EKONOMETRIA ECONOMETRICS 2(36) • 2012
ISSN 1507-3866

Wiktor Ejsmont
Uniwersytet Ekonomiczny we Wrocławiu

WPŁYW WIEDZY ZDOBYTEJ W SZKOLE PODSTAWOWEJ NA PÓŹNIEJSZY PRZYROST WIEDZY W LICEUM*

Streszczenie: W artykule autor próbuje porównać efektywności nauczania w szkole średniej w zależności od rożnego typu liczenia edukacyjnej wartości dodanej. W tym celu autor używa model do danych panelowych zaproponowany przez M. Aitkina i N. Longforda w 1986 r. W części empirycznej porównuje wyniki otrzymane za pomocą różnych metod. Na tej podstawie zostają wyciągnięte wnioski.

Słowa kluczowe: efektywność nauczania, szkoła podstawowa, analiza danych panelowych.

1. Wstęp

Celem artykułu jest zbadanie zależności między efektywnością nauczania a wiedzą zdobytą w szkole podstawowej. Przy badaniu przyrostu wiedzy zdobytej na etapie kończenia liceum należy się przyjrzeć, jak kształtował się przyrost wiedzy na wcześniejszym etapie nauki. W związku z tym zostanie zbadany wpływ wiedzy zdobytej w szkole podstawowej na późniejsze wyniki w szkole średniej. Sprawdzian szóstoklasisty jest egzaminem przeprowadzanym pod koniec szóstej klasy szkoły podstawowej. Tego typu egzamin ma charakter obligatoryjny, tzn. obejmuje zasięgiem wszystkich uczniów kończących szkoły podstawowe. Aby uzyskać świadectwo ukończenia szkoły podstawowej, należy przystąpić do sprawdzianu[1]. Sprawdzian określa wiedzę ucznia. Nie jest podzielony na różne przedmioty, np. część humanistyczną lub ścisłą. Jego zadaniem jest sprawdzenie wiedzy ogólnej ucznia. Rezultat ma informować o stanie wiedzy końcowej. Nie jest on uwzględniony w rekrutacji do gimnazjów, o ile jest to jedyna szkoła w regionie zamieszkania ucznia oraz jeśli nie

* Ten artykuł powstał przy wsparciu Narodowego Centrum Nauki w latach 2010-2012 i był finansowany jako projekt badawczy nr 3361/BH03/2010/38.

[1] Pierwszy sprawdzian odbył się 10 kwietnia 2002 r. Dotyczy to zarówno szkół podstawowych dla dzieci i młodzieży, jak i dla dorosłych. Są oczywiście wyjątki. Do sprawdzianu nie przystępują uczniowie z upośledzeniem umysłowym (w stopniu umiarkowanym lub znacznym). W pewnych szczególnych przypadkach istnieje możliwość zwolnienia ucznia z powinności przystąpienia do egzaminu.

ukończył 16 lat. Nieprzystąpienie do sprawdzianu jest jednoznaczne z koniecznością powtórzenia ostatniej klasy szkoły podstawowej (oraz przystąpienia do sprawdzianu w następnym roku szkolnym). Uczeń musi się wykazać wiedzą i umiejętnościami z zakresu czytania, pisania, logicznego myślenia, korzystania z informacji oraz wykorzystywania wiedzy w praktyce.

Celem opracowania będzie pokazanie, jak zachowuje się edukacyjna wartość dodana – EWD[2] w zależności od różnego typu jej liczenia oraz ukazanie wpływu szkoły podstawowej na wyniki uzyskiwane na egzaminie maturalnym.

2. Opis danych reprezentujących wyniki maturalne oraz gimnazjalne

Tabela 1 przedstawia zagregowane wyniki na poziomie województw z wybranych 844 liceów. Dane z tab. 1 reprezentują tylko wybrane szkoły, w których uczy się przynajmniej 80 uczniów (w trzech kolejnych latach). Drugim rodzajem danych, jakie przeanalizowano, są dane opisujące uczniów kończących polskie technika (tab. 2). Porównywanie takich wyników jest trudne, ponieważ np. uczniowie zdający maturę w 2010 r. pisali egzaminy gimnazjalne w 2006 r., podczas gdy ich rówieśnicy z liceów w 2007 r. Po uzyskanych średnich wyników egzaminów widać, że trudność egzaminów gimnazjalnych zmienia się w czasie. Dlatego dla większej obiektywności analizowano je osobno. Tabela 2 przedstawia średnie wyniki egzaminów gimnazjalnych oraz maturalnych uczniów kończących technika w różnych województwach, przy czym wybrano 262 polskie technika, a minimalna liczba uczniów w danym roczniku była równa 70. Obniżenie tego progu wiązało się z mniejszą liczbą obserwacji. Procedura spowodowała, że do analizy nie zostały wzięte szkoły z województwa warmińsko-mazurskiego.

