

# Paweł Nowak

---

## Rzetelność i trafność skali WISC-R w świetle wyników badań dzieci 11-letnich

---

Acta Universitatis Lodzianis. Folia Psychologica 2, 79-101

---

1998

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej [bazhum.muzhp.pl](http://bazhum.muzhp.pl), gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

PAWEŁ NOWAK  
Instytut Psychologii UŁ

## **Rzetelność i trafność skali WISC-R w świetle wyników badań dzieci 11-letnich**

### Wstęp

Skala WISC (Wechsler Intelligence Scale for Children) jest stosowana w Polsce od ponad 30 lat. Dokonywano wprawdzie oceny jej rzetelności i trafności, ale do dziś nie znormalizowano jej na reprezentatywnej próbie populacji.

Udostępnienie psychologom polskim w 1991 r. Zmodyfikowanej Wersji Skali do Pomiaru Inteligencji Dzieci – WISC-R (Wechsler Intelligence Scale for Children-Revised) poprzedziły kilkuletnie wysiłki psychologów z ośrodka warszawskiego zmierzające do adaptacji tej techniki do polskich warunków i jej normalizacji. Oszacowania rzetelności i trafności skali dokonano jednakże niezbyt wnikliwie – nie dano m. in. odpowiedzi na podstawowe pytanie praktyczne: czy badanie tą skalą pozwala na przewidywanie późniejszych sukcesów lub niepowodzeń w nauce?

Bacząc na to, że skala D. Wechslera jest jedną z podstawowych technik służących do diagnozy upośledzenia umysłowego oraz pamiętając o tym, że na podstawie wyników badań tą techniką podejmuje się wiele ważkich decyzji dotyczących systemu edukacji dziecka (skala masowa czy specjalna), każdy z psychologów korzystających ze skali WISC-R (lub WISC) powinien znać rzetelność i trafność tych technik, mieć świadomość ich ograniczeń i nie popełniać brzemiennej w skutki błędów diagnostycznych.

## Problem i cel badań

W toku normalizacji polskiej i adaptacji skali WISC-R tylko częściowo oszacowano jej rzetelność, natomiast trafność oceniono bardzo pobieżnie. Zastosowane pojedyncze kryterium zewnętrzne (Standardowe Matryce Ravena) niewiele mówi o trafności diagnostycznej. Nie wiadomo też, czy i w jakim stopniu można przewidzieć sukcesy w nauce szkolnej na podstawie wcześniejszego badania skalą WISC-R, ponieważ nie przeprowadzono w Polsce żadnych badań nad trafnością prognostyczną skali. Nie podjęto również badań porównawczych skal WISC oraz WISC-R i nie odpowiedziano na pytanie: czy dopuszczalne jest posługiwanie się starymi amerykańskimi normami do nadal stosowanej skali WISC, kiedy wyniki badań prowadzonych w 14 krajach świata wskazują na ich systematyczną dezaktualizację?

Ponadto autor pracy nie znalazł w piśmiennictwie żadnych informacji o badaniach nad rzetelnością konsensyjną obu skal Wechslera, a system oceny wyników uzyskiwanych w testach: Podobieństwa, Rozumienie i Słownik nie jest zupełnie jednoznaczny i może powodować pewne różnice w ocenie poszczególnych pytań, a w konsekwencji różne wyniki pomiaru za pomocą poszczególnych testów.

Podjęte w ramach wykonywanej pod kierunkiem prof. J. Kostrzewskiego pracy magisterskiej badania miały na celu kompleksowe porównanie właściwości psychometrycznych skal WISC i WISC-R, a w szczególności:

1. Ocenę rzetelności konsensyjnej trzech testów skali słownej obu technik (Podobieństwa, Rozumienie i Słownik).
2. Ocenę trafności czynnikowej obu skal.
3. Oszacowanie trafności diagnostycznej skali WISC-R.
4. Ustalenie trafności prognostycznej obu skal na podstawie:
  - a) wyników pomiaru za pomocą dwóch kryteriów zewnętrznych: testów Wiadomości i Umiejętności Szkolnych oraz Ocen Szkolnych;
  - b) równań regresji wielokrotnej.
5. Porównanie wyników w obu skalach, obliczając:
  - a) istotność różnic między rozkładami wyników w poszczególnych testach oraz trzech skalach obu technik;
  - b) istotność różnic między średnimi wynikami badań za pomocą poszczególnych testów oraz wszystkich skal obu technik.

