

**dr Dorota Węziak-Białowolska**

Instytut Badań Edukacyjnych

## **Skala motywacji testowej – analiza właściwości psychometrycznych**

### **Wprowadzenie**

Motywację człowieka w literaturze przedstawia się jako stan gotowości (wewnętrznej lub zewnętrznej) do podjęcia określonego działania w celu uzyskania określonego efektu (Reykowski 1985). W sytuacji gdy stan gotowości jest stanem uświadomionym, efekt ten określa się mianem celu. Poziom charakteryzującej człowieka motywacji zależy nie tylko od rodzaju podejmowanego działania, ale również od czasu trwania tego działania. Innymi słowy, pewne działania nie wymagają wysokiego poziomu motywacji, inne natomiast – zwłaszcza te długo trwające – charakteryzują się tym, że nawet początkowo wysoki poziom motywacji zwykle obniża się w miarę zbliżania się do końca trwania działania. Ponadto występujący w danym momencie trwania działania poziom motywacji zależy od wysiłku, jaki należy włożyć w osiągnięcie zamierzonego efektu i ryzyka związanego z daną czynnością. Jego wzrost (spadek) wynikać może ze wzrostu (spadku) prawdopodobieństwa sukcesu lub porażki w działaniu ukierunkowanym na osiągnięcie zamierzonego efektu.

W przypadku egzaminów i testów wydaje się, że duże znaczenie dla jakości otrzymywanych wyników ma motywacja wewnętrzna. Pojawia się ona, gdy człowiek dąży do zaspokojenia swoich własnych potrzeb (Reykowski 1985), zaś podejmowane działania i wkładany w nie wysiłek wynikają w dużej mierze z jego wewnętrznego przekonania. Motywację taką w odniesieniu do uczestnictwa w egzaminach określa się w literaturze mianem motywacji testowej. Motywację testową Dolata i Pokropek (2010, s. 86) określają jako działanie mające na celu „podjęcie zadania, wysiłek poznawczy i wytrwałość w dążeniu do rozwiązania zadań testowych”. Natomiast Baumert i Demmrich (2001, s. 441) określają ją jako chęć do zaangażowania się w pracę nad rozwiązaniem testu lub egzaminu, włożenia w te czynności wysiłku, który to wysiłek utrzymać się będzie przez cały czas trwania egzaminu.

Uzyskanie takiego stanu rzeczy jest bardzo trudne. Zdaniem Erwin i Wise (2002, s. 71) zmotywowanie uczniów do rzetelnego podejścia do egzaminu, w sytuacji gdy wyniki tego egzaminu nie mają żadnego lub mają niewielkie znaczenie dla kariery edukacyjnej ucznia, jest najprawdopodobniej największym wyzwaniem, ale i największym utrapieniem dla osób zaangażowanych w proces oceny umiejętności. Jak zauważa Beguin (2000), w przypadku każdego testu, który przeprowadzany jest przed lub po właściwym egzaminie

doniosłym (ang. *high-stakes*), bez względu na jego cel<sup>1</sup>, należy liczyć się z niższą motywacją testową. Ta zaś może, choć nie musi, obniżyć trafność uzyskanych wyników oraz spowodować ich obciążenie. O możliwości wystąpienia takich niedokładności piszą między innymi Kiplinger i Linn (1993), Mislevy (1995), Napoli i Raymond (2004), O'Neil, Sugrue i Baker (1995/1996), Wainer (1993), Wolf i Smith (1995). W Polsce, według najlepszej wiedzy autora, badania mające na celu zbadanie związku między motywacją testową a wynikami egzaminów o niewielkim znaczeniu dla kariery edukacyjnej ucznia nie były podejmowane. Związki takie próbowano oceniać, ale jedynie przy okazji innych badań (zob. Dolata, Pokropek 2010).

Przegląd wyników badań nad związkiem między motywacją testową a uzyskanym wynikiem egzaminu pokazuje, że wnioski w tym zakresie nie są jednoznaczne. Są autorzy, którzy twierdzą, że jeśli uczeń postrzega test lub egzamin jako mało ważny, to może nie być chętny do pełnego zaangażowania się w jego rozwiązanie (Beguin 2000), zaś uzyskane przez niego wyniki są względem rzeczywistych umiejętności zaniżone (DeMars 2000). Niemniej jednak wyniki innych badań wniosku tego nie potwierdzają. Po pierwsze, okazuje się, że związek między poziomem motywacji testowej a uzyskiwanymi wynikami może być bardzo słaby bądź też nie istnieć (Dolata i Pokropek 2010). Po drugie, wyniki badań prowadzonych przez m.in. Baumert i Demmrich (2001), Center for Educational Testing and Evaluation (2001), czy Eklof (2006) pokazują, że nawet jeśli uczniowie postrzegają test jako test o niewielkim znaczeniu, to też mogą być mocno zmotywowani, aby włożyć duży wysiłek w jego jak najlepsze rozwiązanie.