Tabela 1. Średnie wyniki egzaminów maturalnych i gimnazjalnych uzyskanych przez uczniów poszczególnych województw z wybranych 844 polskich liceów*

Województwo	Język polski									Matematyka	
	2008			2009			2010			2010	
	liczba uczniów	średnia G-H	średnia M-P	liczba uczniów	średnia G-H	średnia M-P	liczba uczniów	średnia G-H	średnia M-P	średnia G-MP	średnia M-M
1	2	3	4	5	6	7	8	9	10	11	12
Dolnośląskie	10 810	79,52	63,08	11 110	75,43	63,48	10 347	78,89	66,67	67,37	71,52
Kujawsko-pomorskie	6 542	80,12	63,83	7 078	74,46	63,76	6 094	78,08	68,20	67,55	74,18
Lubelskie	11 518	79,05	63,26	12 105	75,93	62,65	11 440	78,73	64,00	64,67	70,14
Lubuskie	3 817	80,10	60,88	3 508	75,03	60,96	3 445	77,36	65,74	64,41	72,11
Łódzkie	8 075	78,56	63,99	9 138	74,83	64,16	5 340	76,88	66,30	66,24	71,99
Małopolskie	14 973	80,35	65,12	15 376	76,72	64,50	15 133	79,54	66,73	67,30	71,79

[2] To pojęcie zostanie wyjaśnione w dalszej części artykułu.

1	2	3	4	5	6	7	8	9	10	11	12
Mazowieckie	22 046	81,40	61,84	22 331	76,90	61,62	21 353	80,31	65,41	69,26	73,79
Opolskie	3 681	77,50	60,91	3 767	74,88	60,22	3 582	78,14	66,40	66,29	72,11
Podkarpackie	1 0401	79,94	63,30	10 828	76,35	61,51	10 217	78,92	64,48	65,51	71,16
Podlaskie	1 820	72,62	63,77	3 914	76,61	64,83	3 758	79,70	67,85	70,36	75,25
Pomorskie	7 409	80,49	63,21	7 970	74,49	63,85	7 125	77,73	69,19	69,06	73,69
Śląskie	17 393	79,49	64,93	17 040	75,50	65,02	16 566	78,15	68,53	65,92	72,29
Świętokrzyskie	4 929	78,34	60,81	5 529	74,90	61,19	3 192	77,63	65,37	64,04	72,44
Warmińsko--mazurskie	997	71,39	65,32	2 040	75,80	64,46	1 906	78,30	67,46	68,96	73,52
Wielkopolskie	13 604	80,89	62,14	13 398	76,13	62,25	12 954	78,37	64,89	65,66	72,73
Zachodnio-pomorskie	6 606	79,48	61,45	6 388	75,49	61,18	5 999	77,26	64,31	63,51	69,67
Razem	144 621	79,81	63,12	151 520	75,82	62,97	138 451	78,72	66,22	66,77	72,31

* G-H, G-MP oznaczają wyniki egzaminów gimnazjalnych z części humanistycznej oraz matematyczno-przyrodniczej. M-P oraz M-M oznaczają egzaminy maturalne części podstawowej z języka polskiego oraz matematyki.

Źródło: Centralna Komisja Egzaminacyjna w Warszawie (2010).

Tabela 2. Średnie wyniki egzaminów maturalnych i gimnazjalnych uzyskanych przez uczniów poszczególnych województw z wybranych 262 polskich techników*

Województwo	Język polski						Matematyka	
	2009			2010			2010	
	liczba uczniów	średnia G-H	średnia M-P	liczba uczniów	średnia G-H	średnia M-P	średnia G-MP	średnia M-M
Dolnośląskie	1 219	67,21	47,30	1 163	61,93	51,09	47,27	46,89
Kujawsko-pomorskie	1 388	71,14	52,80	1 549	63,66	53,88	52,04	54,39
Lubelskie	1 170	66,87	48,33	1 208	64,08	48,14	46,89	48,05
Lubuskie	444	69,46	47,11	423	64,43	54,35	47,32	50,33
Łódzkie	1 073	66,05	48,35	1 222	62,22	50,84	48,87	51,65
Małopolskie	4 461	70,03	51,63	4 435	66,58	55,08	51,03	54,12
Mazowieckie	3 384	67,58	45,75	3 561	63,05	49,07	47,70	47,01
Opolskie	749	66,77	46,50	793	64,30	51,50	50,84	53,87
Podkarpackie	3 463	68,33	47,08	3 595	65,58	49,74	48,98	50,87
Podlaskie	248	71,23	50,31	433	64,69	58,96	55,78	59,30
Pomorskie	1 787	70,24	50,81	1 767	64,62	56,92	52,07	54,69
Śląskie	5 845	69,29	51,91	5 736	65,58	56,96	51,54	54,84
Świętokrzyskie	1 416	65,18	47,41	1 370	62,80	51,43	46,47	48,93
Wielkopolskie	3 542	68,73	48,82	3 606	64,27	49,52	48,73	51,96
Zachodniopomorskie	746	67,87	45,94	728	64,08	49,25	47,16	51,65
Razem	30 935	68,62	49,27	31 589	64,60	52,63	49,70	52,02

* Oznaczenia takie same jak w tab. 1.