Utylitarnym celem badań było opracowanie sposobu przewidywania wyników w skali WISC-R na podstawie wyników w skali WISC i na odwrót.

## Wyniki dotychczasowych badań nad rzetelnością i trafnością WISC i WISC-R

Ponieważ autor pracy poddawał ocenie głównie skalę WISC-R, wyniki badań nad własnościami psychometrycznymi skali WISC zostaną omówione w sposób bardzo skrótowy.

### Rzetelność i trafność WISC

Ze względu na brak miejsca autor pracy nie zestawiał wyników badań nad własnościami psychometrycznymi WISC. Zostały one zaprezentowane w innych pracach (m. in. Kostrzewski 1970).

### Rzetelność i trafność WISC-R

#### Rzetelność WISC-R

Autorowi pracy znane są wyniki badań nad stabilnością bezwzględną i jednolitością WISC-R.

### Stabilność bezwzględna WISC-R

W podręczniku do omawianej skali D. Wechsler podaje wyniki badań 303 dzieci z trzech poziomów wieku: 6,6–7,6; 10,6–11,6; 14,6–15,6 lat. Dzieci te poddano 2-krotnym badaniom skalą WISC-R, przerwa między pierwszym a drugim testowaniem wynosiła jeden miesiąc. Otrzymane przez Wechslera współczynniki korelacji są bardzo wysokie między wynikami w skali słownej (0,94–0,95) na wszystkich poziomach wieku oraz wysokie między wynikami w skali bezsłownej (0,89–0,90). Średnie współczynniki stabilności wyników poszczególnych testów wahają się od 0,65 (Labirynty) do 0,88 (Wiadomości) i wskazują na znaczną zbieżność (z wyjątkiem Labiryntów, w których zbieżność ta jest istotna).

W drugim badaniu wystąpił wyraźny wzrost średnich I.I. – o ok. 4 jednostki w skali słownej, ok. 10 – w skali bezsłownej i ok. 7 – w skali pełnej. Jednak jak podają Haynes i Howard (1986) oraz Naglieri i Pfeiffer

(1983), którzy przeprowadzili 2-krotne testowanie z przerwą 2-letnią, średnie ilorazy we wszystkich skalach zmieniły się mniej niż o 3 jednostki (Sattler 1988).

Autorki polskiej wersji skali WISC-R obliczyły współczynniki stabilności bezwzględnej jedynie dla testów Powtarzania Cyfr i Kodowania. Badając z miesięczną przerwą 300-osobową grupę dzieci z sześciu poziomów wieku, uzyskały współczynniki korelacji należące do rzędu wysokich (0,70 – Powtarzanie Cyfr i 0,74 – Kodowanie). Wartości otrzymanych współczynników stabilności są zbliżone do obliczonych w badaniach oryginalnych (Matczak, Piotrowska, Ciarkowska 1991).

### Jednolitość WISC-R

Wechsler oszacował równoważność międzypółkową (wykorzystując wzór Spearmana-Browna) wszystkich testów (z wyjątkiem testów Powtarzania Cyfr i Kodowania) oraz trzech skal (stosując wzór Mosiera) i podał współczynniki jednolitości dla 11 poziomów wieku, przy czym oparł się na wynikach badań próbki normalizacyjnej ( $N = 2200$ ). Uzyskane współczynniki korelacji są bardzo wysokie na wszystkich poziomach wieku w skalach słownej i pełnej (0,91–0,94 i 0,95–0,96) oraz bardzo wysokie w skali bezsłownej (0,89–0,91). Średnie współczynniki korelacji dla poszczególnych testów wynoszą: w skali słownej od 0,77 (Rozumienie i Arytmetyka) do 0,86 (Słownik), a w skali niewerbalnej od 0,70 (Układanki) do 0,85 (Wzory z Klocków).

Jednolitość skali oceniono również obliczając współczynniki korelacji między wynikami poszczególnych testów i całych skal oraz między wynikami uzyskanymi w trzech skalach. Wyniki testów skali słownej wyżej korelują ze sobą oraz z wynikami całej skali słownej niż wyniki testów w skali wykonawczej. Skala werbalna jest więc bardziej jednorodna niż skala wykonawcza. Na podstawie wyników w skali słownej możemy też w trafniejszy sposób przewidzieć wynik w skali pełnej. Cała skala werbalna koreluje z całą skalą niewerbalną w sposób umiarkowany (Wechsler 1974; Sattler 1988).