Bez względu na to, jaki w rzeczywistości jest związek między poziomem motywacji testowej a uzyskiwanymi przez uczniów wynikami, warto podejmować działania mające na celu polepszenie postrzegania wagi egzaminu<sup>2</sup>. Niemniej jednak warto również tworzyć takie narzędzia, które zagwarantują, aby uzyskane wyniki były odporne (ang. *robust*) na różnice w poziomie motywacji testowej między egzaminami różnej wagi. W badaniu motywacji testowej stosuje się zwykle dwa podejścia (Dolata i Pokropek 2010): (1) podejście eksperymentalne i (2) podejście korelacyjne. W pierwszym przypadku badanie umiejętności uczniów traktuje się jak eksperyment i manipuluje się poziomami czynników przyjętych za te, które mają znaczenie dla poziomu motywacji testowej w celu jej zwiększenia. Drugie podejście związane jest z konstrukcją odpowiednich narzędzi do korekty wyników egzaminów ze względu na poziom motywacji testowej. Są to zwykle kwestionariusze do samodzielnego wypełnienia po zakończeniu pisania egzaminu służące badaniu motywacji testowej (Eklof 2006, Sundre 1999).

<sup>1</sup> Na przykład może być to badanie mające na celu (1) uzyskanie danych do zrównania wyników egzaminacyjnych, (2) wystandaryzowanie nowej wersji egzaminu/testu, (3) uzyskanie danych do oceny ewaluacyjnej pracy szkoły czy polityki edukacyjnej.

<sup>2</sup> Po pierwsze, badania prowadzone przez O'Neil i in. (1997) pokazały, że w przypadku testu NAEP płacenie za każdą dobrą odpowiedź 1 dolara podnosiło poziom zaangażowania uczniów klas ósmych w rozwiązanie testu, jak również jego wynik. Związku takiego nie stwierdzono dla uczniów klas dwunastych. Po drugie, zdaniem Palomba i Banta (1999) zachętą do dobrowolnego udziału w teście oceny osiągnięć może być: (1) przekonanie uczniów, że ich udział jest ważny, bo pozwala na doskonalenie całego programu badania osiągnięć, (2) możliwość sprawdzenia swojej wiedzy dzięki otrzymaniu informacji zwrotnej o wynikach, (3) możliwość refleksji nad własnym poziomem osiągnięć i kompetencji, (4) otrzymanie drobnej nagrody w postaci gadżetu. Do listy tej, za Dolatą i Pokropkiem (2010, s. 88), dodać można (5) przekonanie ucznia o wysokiej wartości osobistej i społecznej udziału w teście.

Niniejszy artykuł ma na celu zaprezentowanie własności psychometrycznych skali motywacji testowej stworzonej na podstawie odpowiedzi udzielonych przez uczniów klas trzecich gimnazjum na stwierdzenia kwestionariusza ankietowego do samodzielnego wypełnienia. W artykule przedstawiono wyniki analiz eksploracyjnych i confirmacyjnych. Analizy eksploracyjne miały na celu zweryfikowanie, czy zastosowane narzędzie – stwierdzenia kwestionariusza – jest przydatne do stworzenia skali motywacji testowej. Natomiast analizy confirmacyjne posłużyły do zweryfikowania, czy stworzone narzędzie charakteryzuje się odpowiednią dokładnością pomiarową – trafnością i rzetelnością. Artykuł składa się z pięciu części. Najpierw przedstawiono pokrótce podstawowe wyniki badań w obszarze motywacji testowej, następnie źródło danych oraz zastosowane metody analityczne. W kolejnym punkcie zaprezentowano uzyskane wyniki analiz eksploracyjnych w zakresie (1) rzetelności i (2) wymiarowości skali. W kolejnym kroku pokazano wyniki analiz, mających na celu weryfikację modeli konceptualnych motywacji testowej. Artykuł kończy się podsumowaniem uzyskanych wyników.

## Dane i metody

W każdym dniu sesji zrównującej (marzec 2011 r., tj. po zakończeniu części humanistycznej oraz po zakończeniu części matematyczno-przyrodniczej) uczniom rozdane zostały kwestionariusze ankietowe do samodzielnego wypełnienia. Celem badania ankietowego było (1) zmierzenie poziomu motywacji testowej uczniów oraz stopnia przygotowania do sesji zrównującej, (2) uzyskanie informacji o percepcji poziomu trudności arkuszy zrównujących, (3) zweryfikowanie, czy czas przeznaczony na rozwiązanie zadań był w odczuciu uczniów wystarczający oraz (4) zebranie informacji o wcześniejszej znajomości przez uczniów zadań z arkuszy zrównujących. Zebranie tych wszystkich informacji konieczne było w celu zapewnienia większej dokładności zrównania wyników egzaminów.