Źródło: Centralna Komisja Egzaminacyjna w Warszawie (2010).

3. Metodologia badań

3.1. Model Aitkina-Longforda

Dane zaprezentowane w punkcie 2 nazywane są niezbilansowanymi danymi panelowymi. Panele niezbilansowane występują wówczas, gdy ilość obserwacji dla poszczególnych obiektów jest różna, tzn. n_j. W przypadku równej ilości obserwacji mówi się o danych zbilansowanych. W literaturze modele te są stosowane najczęściej do danych przekrojowo-czasowych, tzn. takich, których dany obiekt jest obserwowany w jakimś określonym czasie.

Oznaczenia:

- x_{ij} – liczba punktów gimnazjalnych uzyskanych przez *i*-tego ucznia w *j*-tej szkole (wejście),
- y_{ij} – liczba punktów maturalnych uzyskanych przez *i*-tego ucznia w *j*-tej szkole (wyjście),
- n_j – liczba uczniów w szkole j,
- n – liczba wszystkich uczniów, tzn. $n = n_1 + ... + n_k$,
- k – liczba szkół, tzn. $j \in \{1,...,k\}$,
- $\bar{x}$ – średni wynik gimnazjalny wszystkich uczniów,
- $\bar{y}$ – średni wynik maturalny wszystkich uczniów,
- $\bar{x}_j = \left(\sum_{i=1}^{n_j} x_{ij} \right) / n_j$, $\bar{y}_j = \left(\sum_{i=1}^{n_j} y_{ij} \right) / n_j$ – średni wynik gimnazjalny oraz maturalny na poziomie j-tej szkoły.

Zastosowany model to model z czynnikami losowymi. W ekonometrii model ten zawdzięcza popularność artykułowi Balestry i Nerlove'a [1966], traktującemu o popycie na gaz ziemny. Gdy populacja, którą chcemy opisać, nie jest jednorodna, należy uwzględnić ową niejednorodność w modelu. Jeśli elementy w próbie pochodzą z dużej populacji, lepiej założyć, że indywidualny efekt jednostkowy jest realizacją pewnej zmiennej losowej. W modelu tym występują dwa składniki losowe. Model z czynnikami losowymi znany jest też pod nazwą modelu komponentów wariancyjnych (*variance components model* – VC lub *error component model*). Model ten jest postaci[3]:

$$y_{ij} = \alpha + \beta x_{ij} + \xi_j + e_{ij}.. \tag{1}$$

[3] W kontekście beneficjentów szkolnictwa model ten został pierwszy raz opisany przez Aitkina i Longforda [1986], stąd nazwa podrozdziału.

W modelu zakłada się:

- e_{ij} – zmienna losowa o rozkładzie $N(0,\sigma^2)$,
- ξ_j – zmienna losowa o rozkładzie $N(0,\sigma_I^2)$,
- składniki losowe pochodzące z różnych szkół i dla różnych uczniów są nieskorelowane,
- indywidualny składnik losowy ξ_j jest nieskorelowany ze składnikiem losowym e_{ij} (tzn. $E(\xi_j, e_{is}) = 0$).

Z powyższych założeń wynika:

$$\operatorname{var}(y_{ij}) = \operatorname{var}(\xi_j + e_{ij}) = E(\xi_j + e_{ij})^2 - E^2(\xi_j + e_{ij})$$