W badaniach standaryzacyjnych polskiej wersji WISC-R jednolitość całych skal oszacowano wykorzystując współczynnik Mosiera oraz dodatkowo współczynnik alfa-Cronbacha. Do określenia rzetelności poszczególnych testów zastosowano wzór Spearmana-Browna. Współczynniki jednolitości obliczono oddzielnie dla 11 poziomów wieku, opierając się na wynikach badań całej próby normalizacyjnej ( $N = 1670$ ).

Otrzymane współczynniki Mosiera dla wyników skali słownej i pełnej są bardzo wysokie dla wszystkich poziomów wieku, a dla wyników w skali

bezsłownej – wysokie i bardzo wysokie. Współczynniki te są niemal identyczne z uzyskanymi przez D. Wechslera.

Wartości współczynników alfa-Cronbacha są nieco niższe od wartości współczynników Mosiera i we wszystkich skalach dla wszystkich poziomów wieku należą do rzędu wysokich.

Spośród poszczególnych testów najbardziej rzetelne są pomiary za pomocą testów Słownika i Wzorów z Klocków, a najmniej rzetelne pomiary za pomocą testów Powtarzania Cyfr i Układanek (Maczak, Piotrowska, Ciarkowska 1991).

### **Standardowe błędy pomiaru**

Podane przez Wechslera średnie standardowe błędy pomiaru wahają się dla testów skali słownej od 1,15 punktu przeliczeniowego (Słownik) do 1,44 (Powtarzanie Cyfr); zaś dla testów skali wykonawczej od 1,77 P.P. (Wzory z Klocków) do 1,70 (Układanki i Labirynty).

Stosunkowo najmniejszym błędem pomiaru charakteryzuje się skala pełna (średni – 3,19). W przypadku skali słownej średni standardowy błąd pomiaru wyniósł 3,60, a w skali bezsłownej – 4,66 (Wechsler 1974).

W badaniach polskich standardowe błędy pomiaru (SEM) całych skal obliczono na podstawie współczynników alfa-Cronbacha (współczynnik Mosiera powoduje bowiem zawyżenie rzetelności). Autorki polskiej wersji oszacowały również wielkości standardowych błędów estymacji (SEE). Najmniejszym błędem pomiaru obarczona jest skala słowna, a największym – pełna. Dla wszystkich skal najniższe wartości SEM i SEE uzyskano dla dzieci 6-letnich (Maczak, Piotrowska, Ciarkowska 1991).

### **Trafność WISC-R**

Autorowi pracy znane są wyniki badań prowadzonych na świecie i w Polsce nad trafnością diagnostyczną i czynnikową WISC-R oraz analizy dotyczące trafności prognostycznej skali dokonane przez badaczy zagranicznych.

### **Trafność diagnostyczna**

Jako kryterium zewnętrzne dla oszacowania trafności diagnostycznej skali WISC-R stosowano najczęściej uznane za rzetelne i trafne inne narzędzia

pomiaru inteligencji, m. in. WPPSI (Wechsler Preschool and Primary Scale for Intelligence), WAIS (Wechsler Adult Intelligence Scale), WAIS-R, Stanford-Binet wersja IV oraz L-M., Test Inteligencji Slossona, Skalę Zdolności Umysłowej dla Dzieci McCarthy'ego, Test Zdolności Poznawczych Woodcocka-Johnsona, Obrazkowy Test Słownikowy Peabody'ego.

J. M. Sattler (1988) zebrał wyniki badań korelacyjnych między skalą WISC-R oraz innymi skalami inteligencji i obliczył średnie współczynniki korelacji. Na ich podstawie można wnosić, iż skala WISC-R jest trafnym narzędziem pomiaru inteligencji, zaś najpewniejszym jej wskaźnikiem jest wynik w skali pełnej (Wechsler 1974; Sattler 1988).

Trafność diagnostyczna polskiej wersji WISC-R oceniona została poprzez określenie korelacji między wynikami w tej skali a wynikami badań Standardowymi Matrycami Ravena. Poddając badaniom grupę 572 dzieci uzyskano na wszystkich poziomach wieku (z wyjątkiem dzieci 8-letnich – w przypadku skali słownej niski współczynnik korelacji) umiarkowane współczynniki korelacji. Zbieżność między wynikami poszczególnych testów jest wyraźna, lecz mała (z wyjątkiem wyników testu Wiadomości, które są zbieżne istotnie) (Maczak, Piotrowska, Ciarkowska 1991).