Do zmierzenia poziomu motywacji testowej uczniów oraz stopnia przygotowania do sesji zrównującej przygotowanych zostało 14 stwierdzeń. Uczniowie proszeni byli o ustosunkowanie się do nich na 5-stopniowej skali Likerta. Do wyboru mieli jedną z pięciu następujących odpowiedzi: (1) zdecydowanie się nie zgadzam, (2) raczej się nie zgadzam, (3) ani się zgadzam, ani nie zgadzam, (4) raczej się zgadzam oraz (5) zdecydowanie się zgadzam. Stwierdzenia brzmiały następująco:

- S1. *Byłem(-am) dobrze przygotowany(-a) do tego testu.*
- S2. *Staralbym(-abym) się bardziej, gdyby ten test był egzaminem.*
- S3. *Uczyłem(-am) się kilka dni specjalnie do rozwiązywanego dzisiaj testu.*
- S4. *Jestem dobrze przygotowany(-a) do części matematyczno-przyrodniczej egzaminu gimnazjalnego<sup>3</sup>.*
- S5. *Część zadań była nudna, nie rozwiązywałem(-am) ich wcale.*
- S6. *Spotkałem(-am) się z zadaniami z tego testu lub z zadaniami bardzo podobnymi.*

<sup>3</sup> Kwestionariusz ankietowy wypełniany po zakończeniu rozwiązywania część humanistycznej różnił się wyłącznie treścią stwierdzenia S4.

- S7. *Przygotowując się do egzaminu, rozwiązywałem(-am) arkusze egzaminacyjne z poprzednich lat.*
- S8. *Dzisiaj nie czułem(-am) się na siłach, aby dobrze rozwiązać ten test.*
- S9. *Gdyby ten test był prawdziwym egzaminem, byłbym(-abym) lepiej przygotowany(-a).*
- S10. *Rozwiązywałem(-am) przede wszystkim pytania zamknięte, na resztę nie miałem(-am) ochoty.*
- S11. *Do egzaminu zostało jeszcze tyle czasu, że można się dużo nauczyć.*
- S12. *Do tej pory nie uczyłem(-am) się do egzaminu, dopiero zamierzam to zrobić.*
- S13. *Zadania z tego testu rozwiązałem(-am) najlepiej jak potrafiłem (-am).*
- S14. *Dzięki temu testowi sprawdzę swoje przygotowanie do egzaminu.*

Uczniowie trzecich klas gimnazjalnych ustosunkowywali się do tych stwierdzeń w każdym dniu sesji zrównującej (marzec 2011 r.) po zakończeniu rozwiązywania właściwego dla danego dnia sesji arkusza zrównującego (tj. po przeprowadzeniu części humanistycznej oraz po przeprowadzeniu części matematyczno-przyrodniczej). W ten sposób zebrane zostały dane odnoszące się do (1) motywacji testowej i stopnia przygotowania w części matematyczno-przyrodniczej oraz (2) motywacji testowej oraz stopnia przygotowania w części humanistycznej. W pierwszym przypadku kwestionariusz wypełniło 9223 z 9551 uczniów biorących udział w sesji zrównującej w części matematyczno-przyrodniczej, zaś w drugim – 9408 z 9593 uczniów rozwiązujących część humanistyczną egzaminu gimnazjalnego. Uczniowie pochodzili z 440 losowo wybranych szkół gimnazjalnych z całej Polski<sup>4</sup>.

Do stworzenia skali motywacji testowej zastosowano następujące analizy statystyczne:

#### I. Analizy eksploracyjne

- I.1. analiza głównych składowych z rotacją varimax – w celu rozpoznania możliwości utworzenia z zaproponowanego zestawu zmiennych mniejszego zbioru zmiennych, które dobrze wyjaśniają korelację między tymi zmiennymi i tym samym dobrze odtwarzają informację zawartą w zbiorze wyjściowym; sprawdzony został poziom zmienności wspólnej każdego ze stwierdzeń (min. 0,3) oraz poziom wyjaśnianej wariancji (min. 60%);
- I.2. analiza wartości współczynnika alfa-Cronbacha – w celu sprawdzenia, czy proponowany zestaw stwierdzeń tworzy skalę o satysfakcjonującej rzetelności (min. 0,6);

Punkty 1 i 2 prowadziły do redukcji liczby stwierdzeń.

<sup>4</sup> Szkoły losowano z wyłączeniem szkół specjalnych i przyszpitalnych oraz takich, w których oddziały klas trzecich liczyły więcej niż 10 uczniów; próba została powarstwowana według wielkości miejscowości oraz wyników egzaminacyjnych szkoły z roku 2010 (średnia egzaminu gimnazjalnego z części humanistycznej i matematyczno-przyrodniczej łącznie). Szkoły losowane były proporcjonalnie do liczby uczniów uczących się w klasach trzecich, a w każdej wylosowanej szkole do udziału w badaniu wylosowano w sposób prosty jedną klasę. Więcej na temat doboru próby do badań w Kondratek B., Kulon F., Pokropek A., Szaleniec H., Węziak-Białowska D., *Koncepcja metodologiczna badań umożliwiających długookresową porównywalność wyników egzaminów zewnętrznych*, Pracownia Analiz Osiągnięć Uczniów, IBE, Warszawa, marzec 2011.

