania była różna w różnych szkołach;
- są szkoły, w których testy rozwiązywane były na salach gimnastycznych w grupie 100 uczniów — praca w tak dużych grupach była niesamodzielną;
- termin badania powinien być dniem wolnym od zajęć dla pozostałych klas;
- należy wprowadzić lepszy i bardziej odpowiedzialny system doboru obserwatorów z zewnątrz towarzyszących egzaminom badania osiągnięć uczniów klas ósmych.

Niezwykle istotny element przygotowań egzaminu zewnętrznego stanowią prace związane z tworzeniem narzędzi pomiaru osiągnięć uczniów. O spełnieniu wymogów stawianych ustandaryzowanym testom z pewnością wiele powiedzą nam wyniki analiz ilościowych uwzględniających miary tendencji centralnej, rozrzutu i korelacji, rozkłady liczebności, analizy opuszczeń zadań, atrakcyjności dystraktorów, łatwości zadań i ich mocy różnicującej oraz korelacji. Jedną z podstawowych miar ilościowych testu jest współczynnik jego rzetelności, który może przybierać wartości od 0 do 1. Zadowolająca rzetelność pomiaru jest warunkiem koniecznym, ale niewystarczającym do oceny jego trafności. Informacje o wynikach analizy statystycznej, w tym o rzetelności testu, w świetle teorii Samuela Messicka stanowią jeden z elementów badań trafności narzędzi pomiaru dydaktycznego.

Stąd też w procesie ustalania trafności testów zastosowanych podczas egzaminów zewnętrznych należałoby skupić się na aspekcie treściowym, obejmującym dobór sprawdzanych czynności i techniczną jakość pomiaru. Przeprowadzona analiza programów kształcenia, operacjonalizacji celów kształcenia, planów testów oraz analiza poszczególnych zadań do wykonania umieszczonych na arkuszu testowym pod względem celów, materiału, wymagań stanowi punkt odniesienia analizy poprawności rzeczowo-merytorycznej zadania oraz analizy poprawności konstrukcyjno-językowej każdego zadania.

Mimo starań konstruktorów analiz wyników standaryzacji co roku pojawiały się zastrzeżenia dotyczące zadań umieszczonych na arkuszu badania kompetencji uczniów (głównie z języka polskiego).

Kolejnym elementem analizy trafności, opartym na modelach teoretycznych odpowiednich procesów poznawczych, emocjonalnych i praktycznych, a zarazem sprawdzającym te modele, jest aspekt rzeczowy. Stosowne będzie przeanalizowanie przez autorów testów taksonomii celów poznawczych, psychomotorycznych, emocjonalno-motywacyjnych w kontekście informacji o człowieku, jakich dostarcza nam współczesna psychologia i pedagogika. Autorom narzędzi pomiaru osiągnięć uczniów wykorzystywanych w eksperymencie zarzucano m.in.:

- brak uwzględnienia indywidualnych potrzeb dziecka, np. obniżenia wymagań;
- niewłaściwy dobór tekstu literackiego.

O trafności testu wiele powiedzą nam wyniki analizy aspektu strukturalnego, dotyczącego zgodności schematów punktowania zadań z jego konstrukcją. Taką wierność może wykazywać układ kryteriów norm oceniania zadań otwartych typu rozszerzona odpowiedź wprowadzonych do testów badających umiejętność formułowania wypowiedzi pisemnej, a także system punktowania zadań wyboru wielokrotnego. Poza analizami dotyczącymi zadań testowych ważne będą wyniki badań jakości pracy egzaminatorów sprawdzających prace uczniów.

Aspekt uniwersalizacyjny, dotyczący możliwości uogólniania wyników pomiaru na inne grupy uczniów, należy rozważać w związku z rzetelnością dokonywanych pomiarów osiągnięć uczniów z języka polskiego i zastosowaniem teorii uniwersalizacji wyników pomiaru w trakcie standaryzacji testu.

Aspekt zewnętrzny trafności rozpatrujemy w połączeniu z efektami korelacji wyników pomiaru z ważnymi kryteriami zewnętrznymi. Korelacja ta powinna być wyższa lub niższa, zależnie od pojęciowej bliskości tych kryteriów z wynikami pomiaru. Aspekt zewnętrzny jest szczególnie doniosły w zastosowaniach pomiaru do orientacji szkolnej i selekcji, co miało miejsce w przechodzeniu uczniów klas ósmych do szkoły ponadpodstawowej według procedur wypracowanych i stosowanych w eksperymencie wałbrzyskim.

Aspekt konsekwencyjny trafności obejmuje aktualne i potencjalne konsekwencje pomiaru, a zwłaszcza te, które dotyczą uznanych wartości społeczno-moralnych, które mogą być wspierane lub podważane wynikami testowania. Wartości te zostały wpisane w koncepcję eksperymentu wałbrzyskiego i dotyczyły:

- upowszechnienia wykształcenia średniego;
- wyrównania szans edukacyjnych uczniów pochodzących z różnych środowisk społecznych;
- obiektywizmu wyniku uzależnionego od wysiłku ucznia.