### **Trafność prognostyczna**

J. M. Sattler (1988) zebrał również wyniki badań nad trafnością prognostyczną WISC-R oraz obliczył przeciętne współczynniki korelacji. Dla skali słownej oraz testów Czytania wahają się one od 0,57 do 0,66 (w zależności od rodzaju testu Czytania); dla skali bezsłownej WISC-R z tymi testami od 0,34 do 0,47; zaś dla skali pełnej od 0,56 do 0,65.

Między testami ortografii a skalą słowną i pełną zachodzi zbieżność istotna (współczynniki korelacji od 0,50 do 0,59), a między tymi testami a skalą wykonawczą zbieżność wyraźna, lecz mała.

Wyniki badań otrzymane za pomocą testów arytmetyki korelują z wynikami skali słownej, wykonawczej i pełnej WISC-R w stopniu umiarkowanym.

### **Trafność czynnikowa**

A. S. Kaufman i A. B. Silverstein na podstawie analiz przeprowadzonych na amerykańskiej próbie normalizacyjnej wyodrębnili trzy czynniki: rozumowanie słowne, organizację percepcyjną, pamięć i odporność na dystraktory.

Autorki polskiej wersji skali, wykorzystując metodę głównych składowych z rotacją Varimax, wyodrębniły trzy czynniki całkowicie zgodne z czynnikami wyodrębnionymi przez Kaufmana i Silversteina. Czynniki pierwszy reprezentowany jest przez testy: Wiadomości, Podobieństw, Słownika i Rozumienia. Najbardziej nasycone drugim czynnikiem są testy: Uzupełniania Obrazków, Wzorów z Klocków, Układek i Labiryntów. Czynniki trzeci obejmował testy: Arytmetyki, Powtarzania Cyfr i Kodowania. Wyróżnione czynniki wyjaśniają ogółem 59,96% wariacji, przy czym czynnik pierwszy wyjaśnia 26,49%, drugi 20,05%, a trzeci 13,51% (Maczak, Piotrowska, Ciarkowska 1991).

### **Badania nad systematycznym wzrostem ilorazów inteligencji we współczesnych społeczeństwach**

J. R. Flynn (1987) zebrał wyniki badań prowadzonych w 14 krajach nad stopniowym wzrostem ilorazów inteligencji. Do oceny inteligencji stosowano różne testy, badano dzieci i dorosłych, różny też był okres przerwy między badaniami, na podstawie których wnioskowano o wzroście lub spadku ilorazów inteligencji lub innych jej wskaźników. Ze wszystkich tych badań (poza jednym) można wysnuć oczywisty wniosek – z biegiem lat we współczesnych, rozwiniętych cywilizacyjnie społeczeństwach następuje systematyczny wzrost ilorazów inteligencji lub innych jej wskaźników.

Tabela 1

Otrzymany lub przewidywany wzrost ilorazów inteligencji w ciągu 30 lat w pięciu wybranych krajach (na podstawie wyników badań uzyskanych za pomocą skal Wechslera)

Kraj	Przyrost I.I. w skalach		
	słownej	wykonawczej	pełnej
USA	9	8	10
Japonia	25	20	22
Austria	25	20	28
RFN	22	12	28
Francja	11	3	19

Źródło: J. R. Flynn (1987), s. 186.



## Hipotezy, metody, przebieg i teren badań

### Hipotezy badawcze

Sformułowano następujące hipotezy badawcze:

1. Rzetelność konsensusyjna testów: Podobieństw, Rozumienia i Słownika w skalach WISC-R i WISC jest wysoka.
2. Skale WISC i WISC-R mierzą trzy podstawowe czynniki: werbalny, wykonawczy oraz koncentracji uwagi i pamięci.
3. Skala WISC-R jest trafnym narzędziem pomiaru inteligencji.
4. Wyniki w skalach WISC-R i WISC umożliwiają przewidywanie osiągnięć w nauce szkolnej w zakresie: jęz. polskiego, matematyki i innych przedmiotów.
5. Wyniki otrzymane w skali WISC-R są niższe niż wyniki pomiaru skalą WISC.