## II. Analizy konfirmacyjne

### II.1. Weryfikacja modelu konceptualnego

Dla wyselekcjonowanych stwierdzeń (na podstawie wniosków płynących z punktów I.1 i I.2) przeprowadzono konfirmacyjną analizę czynnikową. Wspecyfikowano model konfirmacyjnej analizy czynnikowej dla motywacji testowej uczniów. Model ten został zweryfikowany m.in. pod względem istotności zmiennych wskaźnikowych (stwierdzeń) oraz jakości dopasowania. Ponadto za pomocą współczynnika rho-Joreskoga sprawdzono również rzetelność uzyskanego rozwiązania. Do estymacji zastosowano metody ograniczonej największej wiarygodności (ang. *restricted maximum likelihood*)<sup>5</sup> oraz WLSMV (ważona metoda największych kwadratów oparta na macierzach korelacji odpowiednio polichorycznych i tetrachorycznych)<sup>6</sup>.

### II.2. Ocena różnic w poziomie motywacji w części matematyczno-przyrodniczej i humanistycznej

W tym kroku oszacowano model motywacji testowej dla obu części egzaminu razem w celu sprawdzenia, czy między poziomami motywacji prezentowanej przez uczniów w różnych częściach egzaminu gimnazjalnego wystąpiły statystycznie istotne różnice. Zastosowano analizę konfirmacyjną dla wielu grup. Ponadto, celem zachowania porównywalności skal motywacji testowej, w każdej części egzaminu gimnazjalnego sprawdzono, czy zachodzą (van der Veld 2009, Davidov 2008):

- zgodność definicyjna (ang. *configural invariance*),
- zgodność skali lub zgodność metryki (ang. *metric invariance, factorial invariance*),
- zgodność skalarna lub zgodność punktu referencyjnego skali (ang. *scalar invariance*).

Do analiz zastosowano program Mplus version 6.

## Analizy eksploracyjne

### Analiza rzetelności

W pierwszym kroku przeprowadzono krytyczną analizę treści stwierdzeń. Usunięto te z nich, które nie są jednoznacznie sformułowane lub też nie dotyczą bezpośrednio kwestii motywacji testowej i przygotowania do sesji zrównującej. W konsekwencji analizy statystyczne przeprowadzono na stwierdzeniach: S1, S2, S3, S4, S5, S7, S8, S9, S13, S14 (w sumie 10 stwierdzeń). Odpowiedzi udzielone na te stwierdzenia tworzyły zmienne, które w analizie pełniły rolę zmiennych wskaźnikowych. Ponadto przed przystąpieniem do

<sup>5</sup> metoda największej wiarygodności odporna na niespełnienie założeń o normalności rozkładu zmiennych wskaźnikowych oraz niezależności i identyczności rozkładów zmiennych wskaźnikowych (MLR – maximum likelihood with standard errors and a chi-square test statistics that are robust to non-normality and non-independence of observations. The MLR standard errors are computed using a sandwich estimator).

<sup>6</sup> WLSMV – weighted least square parameter estimates using a diagonal weight matrix with standard errors and mean- and variance-adjusted chi-square test statistic that use a full weight matrix.

analiz przekodowano stwierdzenia S2, S5, S8 oraz S9 tak, aby ich orientacja znaczeniowa odpowiadała orientacji znaczeniowej pozostałych stwierdzeń. Przekodowanie odpowiadało odwróceniu punktacji.

Na przygotowanym w ten sposób zbiorze zmiennych wskaźnikowych przeprowadzono analizę rzetelności z wykorzystaniem współczynnika alfa-Cronbacha. Dla skali motywacji testowej i stopnia przygotowania do sesji zrównującej w części matematyczno-przyrodniczej współczynnik ten wyniósł 0,673, zaś dla skali motywacji testowej i stopnia przygotowania do sesji zrównującej w części humanistycznej była na poziomie 0,636. Co prawda przeciętna korelacja między zmiennymi wskaźnikowymi była niska (odpowiednio 0,178 dla części matematyczno-przyrodniczej oraz 0,160 dla części humanistycznej), to jednak w obu przypadkach współczynnik alfa-Cronbacha okazał się być statystycznie istotny ( $p = 0,000$ )<sup>7</sup>.

Ponadto analiza wrażliwości wartości tego współczynnika na usuwanie poszczególnych zmiennych wskaźnikowych z analizowanego zbioru zmiennych ujawniła, że usunięcie którejkolwiek z nich spowodowałoby obniżenie wartości współczynnika. Wniosek ten dotyczy obu skal motywacji testowej i stopnia przygotowania do sesji zrównującej.

### *Analiza wymiarowości skali*

W celu sprawdzenia, na ile dobrze zmienne wskaźnikowe kwantyfikują odpowiednie konstrukty (zmienne o charakterze latentnym odpowiadające skalom motywacji testowej i stopnia przygotowania do sesji zrównującej w każdej z części egzaminu gimnazjalnego), zastosowano analizę głównych składowych oraz eksploracyjną analizę czynnikową (metoda głównych osi wymiarów). W obu podejściach zastosowano rotację ortogonalną varmax.