$$= E(\xi_j^2 + 2\xi_j e_{ij} + e_{ij}^2) = \sigma_I^2 + \sigma^2, \quad (2)$$

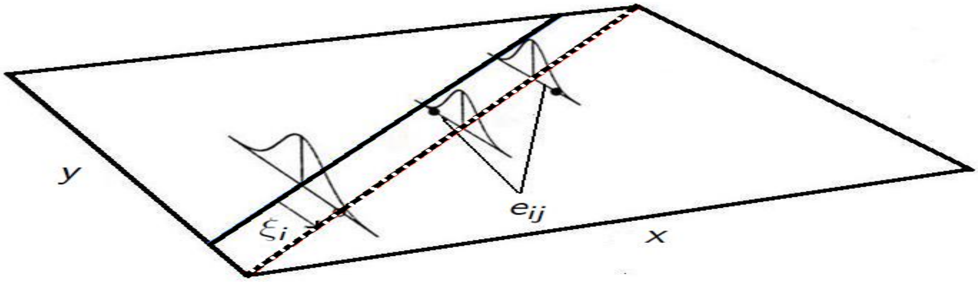
$$\operatorname{cov}(y_{ij}, y_{pj}) = \operatorname{cov}((\xi_j + e_{ij}),(\xi_j + e_{pj})) = E(\xi_j^2 + \xi_j e_{ij} + \xi_j e_{pj} + e_{ij} e_{pj}) = \sigma_I^2,$$

$$\rho = cor(y_{ij}, y_{pj}) = \frac{\sigma_I^2}{\sigma_I^2 + \sigma^2}. \quad (3)$$

Współczynniki modelu (1) szacujemy za pomocą największej wiarygodności (np. [Atkin, Longford 1986]) lub uogólnioną metodą najmniejszych kwadratów (np. [Baltagi 2005]). Estymator parametrów α oraz β jest postaci:

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{k} w_j & \sum_{j=1}^{k} w_j \bar{x}_j \\ \sum_{j=1}^{k} w_j \bar{x}_j & \sum_{j=1}^{k}\sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 + \sum_{j=1}^{k} w_j \bar{x}_j^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j=1}^{k} w_j \bar{y}_j \\ \sum_{j=1}^{k}\sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)(x_{ij} - \bar{x}_j) + \sum_{j=1}^{k} w_j \bar{x}_j \bar{y}_j \end{bmatrix}, \quad (4)$$

gdzie $w_j = n_j\sigma^2 / (\sigma^2 + n_j\sigma_I^2)$. Więcej na temat estymacji parametrów modeli panelowych można przeczytać w pracy [Ejsmont 2009], gdzie szczegółowo opisany jest cały algorytm estymacji, w tym również komponentów wariancji σ^2 i σ_I^2. Poniżej opisano procedurę liczenia efektywności uczenia (zastosowanych przez Aitkina i Longforda). Zestawienie obiektów odbywa się za pomocą porównania wartości oczekiwanej składnika losowego ξ_j (wzór 1). Składnik ten mówi, o ile od uśrednionego wyniku całej populacji odchyla się uśredniony wynik j-tego obiektu. Na rys. 1 przerywaną linią został oznaczony uśredniony wynik j-tego obiektu, ciągła linia zaś przedstawia uśredniony wynik całej populacji (czynnik e_{ij} odpowiada za odchylenie od uśrednianego wyniku na poziomie j-tego obiektu). Jeżeli wartość ξ_j jest dodatnia, wówczas możemy powiedzieć, że j-ty badany obiekt poczynił postęp w stosunku do uśrednionego wyniku całej populacji, jeśli zaś jest ujemna, wówczas uzyskał wynik niższy niż uśredniony wynik badanej populacji.



Rys. 1. Schemat przedstawiający ideę mierzenia przyrostu wiedzy modelem Aitkina-Longforda

Źródło: opracowanie własne na podstawie [Skrondal, Rabe-Hesketh 2008, s. 96].

Aby oszacowywać wartość składnika ξ_j (nie jest ona znana), wykorzystamy poniżej cytowane twierdzenie o błędzie średniokwadratowym (np. [Jakubowski, Sztencel 2004, s. 135]).

Twierdzenie. Załóżmy, że dany jest wektor losowy (A, B), gdzie zmienna A jest obserwowana, zaś zmiennej B nie możemy obserwować. Jeżeli $E(B^2) < \infty$, wtedy optymalna prognoza (dla B) w sensie błędu średniokwadratowego istnieje i można wziąć $E(B / A)$.

Ponieważ składniki σ^2 oraz σ_I^2 są znane przed oszacowaniem modelu, możemy więc tę informację wykorzystać jako informację *a priori*. Wyznaczymy rozkład warunkowej zmiennej losowej ξ_j pod warunkiem $\bar{y}_j$ (podejście Bayesowskie). Ze wzoru (1) średnia na poziomie j-tej szkoły wyraża się wzorem:

$$\bar{y}_j = \alpha + \beta \bar{x}_j + \xi_j + \bar{e}_j \ . \tag{5}$$

Przy poczynionych założeniach $\bar{y}_j$ ma rozkład normalny $N(\alpha + \beta \bar{x}_j, \sigma_I^2 + \sigma^2 / n_j)$. Ten rozkład przyjęto jako rozkład *a priori*. Ponieważ ξ_j jest zmienną losową z rozkładu $N(0, \sigma_I^2)$, więc rozkład warunkowy $f(\xi_j / \bar{y}_j)$ też będzie rozkładem normalnym.