### Metody i sposób badań

Podstawowe metody badań stanowiły skale do badania inteligencji: Wechsler Intelligence Scale for Children (WISC) oraz Wechsler Intelligence Scale for Children-Revised (WISC-R). Metodami dopełniającymi były testy Wiadomości i Umiejętności Szkolnych, a metodą uzupełniającą – wywiad.

### Skale WISC i WISC-R

Skale Wechslera do badania rozwoju intelektualnego oparte są na teorii zakładającej, iż inteligencja jest globalną zdolnością do racjonalnego myślenia, celowego działania i skutecznego radzenia sobie w swoim środowisku. Składa się na nią wiele specyficznych zdolności, które najogólniej podzielić można na zdolności werbalne i wykonawcze.

Z przyjętej teorii wynika konstrukcja omawianych skal Wechslera. Składają się one z 10 testów podstawowych i dwóch zastępczych, z których połowa tworzy skalę słowną, a połowa bezsłowną. W skład skal WISC i WISC-R wchodzi następujące testy: Wiadomości, Podobieństwa, Arytmetyka, Rozumienie, Słownik, Powtarzanie Cyfr (testy skali słownej) oraz Braki na Obrazkach, Porządkowanie Obrazków, Klocki, Układanki, Symbole Figur i Cyfr (Kodowanie) i Labirynty (testy skali wykonawczej).

Skala WISC służy do badania dzieci w wieku 5,0–15,11 lat, zaś WISC-R przeznaczona jest dla dzieci w wieku 6,0–16,11 lat. Podstawowym wskaźnikiem

rozwoju intelektualnego badanego dziecka jest obliczany z wyników 10 testów iloraz inteligencji w skali pełnej; jako miary uzupełniające stosuje się ilorazy inteligencji w skalach: werbalnej i niewerbalnej. Ze względu na objętość pracy bardziej szczegółowa charakterystyka skal WISC i WISC-R nie jest możliwa; nie wydaje się ona również konieczna, ponieważ omawiane skale są powszechnie znanymi narzędziami.

### Charakterystyka testów Wiadomości i Umiejętności Szkolnych

Osiągnięcia szkolne autor pracy badał baterią testów Wiadomości i Umiejętności Szkolnych, którą tworzyły testy: Czytanie i Ortografia M. Grzywak-Kaczyńskiej, Sprawdzian z Matematyki z Zakresu Klasy VI oraz Matematyka kl. V. Jako kryterium dodatkowe uwzględniono oceny szkolne z sześciu przedmiotów.

#### Test Czytanie

Test przeznaczony jest do badania cichego czytania ze zrozumieniem. Można go wykorzystywać do indywidualnego bądź zbiorowego badania uczniów klas I–VII. Zlicza się i interpretuje poprawnie wykonane zadania w ciągu 8 min. Uwzględnia się też liczbę błędnie wykonanych zadań, które wskazują na sposób czytania danego dziecka. Test cechuje umiarkowana stałość wyników 2-krotnego badania ( $r = 0,84$ ).

Zastosowano obrazkową, krótszą wersję testu z ograniczeniem czasu rozwiązywania do 5 min. Określono jego rzetelność na podstawie wyników dwóch badań w odstępie miesiąca, uzyskując umiarkowany współczynnik ( $r = 0,87$ ), który nie różni się od współczynnika otrzymanego przez M. Sobolewską (1988). Oszacowano trafność wewnętrzną testu oraz trafność diagnostyczną. Uzyskane współczynniki  $r$  wskazują na umiarkowaną korelację między wynikami testu Czytanie i oceną szkolną (jęz. polski) i niską korelację między wskaźnikami testu Kaczyńskiej i wynikami testu Ravena.

Test Czytanie znormalizowano na grupie 1595 dzieci (Sobolewska 1988).

#### Test Ortografia M. Grzywak-Kaczyńskiej

Test przeznaczony jest do badania indywidualnego bądź zbiorowego uczniów wszystkich klas szkoły podstawowej. Czas rozwiązywania testu wynosi 8 min. Punktuje się poprawnie wykonane zadania. Badania autorki testu

nie pozwalają rozstrzygnąć, czy test jest rzetelnym i trafnym narzędziem pomiaru umiejętności poprawnego ortograficznie pisania, dlatego też autor ocenił częściowo rzetelność i trafność omawianego narzędzia. Wyliczony współczynnik stabilności bezwzględnej wyniósł 0,88 (2-krotne badanie 59 dzieci z przerwą miesięczną). Trafność wewnętrzną testu oszacowało czterech sędziów (nauczycielki jęz. polskiego), oceniając jego przydatność, jako sprawdzianu umiejętności poprawnego ortograficznie pisania w stopniach: „bardzo dobry”, „dobry”, „słaby”, „bardzo słaby”. Trafność diagnostyczną testu oceniono korelując uzyskane w nim wyniki z ocenami z języków polskiego i angielskiego. Uwzględniono dwa wskaźniki: liczbę poprawnie wykonanych zadań w teście oraz liczbę błędów i opuszczeń. Dla obu wskaźników uzyskano umiarkowane współczynniki korelacji z ocenami z języków polskiego i angielskiego.