W przypadku obu skal definiujących motywację testową i stopień przygotowania do sesji zrównującej w odpowiedniej części egzaminu gimnazjalnego adekwatność zbioru zmiennych wskaźnikowych do założeń analizy czynnikowej mierzona współczynnikiem K-M-O była satysfakcjonująca. Wskaźniki K-M-O wyniosły odpowiednio:

- dla części matematyczno-przyrodniczej – 0,676 dla całej baterii stwierdzeń oraz nie były niższe dla każdego ze stwierdzeń osobno od 0,547;
- dla części humanistycznej – 0,691 dla całej baterii stwierdzeń oraz nie były niższe dla każdego ze stwierdzeń osobno od 0,538.

Analiza zmienności zbioru zmiennych wskaźnikowych wykonana w toku analizy głównych składowych pokazała, że dla każdej z części egzaminu gimnazjalnego cztery wartości własne są większe od 1<sup>8</sup>, co oznaczało, że zbiór 10 zmiennych wskaźnikowych może być wiarygodnie zastąpiony czterema nowymi zmiennymi (głównymi składowymi). W przypadku części matematyczno-przyrodniczej te cztery zmienne razem odtwarzają 65,85% wariancji zbioru zmiennych wskaźnikowych, zaś w przypadku części humanistycznej – 62,22%.

<sup>7</sup> Do oceny istotności uzyskanego rozwiązania zastosowano chi-kwadrat Friedmanna.

<sup>8</sup> Kryterium wartości własnej.

Takie wyniki przemawiały za uznaniem, że struktura zbioru zmiennych wskaźnikowych jest czterowymiarowa.

Analiza zmienności wspólnej (przy przyjęciu założenia o czterowymiarowości zbioru zmiennych wskaźnikowych) pokazała, że każda ze zmiennych charakteryzuje się zmiennością wspólną z przedziału  $\langle 0,467; 0,759 \rangle$  dla części matematyczno-przyrodniczej oraz  $\langle 0,477; 0,713 \rangle$  dla części humanistycznej.

W następnym kroku przeanalizowano układ ładunków czynnikowych występujący przy założeniu występowania czterech zmiennych ukrytych. Okazało się, że w przypadku obu skal motywacji testowej i stopnia przygotowania do sesji zrównującej układ ładunków czynnikowych o najwyższych wartościach bezwzględnych<sup>9</sup>, odpowiadający układowi zmiennych wskaźnikowych najsilniej ładujących do odpowiednich zmiennych ukrytych, jest taki sam. Przeprowadzone analizy pokazały, że układ ten jest następujący:

- Z pierwszą główną składową najsilniej skorelowane były stwierdzenia: S5\_r. Część zadań była nudna, nie rozwiązywałem(-am) ich wcale<sup>10</sup>, S13. Zadania z tego testu rozwiązałem(-am) najlepiej jak potrafiłem(-am), S14. Dzięki temu testowi sprawdź swoje przygotowanie do egzaminu.
- Z drugą główną składową najsilniej skorelowane były stwierdzenia: S1. Byłem(-am) dobrze przygotowany(-a) do tego testu, S4. Jestem dobrze przygotowany(-a) do części matematyczno-przyrodniczej egzaminu gimnazjalnego, S8\_r. Dzisiaj nie czułem(-am) się na siłach, aby dobrze rozwiązać ten test<sup>11</sup>.
- Z trzecią główną składową najsilniej skorelowane były stwierdzenia: S2\_r. Starałbym(-abym) się bardziej, gdyby ten test był egzaminem, S9\_r. Gdyby ten test był prawdziwym egzaminem, byłbym(-abym) lepiej przygotowany(-a).
- Z czwartą główną składową najsilniej skorelowane były stwierdzenia: S3. Uczylem(-am) się kilka dni specjalnie do rozwiązywanego dzisiaj testu, S7. Przygotowując się do egzaminu, rozwiązywałem(-am) arkusze egzaminacyjne z poprzednich lat.

Tym samym można przyjąć, że odpowiednie główne składowe kwantyfikowały następujące wymiary motywacji testowej i stopnia przygotowania do sesji zrównującej: (I) pierwsza główna składowa – zaangażowanie w rozwiązanie testu zrównującego, (II) druga główna składowa – subiektywna ocena stopnia aktualnie posiadanej wiedzy niezbędnej do rozwiązania testu zrównującego i właściwego egzaminu, (III) trzecia główna składowa – subiektywna ocena stopnia przygotowania do właściwego egzaminu, (IV) czwarta główna składowa – ocena stopnia przygotowania się do rozwiązania testu zrównującego. Taki opis wymiarów motywacji testowej i stopnia przygotowania do sesji zrównującej sugeruje, że wprost do motywacji, jaką wykazywali się uczniowie, rozwiązując test zrównujący, odnosi się wymiar I. Natomiast wymiar III odpowiada

---

<sup>9</sup> Zawsze były to ładunki o znaku dodatnim.

<sup>10</sup> Odpowiedzi zostały zdekodowane na przeciwne.

<sup>11</sup> Odpowiedzi zostały zdekodowane na przeciwne.

za potencjalną motywację, którą uczniowie posiadaliby, gdyby test zrównujący był właściwym egzaminem (ang. *high-stake exam*), a więc jego wyniki były znaczące dla ich dalszej kariery szkolnej i zawodowej. Wymiary II i IV opisują proces przygotowania oraz jego efekty (posiadaną wiedzę).