Uwaga. Znany jest następujący fakt z rachunku prawdopodobieństwa. Jeżeli zmienne losowe $X_1 \sim N(\mu_1, \sigma_1^2)$ i $X_2 \sim N(\mu_2, \sigma_2^2)$ oraz $\rho_{1,2} = cor(X_1, X_2)$, to rozkład warunkowy X_1 / X_2 jest postaci

$$N\left(\mu_1 + \rho_{1,2}\frac{\sigma_1}{\sigma_2}(X_2 - \mu_2), \sigma_1^2(1-\rho_{1,2}^2)\right).$$

Stąd uwzględniając fakt $\rho' = cor(\xi_j, \bar{y}_j) = \sigma_I^2 / (\sigma_I \sqrt{\sigma_I^2 + \sigma^2 / n_j})$, wnioskujemy, że $f(\xi_j / \bar{y}_j)$ ma rozkład normalny w postaci:

$$N\left(\rho' \frac{\sigma_I}{\sqrt{\sigma_I^2 + \sigma^2 / n_j}}(\bar{y}_j - \alpha - \beta\bar{x}_j), \sigma_I^2(1-\rho'^2)\right)$$ lub w innym zapisie

$$N\left(\rho n_j^*(\bar{y}_j - \alpha - \beta\bar{x}_j), n_j^*(1-\rho)\sigma_I^2 / n_j\right), \tag{6}$$

gdzie $n_j^* = w_j / (1-\rho)$. Porównanie szkół będzie się opierało na porównaniu wartości średnich z rozkładu warunkowego zadanego wzorem (6). Stąd edukacyjną wartość dodaną (EWD) zdefiniowano w postaci

$$e_j = \hat{\rho} n_j^*(\bar{y}_j - \hat{\alpha} - \hat{\beta}\bar{x}_j). \tag{7}$$

W celu sprawdzenia, czy uzyskane efekty losowe są istotne, użyjemy testu Breuscha-Pagana (np. [Baltagi 2005]). Jest to test mnożników Lagrange'a, w którym mamy hipotezy:

$H_0: \sigma_I^2 = 0$

$H_1: \sigma_I^2 \neq 0$.

Statystyka testowa jest postaci

$$LM = \frac{\left(\sum_{j=1}^{k} n_j\right)^2}{\left(2\sum_{j=1}^{k} n_j(n_j - 1)\right)} \left[\frac{\sum_{j=1}^{k}\left(\sum_{i=1}^{n_j} e'_{ij}\right)^2}{\sum_{j=1}^{k}\sum_{i=1}^{n_j} e'^2_{ij}} - 1\right]^2 \sim \chi^2(1), \tag{8}$$

gdzie e'_{ij} są to reszty otrzymane w wyniku zastosowania metody MNK do wszystkich danych (niezależnie od szkół). Powyższy wzór mówi, że przy założeniu prawdziwości hipotezy zerowej statystyka testowa LM ma asymptotyczny rozkład chi-kwadrat z jednym stopniem swobody. Hipotezę zerową odrzucamy, jeżeli wartość statystyki LM należy do prawostronnego obszaru krytycznego.

3.2. Przyrost wiedzy w gimnazjum

Tabela 3 prezentuje wyniki badania przy wykorzystaniu danych prezentowanych w tab. 1 i 2. Niezależnie od badanego rocznika oraz rodzaju uczniów (kończący licea lub technika) współczynniki beta dla języka polskiego są zbliżone do poziomu 0,5. Wyraźnie wyższe są one dla matematyki, stąd tempo wzrostu wiedzy w 2010 r. było wyższe w przypadku matematyki. Wariancja, która szacuje zróżnicowanie wewnątrzszkolne, była wyraźnie większa dla matematyki.