### **Sprawdzian z Matematyki z Zakresu Klasy VI**

Zadania testu dotyczą podstawowych treści programu nauczania matematyki w klasie szóstej. Test może być wykorzystywany do badań indywidualnych i zbiorowych bez limitu czasu. Ponieważ dotąd nie oszacowano jego rzetelności i trafności, autor ocenił zgodność wewnętrzną testu za pomocą wzoru Kudera-Richardsona, uzyskując współczynnik korelacji  $r = 0,81$  oraz określił trafność diagnostyczną (pięciu nauczycieli matematyki w klasach szóstych oceniło test jako całkowicie zgodny z aktualnym programem nauczania), a uzyskany współczynnik korelacji wyników testu i ocen z matematyki okazał się wysoki (0,83).

### **Test Matematyka kl. V**

Jego autorami są B. Chrzanowska, M. Kicińska, B. Olecka i H. Radziwiłł. Określili oni zgodność wewnętrzną testu (wykorzystując wzór Kudera-Richardsona uzyskali współczynnik  $r = 0,88$ ) oraz trafność wewnętrzną i diagnostyczną. (Współczynniki korelacji wyników testu oraz ocen szkolnych z matematyki wahały się od 0,21–0,73).

### **Etapy i przebieg badań**

W pierwszym etapie połowę grupy przebadano skalą WISC, a połowę WISC-R. W etapie drugim zamieniono skale. Przerwa między badaniami obu skalami wynosiła dla poszczególnych dzieci od 30 do 44 dni.

W trzecim etapie dokonano zbiorowego pomiaru wiadomości i umiejętności szkolnych, w grupach liczących do 12 osób. Podawano jednorazowo wszystkie testy, zawsze w tej samej kolejności: Czytanie, Ortografia i odpowiedni test z matematyki. Czwarty etap obejmował powtórne badanie grupy 59 dzieci testami Czytanie i Ortografia. W ostatnim etapie czterech niezależnych sędziów dokonało oceny odpowiedzi poszczególnych dzieci na pytania testowe w Rozumieniu, Podobieństwach i Słowniku w obu skalach: WISC i WISC-R.

Ostatecznie spośród 70 osób 2-krotnie przebadano 68. Wyniki sześciorga dzieci uznano jednak za mało wiarygodne. Obliczenia wykonano opierając się na próbce 62 dzieci (oprócz obliczeń współczynników rzetelności testów Czytanie i Ortografia, które opierają się na grupie 59 dzieci).

Badania przeprowadzone były przez autora w dwóch szkołach podstawowych.

### Zasady doboru próbki i charakterystyka badanych osób

Aby wyniki badań były wiarygodne, dobrano taką próbę osób badanych, która przypominała strukturę próbki normalizacyjnej pod względem płci uczniów, wykształcenia rodziców oraz miejsca zamieszkania. Ze względu na ograniczone możliwości techniczne skoncentrowano się na przedziale wieku 11,0–11,11 lat. Zrezygnowano z kontroli zmiennej „miejsce zamieszkania” (ze względu na nieistotne różnice między wsią i małym miastem oraz małym miastem i dużym w badaniach normalizacyjnych polskiej wersji skali WISC-R).

Badaniami objęto uczniów zamieszkałych w małym mieście w centrum Polski; wyłączono dzieci upośledzone umysłowo oraz ze stwierdzonymi silnymi zaburzeniami emocjonalnymi, a także dzieci z rodzin o bardzo trudnych warunkach bytowych.

Średni wiek dzieci pierwszy raz badanych skalami Wechslera wynosił 11,5; średni wiek uczniów drugi raz badanych – 11,6.