Identyczny układ ładunków czynnikowych wystąpił w przypadku analizowania odpowiedzi dotyczących motywacji i stopnia przygotowania w części humanistycznej. Jedyna różnica dotyczyła stwierdzenia S8\_r, które co prawda ładowało najsilniej do drugiej głównej składowej, lecz miało niewiele niższy ładunek czynnikowy odpowiadający trzeciej głównej składowej.

Przeprowadzone analizy eksploracyjne pokazały, że zestaw 10 stwierdzeń proponowanych do pomiaru motywacji testowej i stopnia przygotowania do sesji zrównującej może być wykorzystany do budowy skali motywacji testowej. Na podstawie uzyskanych wyników do stworzenia skali motywacji testowej zakwalifikowano stwierdzenia:

S5\_r. *Część zadań była nudna, nie rozwiązywałem(-am) ich wcale*<sup>12</sup>,

S13. *Zadania z tego testu rozwiązałem(-am) najlepiej jak potrafiłem(-am)*,

S14. *Dzięki temu testowi sprawdzę swoje przygotowanie do egzaminu.*

W celu potwierdzenia powyższych wniosków oraz propozycji w kolejnym kroku przeprowadzono analizy konfirmacyjne. Ich celem było zweryfikowanie:

1. czy proponowany zestaw stwierdzeń (kwantyfikowanych przez odpowiednie zmienne wskaźnikowe) może posłużyć do budowy modelu pomiarowego motywacji testowej oraz stopnia przygotowania do sesji zrównującej,
2. czy model pomiarowy motywacji testowej oraz stopnia przygotowania do sesji zrównującej charakteryzuje się odpowiednią jakością (trafnością i rzetelnością),
3. czy proponowany zestaw stwierdzeń (S5\_r, S13 oraz S14) może posłużyć do budowy modelu pomiarowego motywacji testowej,
4. czy model pomiarowy motywacji testowej charakteryzuje się odpowiednią jakością (trafnością i rzetelnością),
5. czy model pomiarowy motywacji testowej charakteryzuje się stabilnością parametrów ze względu na grupę badawczą – weryfikacja zgodności pomiaru dla obu części egzaminu gimnazjalnego.

Ze względu na tematykę artykułu przedstawione zostaną wyłącznie wyniki analiz dotyczące punktów 3. – 5.

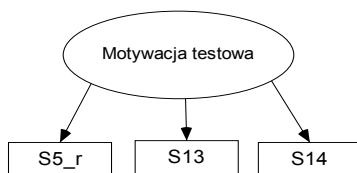
## **Analizy konfirmacyjne**

### *Weryfikacja modelu konceptualnego motywacji testowej*

Oceniając różne konfiguracje zmiennych wskaźnikowych pod względem rzetelności, adekwatności oraz jednowymiarowości budowanego na ich podstawie konstruktów, zaproponowano następujący model konceptualny motywacji testowej (rysunek 1):

<sup>12</sup> Odpowiedzi zostały zrekodowane na przeciwne.





**Rysunek 1. Model konceptualny motywacji testowej**

*Źródło: opracowanie własne*

Jako wskaźniki motywacji testowej wyselekcjonowano zmienne wskaźnikowe odpowiadające stwierdzeniom S5\_r, S13 oraz S14. Model pomiarowy, w którym występują tylko trzy zmienne wskaźnikowe, jest zawsze bardzo dobrze dopasowany do danych empirycznych (por. Brown 2006, Kaplan 2009). Z tego względu zaniechano oceny dopasowania tego modelu, zaś oszacowania jego współczynników przedstawiono wyłącznie dla modelu szacowanego dla obu części egzaminu gimnazjalnego łącznie.

Ocena różnic w poziomie motywacji w części matematyczno-przyrodniczej i humanistycznej

W celu sprawdzenia, czy między poziomami motywacji prezentowanej przez uczniów w różnych częściach egzaminu gimnazjalnego wystąpiły statystycznie istotne różnice, za pomocą analizy konfirmacyjnej dla wielu grup:

1. sprawdzono, czy model pomiarowy motywacji testowej charakteryzuje się własnością niezmienności pomiaru ze względu na badaną grupę uczniów,
2. skwantyfikowano różnice między poziomami motywacji testowej w części humanistycznej oraz matematyczno-przyrodniczej.

Potwierdzeniem istnienia zgodności definicyjnej było uzyskanie w analizach eksploracyjnych skali motywacji testowej oraz stopnia przygotowania do sesji zrównującej tego samego zestawu zmiennych wskaźnikowych stojących przy pierwszej głównej składowej. Składowa ta została zidentyfikowana jako odpowiadająca wprost motywacji testowej. Jak wspomniano powyżej, analizy konfirmacyjne modelu motywacji testowej nie zostały przeprowadzone ze względu na ograniczoną do trzech liczbę zmiennych wskaźnikowych dla tego konstruktów.