Tabela 3. Podstawowe charakterystyki statystyczne modelu efektów losowych

<table>
<tr><td></td><td colspan="4">Uczniowie kończący licea</td><td colspan="5">Uczniowie kończący technika</td></tr>
<tr><td>Przedmiot</td><td colspan="3">język polski</td><td>matematyka</td><td colspan="3">język polski</td><td colspan="2">matematyka</td></tr>
<tr><td>Charakterystyki</td><td>2008</td><td>2009</td><td>2010</td><td>2010</td><td>2009</td><td colspan="2">2010</td><td colspan="2">2010</td></tr>
<tr><td>Współczynnik korelacji (Pearsona)</td><td>0,452</td><td>0,457</td><td>0,412</td><td>0,668</td><td>0,476</td><td colspan="2">0,453</td><td colspan="2">0,665</td></tr>
<tr><td>Wariancja składnika losowego – σ^2</td><td>121,3</td><td>119,04</td><td>146,568</td><td>170,744</td><td>117,854</td><td colspan="2">149,483</td><td colspan="2">201,432</td></tr>
<tr><td>Wariancja międzyszkolna – σ_I^2</td><td>14,520</td><td>11,674</td><td>17,238</td><td>11,522</td><td>13,487</td><td colspan="2">23,566</td><td colspan="2">22,582</td></tr>
<tr><td>p-value normalność</td><td>>0,01</td><td>>0,01</td><td>>0,01</td><td>>0,01</td><td>>0,01</td><td>>0,01</td><td colspan="2">>0,01</td><td>>0,01</td></tr>
<tr><td>Współczynnik beta</td><td>0,516</td><td>0,485</td><td>0,516</td><td>0,600</td><td>0,429</td><td>0,488</td><td colspan="2">0,538</td><td>0,784</td></tr>
<tr><td>Współczynnik alfa</td><td>21,752</td><td>25,868</td><td>32,145</td><td>31,898</td><td>24,987</td><td>15,745</td><td colspan="2">17,847</td><td>13,005</td></tr>
<tr><td>LM – p-value</td><td><0,01</td><td><0,01</td><td><0,01</td><td><0,01</td><td><0,01</td><td><0,01</td><td colspan="2"><0,01</td><td><0,01</td></tr>
</table>

Źródło: obliczenia własne za pomocą programu Excel oraz R-project na podstawie danych z Centralnej Komisji Egzaminacyjnej w Warszawie (2010).

Celem jest pokazanie, jak kształtuje się efektywność nauczania w liceum w zależności od różnego jej typu liczenia. Eksperyment polega na tym, aby w pierwszej kolejności wyliczyć przyrost wiedzy w gimnazjum. Przyrost obliczono na podstawie różnicy między egzaminem gimnazjalnym a prostą regresji wyznaczoną na podstawie równania regresji między wynikiem testu szóstoklasisty a wynikiem gimnazjalnym odpowiednio analizowanej części (np. humanistycznej)[4]. Stosując ową procedurę, otrzymano wynik, który odpowiada przyrostowi wiedzy pojedynczego ucznia w gimnazjum (względem wiedzy w szkole podstawowej). Następnie zastosowano model ze wzoru (1), zastępując x_{ij} zmienną g_{ij} równą $g_{ij} = x_{ij} - \mu_1 - \mu_2 p_{ij}$, gdzie μ_1, μ_2 są współczynnikami prostej regresji utworzonej na podstawie wyniku testu szóstoklasisty p_{ij} a egzaminem gimnazjalnym x_{ij}.

[4] Dopasowanych do wszystkich danych niezależnie od szkoły.

Tabela 4 przedstawia średnie wyniki sprawdzianu szóstoklasisty uczniów, których wyniki maturalne oraz gimnazjalne są opisane w tab. 1 i 2. Dane, podobnie jak to miało miejsce w poprzednich typach egzaminów, zostały przeskalowane do poziomu 100 punktów (na sprawdzianie szóstoklasisty można otrzymać od 0 do 40 punktów). Warte zaznaczenia jest to, że najlepsze wyniki (niezależnie od rocznika) otrzymali uczniowie uczący się w szkołach podstawowych, które odpowiednio reprezentują takie województwa, jak dolnośląskie, pomorskie lub kujawsko-pomorskie. Nie należałoby w tym przypadku wyciągać zbyt pochopnych wniosków odnośnie do najlepiej uczących szkół podstawowych, ze względu na to, że średnie wyniki znacząco od siebie nie odstają.

Tabela 4. Uśrednione wyniki sprawdzianu szóstoklasisty uczniów, których wyniki maturalne oraz gimnazjalne są opisane w tab. 1 i 2*

Województwo	Uczniowie kończący licea			Uczniowie kończący technika	
	rocznik zdawania matury			rocznik zdawania matury	
	2008	2009	2010	2009	2010
Dolnośląskie	87,469	85,941	80,981	77,875	74,703
Kujawsko-pomorskie	86,363	83,924	80,399	79,582	75,045
Lubelskie	84,811	83,657	76,502	73,682	72,831
Lubuskie	83,862	82,205	78,377	75,738	73,958
Łódzkie	85,797	84,036	78,975	73,920	70,659
Małopolskie	86,318	84,488	79,108	77,714	75,411
Mazowieckie	87,187	81,030	79,723	75,970	68,595
Opolskie	86,616	84,934	80,345	77,727	76,371
Podkarpackie	85,771	83,555	77,543	76,113	74,112
Podlaskie	80,560	85,100	81,583	79,798	76,468
Pomorskie	86,829	85,413	80,814	77,487	75,854
Śląskie	85,822	84,723	79,331	76,614	75,702
Świętokrzyskie	84,513	82,808	77,906	74,258	73,650
Warmińsko-mazurskie	80,093	84,662	80,699		
Wielkopolskie	84,957	82,786	78,797	74,339	72,450
Zachodniopomorskie	83,850	82,294	77,987	72,941	71,606
Razem	85,842	83,645	79,176	76,360	73,844

* Część uczniów z tab. 2 została pominięta ze względu na to, że nie posiadano w ich przypadku wyniku testu szóstoklasisty, co ostatecznie złożyło się na 242 technika, w których prześledzono dane przynajmniej 70 uczniów.