Tabela 2

Struktura badanej próbki pod względem płci i wykształcenia rodziców

Wykształcenie rodziców	Podstawowe lub niepełne podstawowe	Zasadnicze zawodowe	Średnie lub niepełne wyższe	Wyższe	Ogółem
Chłopcy	4	12	12	7	35
Dziewczęta	4	12	12	7	35
Razem	8	24	24	14	70

## Wyniki własnych badań

### Wyniki badań uzyskane przy zastosowaniu skal WISC i WISC-R

Średnie odchylenia standardowe dla poszczególnych testów oraz skal WISC i WISC-R, wartości testu t oraz poziomy istotności uzyskanych różnic zawiera tab. 3.

Tabela 3

Średnie arytmetyczne  $\bar{x}$ , odchylenia standardowe  $\sigma$  z wyników uzyskanych w skalach WISC i WISC-R, wartości testu t oraz poziomy istotności różnic między średnimi

Porównywane testy i skale	WISC		WISC-R		t	p
	$\bar{x}_1$	$\sigma_1$	$\bar{x}_2$	$\sigma_2$		
Wiadomości	9,71	2,96	10,27	3,11	-2,94	0,005
Rozumienie	10,37	2,38	11,73	1,90	-4,87	0,0001
Arytmetyka	10,60	1,95	9,90	3,09	2,20	0,031
Podobieństwa	12,71	2,63	10,27	2,58	8,54	0,0001
Słownik	15,15	3,21	11,32	2,43	13,97	0,0001
Powtarzanie Cyfr	8,15	1,98	9,53	2,53	-5,06	0,0001
Uzupełnianie Obrazków	11,18	3,19	10,76	2,70	1,22	ni
Porządkowanie Obrazków	9,63	2,38	11,42	3,01	-4,39	0,0001
Wzory z Klocków	12,69	2,71	10,92	2,72	6,61	0,0001
Układanki	10,85	3,02	9,06	3,08	4,30	0,0001
Kodowanie	12,68	3,27	11,85	2,58	2,38	0,021
Labirynty	10,89	2,71	11,14	2,85	-1,25	ni
Skala słowna	110,65	13,32	104,44	13,63	7,63	0,0001
Skala bezsłowna	109,60	15,25	105,43	13,95	2,66	0,01
Skala pełna	111,18	13,88	105,48	13,60	5,79	0,0001

Jak wynika z danych zawartych w tab. 3, we wszystkich skalach oraz w sześciu testach skali WISC-R (Arytmetyka, Podobieństwa, Słownik, Wzory z Klocków, Układanki i Kodowanie) uzyskano niższe wyniki w porównaniu z wynikami skali WISC. Większe różnice między wynikami skal słownych WISC i WISC-R (niż skal wykonawczych porównywanych technik) spowodowane są zapewne tym, że to właśnie skala słowna WISC-R została w głównej mierze zmodyfikowana. Znaczącą rolę w obniżeniu się średniego wyniku w skali słownej WISC-R, w porównaniu ze średnimi wynikami uzyskanymi w skali WISC, odegrała zapewne polska normalizacja zmodyfikowanej WISC-R. WISC stosowana przez polskich psychologów ma amerykańskie normy z 1949 r., a jak wynika z badań, których rezultaty zestawiał J. R. Flynn (1987), stare normy szybko się dezaktualizują.

Na niższe średnie wyniki w skali bezsłownej WISC-R wydają się wpływać: wspomniana już polska normalizacja oraz zmiana systemu punktowania poszczególnych testów, wprowadzona przez autorki polskiej wersji skali.

Niewątpliwie w Zmodyfikowanej Wersji Skali Wechslera dla Dzieci pozytywnie należy ocenić uwspółcześnienie pytań w Wiadomościach, Słowniku oraz Arytmetyce, a także wprowadzenie przez Wechslera w skali WISC-R nowego systemu oceny testu Powtarzanie Cyfr. Spowodowało to przybliżenie osiąganego wyniku średniego do średniej teoretycznej. Podobne konsekwencje miały modyfikacje zastosowane w Podobieństwach, przy czym sposób punktowania podany przez autorki polskiej wersji skali nie jest zupełnie zgodny ze sposobem oceny tego testu w używanej w Polsce skali WISC. Doprowadziło to prawdopodobnie do tego, że test Podobieństw mierzy w obu skalach nieco inne zdolności (tylko w 39,7% na wynik w tym teście w skali WISC-R mają wpływ te same przyczyny, które brane są pod uwagę w teście Podobieństw w skali WISC).