Oszacowany model motywacji testowej dla obu części egzaminu gimnazjalnego z pełną zgodnością pomiaru (tj. z narzuconymi warunkami równości na ładunki czynnikowe stojące przy odpowiednich zmiennych wskaźnikowych, jak i odpowiadające tym zmiennym wyrazy wolne między częścią humanistyczną i matematyczno-przyrodniczą) okazał się być bardzo dobrze dopasowany do danych. Wartości statystyk dobroci dopasowania<sup>13</sup> były na satysfakcjonującym poziomie przewyższających znacznie 0,95 (CFI = 0,989, TLI = 0,983), zaś statystyki braku dopasowania nie przekraczały granicznej wartości 0,08 (Browne,

<sup>13</sup> CFI – Comparative Fit Index (>0,9), TLI – Tucker-Lewis index (>0,9), RMSEA – Root Mean Square Error of Approximation (<0,08), SRMR – Standardized Root Mean Square Residual (<0,08).

Cudeck 1993) (RMSEA = 0,046<sup>14</sup>; SRMR = 0,021). Ponadto wszystkie zmienne wskaźnikowe były statystycznie istotne, zaś znaki stojących przy nich ładunków czynnikowych były dodatnie, co było zgodne z oczekiwaniami. Jednak rzetelność takiego modelu mierzona współczynnikiem Rho-Joreskoga była dosyć niska ( $\rho = 0,48$ ).

Niemniej jednak dobre dopasowanie modelu do danych empirycznych – mierzone statystykami dopasowania – świadczyło za istnieniem niezmienności pomiaru ze względu na badaną grupę uczniów. Oznaczało to, że szacowany model pomiarowy motywacji testowej ma takie same parametry w każdej z części egzaminu gimnazjalnego, zaś skala, na której mierzona jest motywacja testowa, w każdej z badanych grup uczniów ma tę samą jednostkę z tak samo na niej usytuowanym punktem zero. Tym samym na podstawie takiego modelu w pełni uprawnione metodycznie jest dokonywanie porównań przeciętnego poziomu motywacji testowej w analizowanych grupach.

Porównanie takie pokazało, że badani uczniowie pisząc część humanistyczną egzaminu gimnazjalnego w sesji zrównującej, byli bardziej zmotywowani niż w trakcie pisania części matematyczno-przyrodniczej. Różnica w przeciętnym poziomie motywacji wyrażona wielkością odchylenia standardowego była rzędu 0,101, zaś wyrażona w jednostkach skali motywacji testowej – 0,081. Ponadto różnica ta była statystycznie istotna.

Wynik taki należy jednak traktować z ostrożnością. Choć obie części egzaminu gimnazjalnego różnią się od siebie, to mógł wystąpić efekt obniżenia motywacji związany z eliminacją efektu nowości, który wystąpił w przypadku części matematyczno-przyrodniczej, która pisana była jako druga (drugiego dnia).

## Podsumowanie

Celem artykułu było przedstawienie wyników pierwszych analiz psychometrycznych mających na celu sprawdzenie, czy zestaw proponowanych stwierdzeń w założeniu mierzących poziom motywacji testowej i stopień przygotowania do sesji zrównującej może zostać zastosowany do pomiaru poziomu motywacji testowej. W tym celu przeprowadzono odpowiednie analizy o charakterze eksploracyjnym i confirmacyjnym. Do analizy wyselekcjonowano 10 z 14 stwierdzeń kwestionariusza ankietowego wypełnianego przez uczniów trzech klas gimnazjum po każdej z części sesji zrównującej.

Wyniki analizy rzetelności dla obu skal były satysfakcjonujące. Analiza wymiarowości skal pokazała, że można wyróżnić cztery niezależne wymiary skali motywacji testowej i stopnia przygotowania do sesji zrównującej. Są to: (I) zaangażowanie w rozwiązanie testu zrównującego, (II) subiektywna ocena stopnia aktualnie posiadanej wiedzy niezbędnej do rozwiązania testu zrównującego i właściwego egzaminu, (III) subiektywna ocena stopnia potencjalnego przygotowania do właściwego egzaminu oraz (IV) ocena stopnia przygotowania się do rozwiązania testu zrównującego.

Na podstawie tak wyróżnionych wymiarów zaproponowano model

<sup>14</sup> Ponadto krytyczny poziom istotności dla statystyki testującej dla hipotezy zerowej głoszącej, że RMSEA  $\leq 0,05$  osiągnął poziom 0,769.

konceptualny skali motywacji testowej. Taka konceptualizacji zyskała potwierdzenie w wynikach analiz empirycznych. Oszacowany model motywacji testowej dla obu części egzaminu gimnazjalnego z pełną zgodnością pomiaru (tj. z narzuconymi warunkami równości na ładunki czynnikowe stojące przy odpowiednich zmiennych wskaźnikowych, jak i odpowiadające tym zmiennym wyrazy wolne między częścią humanistyczną i matematyczno-przyrodniczą) okazał się być bardzo dobrze dopasowany do danych. Udało się również potwierdzić istnienie zgodności definicyjnej oraz pełnej zgodności skalarnej. Tym samym wykazano, że model pomiarowy motywacji testowej charakteryzuje się stabilnością parametrów ze względu na grupę badawczą.