Mimo starannie prowadzonych działań przygotowawczych jako wady eksperymentu wałbrzyskiego dyrektorzy szkół podstawowych wymienili:

- niedostosowanie zadań otwartych do możliwości kulturowych uczniów z małych środowisk;
- długotrwałą niepewność co do przyjęcia dziecka do wybranej szkoły;
- duży nakład kosztów i czasu rodzica obchodzącego szkoły z wynikami swego dziecka i zebranie o miejsce w Trójmieście;
- dezorganizację pracy szkoły;
- koszty, ponoszone przez gminę i szkołę, placące za zastępstwa nauczycieli, którzy biorą udział w egzaminie, oraz za delegacje obserwatorów.

ETAP II. PRZEBIEG TESTOWANIA OSIĄGNIĘĆ UCZNIÓW

Na niewiele zdadzą się rzetelnie przygotowane regulaminy i instrukcje zawierające szczegółowe procedury, jeśli zawiodą ludzie odpowiedzialni za przebieg testowania osiągnięć uczniów. W wyznaczonym na egzamin czasie wiele się może zdarzyć — dobrego, ale też i złego. Nie pozostaje nic innego jak konsekwentne budowanie etyki uczestnika egzaminu i egzaminatora odpowiedzialnego za jakość przebiegu testowania. Jest to zadanie niełatwe,

które przyjdzie realizować czasami w niezwykle trudnych okolicznościach. Ważna jest świadomość tego, jak istotne dla trafności egzaminów zewnętrznych jest prawidłowe przeprowadzenie badań osiągnięć uczniów. W ustalaniu trafności przebiegu testowania pomocne mogą być także diagramy rozsadzenia uczniów oraz protokoły przebiegu egzaminu.

ETAP III. PRZEKAŻANIE INFORMACJI ZWROTNEJ

Jest to niezwykle ważny element egzaminów zewnętrznych. Podejmując decyzję o tym, co i komu powiemy na temat wyników, jesteśmy odpowiedzialni za skutki, jakie ta informacja wywoła. Zarówno te natychmiastowe, jak i późniejsze. To od formy, użyteczności i jakości informacji przekazywanej na podstawie wyników testowania zależy, jak będą postrzegane egzaminy zewnętrzne przez uczniów, ich nauczycieli oraz rodziców, a także instytucje wspomagające rozwój ucznia, media, ośrodki badawczo-naukowe i wreszcie przez samych inicjatorów systemu egzaminacyjnego. Pozytywny odbiór informacji zwrotnej, możliwość jej wykorzystania w dalszej edukacji zarówno przez ucznia, jak i nauczycieli będą wskaźnikiem wysokiej trafności. W przypadku niespełnienia oczekiwań uczestników procesu egzaminacyjnego, wystąpienia napięć i niepokoїв trudno będzie mówić o trafności działań związanych z przekazywaniem informacji zwrotnej. Eksperyment wałbrzyski pokazał, iż operowanie jedynie punktami informującymi o uzyskanym wyniku nie wpływa na głębsze refleksje ich odbiorców. Każde wzbogacenie informacji cyfrowej o informacje merytoryczno-rzeczowe wpływa na pogłębienie odbioru i owocuje praktycznym wykorzystaniem przez nauczyciela w bieżącej pracy z uczniem. Autorzy informacji zwrotnej opracowywanej na użytek uczestników eksperymentu wałbrzyskiego ewaluowali od punktów, poprzez zamieszczane na zaświadczeniu teksty o charakterze wybitnie motywującym ucznia, aż do bardziej rzeczowej informacji o stanie osiągnięć. Jest to kierunek wart rozwijania. Dla trafności egzaminów zewnętrznych istotne będzie także merytoryczne przygotowanie odbiorców informacji o wynikach testowania w celu odpowiedniego ich spożytkowania — z korzyścią dla ucznia, jego nauczycieli — w codziennej pracy szkoły.

LITERATURA

- Badania ankietowe dyrektorów szkół podstawowych*, 2000, Słupsk (maszynopis).
Niemierko B., 1999, *Pomiar wyników kształcenia*, Warszawa.
Niemierko B., 2001, *Między oceną szkolną a dydaktyką. Bliżej dydaktyki*, Warszawa.
Potulicka E., 1996, *Nowa prawica a edukacja*, Poznań.
Rutkowiak J. (red.), 1995, *Odmiany myślenia o edukacji*, Kraków.
Śliwerski B., 1998, *Współczesne teorie i nurty wychowania*, Kraków.
Wymagania z języka polskiego i matematyki dla ucznia kończącego szkołę podstawową w roku szkolnym 1999/2000, 1999, Wałbrzych.
Żuk B., Tyborczyk J., 1997, *Biuletyn informacyjny — badanie kompetencji uczniów klas ósmych*, Słupsk.