Źródło: Centralna Komisja Egzaminacyjna w Warszawie (2010).

Tabela 5 przedstawia charakterystyki modeli liczonych względem przyrostu wiedzy w gimnazjum. Porównując modele efektów losowych liczonych na podstawie wiedzy zdobytej w gimnazjum (tab. 3), otrzymano, że wszystkie współczynniki beta są mniejsze, stąd wniosek, że przyrost wiedzy w liceum był „wolniejszy" niż liczony względem wyników gimnazjalnych.

Tabela 5. Podstawowe charakterystyki statystyczne modeli efektów losowych liczonych względem przyrostu wiedzy w gimnazjum

Przedmiot	Uczniowie kończący licea				Uczniowie kończący technika		
	język polski			matematyka 2010	język polski		matematyka
Rocznik	2008	2009	2010		2009	2010	2010
Wariancja składnika losowego – σ^2	131,626	127,162	151,514	209,394	133,601	167,168	260,675
Wariancja międzyszkolna – σ_I^2	13,447	10,128	14,678	14,692	15,234	32,517	40,089
p- value normalność	>0,01	>0,01	>0,01	>0,01	>0,01	>0,01	>0,01
Współczynnik beta	0,376	0,407	0,347	0,615	0,369	0,397	0,654
Współczynnik alfa	30,479	28,579	38,416	22,983	49,277	52,594	51,909
LM – *p-value*	<0,01	<0,01	<0,01	<0,01	<0,01	<0,01	<0,01

Źródło: obliczenia własne za pomocą programu Excel oraz R-project na podstawie danych z Centralnej Komisji Egzaminacyjnej w Warszawie (2010).

4. Otrzymane rezultaty oraz wnioski

Tabela 6 przedstawia zależności korelacyjne między różnymi typami liczenia EWD. Należy tutaj wytłumaczyć, że np. w przypadku języka polskiego EWD było liczone dwojako. W pierwszej kolejności względem części humanistycznej egzaminu gimnazjalnego[5], w drugiej kolejności względem przyrostu wiedzy w gimnazjum (liczonego w taki sposób, jak opisany w poprzednim podpunkcie). W przypadku matematyki sytuacja jest analogiczna, z tym wyjątkiem, że obliczano względem części matematyczno-przyrodniczej. W ten sposób wyliczono EWD dla każdej szkoły i zbadano zależność korelacyjną – na wektorze długości 844 (liczba liceów). Zauważalne jest „mocne" skorelowanie pomiędzy EWD liczonym względem egzaminów gimnazjalnych oraz maturalnych w przypadku tych samych przedmiotów w tym samym czasie (wartości przekątnej). Przekraczają one w każdym z przypadków 0,9. Otrzymane rezultaty pokazują, że liczenie EWD względem wyników gimnazjalnych oraz przyrostu wiedzy z poprzedniego okresu daje bardzo podobne rezultaty. Wynika stąd, że przyrost wiedzy na etapie gimnazjum jest czynnikiem determinującym efek-

[5] W tabeli jest to oznaczone w części pionowej.

ty maturalne na etapie późniejszej nauki. Zauważalne są również słabsze korelacje pokazujące zmienność w czasie między różnymi typami liczenia EWD.

Tabela 6. Korelacyjna macierz zależności pomiędzy EWD liczonym względem wyniku gimnazjalnego oraz przyrostu wiedzy w gimnazjum dla wybranych polskich liceów

		EWD liczone dla liceum względem przyrostu wiedzy w gimnazjum			
		język polski 2008	język polski 2009	język polski 2010	matematyka 2010
EWD liczone dla liceum względem egzaminu gimnazjalnego	Język polski 2008	0,971	0,533	0,478	0,476
	Język polski 2009	0,551	0,954	0,521	0,508
	Język polski 2010	0,494	0,512	0,981	0,498
	Matematyka 2010	0,469	0,507	0,465	0,934

Źródło: obliczenia własne na podstawie programu Excel na podstawie danych z Centralnej Komisji Egzaminacyjnej w Warszawie (2010).