Takie wyniki pozwoliły na przeprowadzanie porównania poziomów motywacji testowej w dwóch częściach egzaminu gimnazjalnego rozwiązywanego w sesji zrównującej. Okazało się, że badani uczniowie byli bardziej zmotywowani, pisząc część humanistyczną egzaminu gimnazjalnego w sesji zrównującej niż w trakcie pisania części matematyczno-przyrodniczej. Różnica ta była statystycznie istotna. Wynik taki należy jednak traktować z ostrożnością. Chociaż obie części egzaminu gimnazjalnego różnią się od siebie, to mógł wystąpić efekt obniżenia motywacji związany z eliminacją efektu nowości, który wystąpił w przypadku części matematyczno-przyrodniczej, która pisana była jako druga (drugiego dnia). Należy wyraźnie zaznaczyć, że wszystkie sformułowane powyżej wnioski wymagają jeszcze weryfikacji w dalszych badaniach.

### Bibliografia:

1. Baumert J., Demmrich A. (2001) *Test motivation in the assessment of student skills: The effects of incentives on motivation and performance*. *European Journal of Psychology of Education*. Vol. 16, 441-462.
2. Beguin A. A. (2000) *High-Stakes Tests*, FEBODRUK B.V., Enschede.
3. Brown T. A. (2006) *Confirmatory Factor Analysis for Applied Research*, The Guilford Press, New York, London.
4. Browne M. W., Cudeck, R. (1993) *Alternative ways of assessing model fit* [w:] Bollen K. A., Long, J. S. (Eds.), *Testing Structural Equation Models*, 136-162, Beverly Hills, Sage.
5. Center for Educational Testing and Evaluation (2001) *Student test taking motivation and performance: Grade 10 mathematics and science and grade 11 social studies*. University of Kansas, KS: School of Education.
6. Davidov E. (2008) *A Cross-Country and Cross-Time Comparison of the Human Values Measurements with the Second Round of the European Social Survey*, "Survey Research Methods", Vol.2, No.1, 33-46.
7. DeMars Ch. (2000) *Test stakes and item format interactions*. *Applied Measurement in Education*, 13, 55-77.
8. Dolata R., Pokropek A. (2010) *Motywacja a wynik testu z nauk przyrodniczych. Studium na przykładzie PISA 2006* [w:] Niemierko B., Szmigel M., *Teraźniejszość i przyszłość oceniania szkolnego*, TOMAMI, Toruń 2010, 86-97.

9. Eklof H. (2006) *Development and validation of scores from an instrument measuring student testtaking motivation*. *Educational and Psychological Measurement*, 66, 643-656.
10. Erwin, T. D., Wise, S. L. (2002) *A scholar-practitioner model for assessment*. In T. W. Banta (Ed.), *Building a scholarship of assessment*. San Francisco: Jossey-Bass, 67-81.
11. Kaplan D., 2009, *Structural equation modeling. Foundation and Extentions*, Sage, Los Angeles.
12. Kiplinger, V. L., & Linn, R. L. (1993) *Raising the stakes of test administration: The impact on student performance on NAEP* (CSE Technical Report 360). Los Angeles, CA: University of California, CRESST.
13. Kondrątek B., Kulon F., Pokropek A., Szaleniec H., Węziak-Białowska D., *Koncepcja metodologiczna badań umożliwiających długookresową porównywalność wyników egzaminów zewnętrznych*, Pracownia Analiz Osiągnięć Uczniów, IBE, Warszawa, marzec 2011.
14. Mislavy, R. J. (1995) *What can we learn from international assessments?* *Educational Evaluation and Policy Analysis*, 4, 419-437.
15. Napoli, A. R., Raymond, L. A. (2004) *How reliable are our assessment data? A comparison of the reliability of data produced in graded and ungraded conditions*. *Research in Higher Education*, 45, 921-929.
16. O'Neil, H. F., Jr., Sugrue, B., Abedi, J., Baker, E. L., Golan, S. (1997) *Final report of experimental studies on motivation and NAEP test performance* (CSE Technical Report 427). Los Angeles, CA: University of California, CRESST.
17. O'Neil, H. F., Sugrue, B., Baker, E. L. (1995/1996) *Effects of motivational interventions on the National Assessment of Educational Progress Mathematics performance*. *Educational Assessment*, 3, 135-157.
18. Palomba, C. A., & Banta, T. W. (1999) *Assessment essentials: Planning implementing and improving assessment in higher education*. San Francisco: Jossey-Bass.
19. Reykowski J., 1985, *Emocje i motywacja* [w:] T. Tomaszewski (red), *Psychologia*, PWN, Warszawa.
20. Sundre D., L. (1999) *Does examinee motivation moderate the relationship between 10. test consequences and test performance?* Paper presented at the annual meeting of the American Educational Research Association, Montreal, (ERIC Document Reproduction Service No. ED432588).
21. Wainer, H. (1993) *Measurement problems*. *Journal of Educational Measurement*, 30, 1-21.
22. Wolf, L. F., Smith, J. K. (1995) *The consequence of consequence: Motivation, anxiety, and test performance*. *Applied Measurement in Education*, 8, 227-242.