Tabela 7. Korelacyjna macierz zależności pomiędzy EWD liczonym względem wyniku gimnazjalnego oraz przyrostu wiedzy w gimnazjum dla wybranych polskich techników

		EWD liczone dla technikum względem przyrostu wiedzy w gimnazjum		
		język polski 2009	język polski 2010	matematyka 2010
EWD liczone dla liceum względem egzaminu gimnazjalnego	Język polski 2009	0,975	0,434	0,325
	Język polski 2010	0,443	0,983	0,384
	Matematyka 2010	0,509	0,551	0,956

Źródło: obliczenia własne na podstawie programu Excel na podstawie danych z Centralnej Komisji Egzaminacyjnej w Warszawie (2010).

Analogiczne obliczenia zostały przeprowadzone również dla uczniów kończących technika, których uśrednione wyniki gimnazjalne oraz maturalne prezentuje tab. 2. Korelacyjna macierz (tab. 7) zależności pomiędzy EWD liczonym względem wyniku gimnazjalnego oraz przyrostu wiedzy w gimnazjum pokazuje, że analogiczne wyniki są zauważane w przypadku uczniów kończących technika. Otrzymane rezultaty pokazują, że liczenie „przyrostu względem przyrostu” daje podobne rezultaty jak liczenie na podstawie wiedzy wejściowej. Na koniec warto jeszcze wspomnieć, że podobnego typu badania były prowadzone również przez innych. Carpita, D’Ambra, Vichi oraz Vittadini [2006] w pierwszej części swojej książki (rozdz. V, s. 125) pokazali przyrosty wiedzy różnych uczniów na czterech różnych etapach nauki. Otrzymane wyniki tworzą układ rosnący. Powodem jest inna specyfika analizowanych przez nich danych.

Ukazane zależności niosą ze sobą także inne przesłanie. Przede wszystkim należałoby podkreślić dużą wiarygodność przeprowadzonych analiz. Niezależnie od rocznika oraz rodzaju szkoły (liceum, technikum) zawsze otrzymano wysokie współczynniki korelacji. Stąd wniosek, że poczynione badania można również odnieść do szkół wyższych. Mianowicie duży przyrost wiedzy w szkole ponadgimnazjalnej będzie miał istotny wpływ na przyrost wiedzy przyszłych studentów. Coraz częściej można usłyszeć, że uczelnie wyższe mają problemy z uczniami, którzy przychodzą na studia, mając małą wiedzę ścisłą (matematyczną). Przeprowadzone rozumowanie ilościowe pokazuje, że na to, jak będą się kształtowały wyniki uczniów, ma wpływ wcześniej zdobyta wiedza. Należałoby się więc spodziewać, że słaby system polskiego szkolnictwa ponadgimnazjalnego jest przyczyną posiadania mniejszej wiedzy przez przyszłych absolwentów wyższych uczelni. Stad wniosek, że jeżeli państwo chce mieć wysoko wykwalifikowanych absolwentów uczelni wyższych, należy zwrócić większą uwagę na szkodnictwo „niższych szczebli”.

Literatura

Aitkin M., Longford N., *Statistical modelling issues in school effectiveness studies,* “Journal of the Royal Statistical Society” 1986, vol. 149, no. 1.

Balestra P., Nerlove M., *Pooling cross section and time series data in the estimation of a dynamic model: The demand for natural gas*, “Econometrica” 1966, vol. 34, no. 3.

Baltagi B., *Econometric Analysis of Panel Data*, John Wiley & Sons Ltd 2005.

Carpita M., D’Ambra L., Vichie M., Vittadini G., *Valutare la qualità: i servizi di pubblica utilità alla persona*, Edizioni Guerini e Associati, Milano 2006.

Ejsmont W., *Efektywność nauczania we wrocławskich liceach*, Didactics of Mathematics 5-6 (9-10), Wydawnictwo UE, Wrocław 2009.

Jakubowski J., Sztencel R., *Wstęp do teorii prawdopodobieństwa*, SCRIPT, Warszawa 2004.

Skrondal A., Rabe-Hesketh S., *Multilevel and Longitudinal Modeling Using Stata*, College Station, Texas: Stata Press Publication – StataCorp LP. 2008.

IMPACT OF KNOWLEDGE ACQUIRED IN ELEMENTARY SCHOOL ON THE SUBSEQUENT INCREASE OF KNOWLEDGE IN HIGH SCHOOL

Summary: In the article the author tries to show that knowledge acquired in elementary school is important in high school. For this purpose he uses models for panel data. The use of panel models to measure the effectiveness was launched by M. Aitkins and N. Longford in 1986. In the second part of the article the author applies data to the model and draws conclusions from the obtained results.

Keywords: school effectiveness, elementary school, analysis of panel data.