Podobnie rzecz się ma przy testach skali bezsłownej – Porządkowania Obrazków i Układanek. Modyfikacje sposobów punktowania w obu testach oraz zwiększenie liczby łatwiejszych zadań w teście Porządkowanie Obrazków spowodowało niewielki udział wspólnych czynników w wyjaśnianiu wariancji tychże testów w skalach WISC i WISC-R.

Natomiast zmiany w testach Klocków oraz Kodowania doprowadziły do zbliżenia otrzymanych wyników do teoretycznej średniej.

#### **Przewidywanie wyników w skali WISC na podstawie wyników w skali WISC-R (i odwrotnie)**

Znajomość sposobu przeliczania wyników uzyskanych w nieznormalizowanej – a zdarza się, że nadal stosowanej skali WISC – na wyniki w znormalizowanej na reprezentatywnej próbie polskiej populacji skali WISC-R, umożliwia w pełni odpowiedzialne interpretowanie otrzymanych wyników w skali WISC. Dlatego autor obliczył równania regresji:

- 1) w skali słownej:  $Y = 0,867X + 20,01$ ,  $X = 0,907Y + 4,08$ ;
- 2) w skali wykonawczej:  $Y = 0,709X + 34,877$ ,  $X = 0,593Y + 40,497$ ;
- 3) w skali pełnej:  $Y = 0,858X + 20,659$ ,  $X = 0,824Y + 3,85$ ;

gdzie: Y – wynik oczekiwany w WISC, X – wynik oczekiwany w WISC-R.

Tabela 4

Współczynniki korelacji  $r_{xy}$ , współczynniki alienacji  $k_{xy}$  i wskaźniki efektywności przewidywania E dla wyników trzech skal porównawczych technik

Współczynnik	Skale		
	słowna WISC i WISC-R	wykonawcza WISC i WISC-R	pełna WISC i WISC-R
$r_{xy}$	0,887	0,648	0,841
$k_{xy}$	0,461	0,762	0,541
E	54,9%	23,8%	45,9%

Jak wynika z danych zawartych w tab. 4, skuteczność przewidywań wyników w skali WISC na podstawie wyników w skali werbalnej WISC-R i odwrotnie wynosi 54,9%. Przewidywanie wyników jednej skali na podstawie rezultatów w drugiej jest obarczone większym błędem w skali pełnej; najwięcej błędów można popełnić prognozując wyniki w skali bezsłownej WISC z wyników w tej skali WISC-R (lub odwrotnie).

#### Weryfikacja reguł kończenia poszczególnych testów w skali WISC-R

Skala WISC-R posiada sprecyzowane reguły rozpoczynania i kończenia poszczególnych testów, co pozwala na oszczędność czasu podczas badania. Autor sprawdził, na ile podane przez autorki polskiej wersji WISC-R zasady zakończenia poszczególnych testów znajdują potwierdzenie w praktyce. Stosował więc każdorazowo wszystkie zadania testowe obu skal. Czterej oceniający sędziowie stwierdzili, że w teście Wiadomości aż 11 osób (17,6%) przekroczyło limit 5 kolejnych błędnych odpowiedzi, po których nie można się już spodziewać odpowiedzi poprawnej. W Słowniku zdarzyły się 3 odpowiedzi ocenione pozytywnie po 8 kolejnych błędach. Ogółem w Słowniku aż 13 osób (21% wszystkich badanych) przekroczyło limit błędów przyjęty przez autorki polskiej wersji skali, po którym A. Matczak, A. Piotrowska i W. Ciarkowska zalecają przerwanie badania testem. W teście Rozumienie dwie osoby odpowiedziały na pytanie testowe po wcześniejszych czterech kolejnych błędach. W pozostałych testach skali WISC-R zadania są ułożone we właściwej kolejności, a reguły przerywania tych testów w pełni sprawdzają się.

### Wyniki własnych badań nad rzetelnością konsensusyjną skal WISC i WISC-R

Estymacji konsensusyjności dokonano po to, aby określić, w jakim stopniu system oceny odpowiedzi w poszczególnych testach obu skal pozwala na jednakowe punktowanie tychże odpowiedzi przez osoby oceniające. Jest to bardzo ważny problem dla praktyki psychologicznej, decyduje również o obiektywności testu.

Niezależnej oceny, zgodnie z kryteriami podanymi w instrukcjach do obu skal, dokonali 4 kompetentni sędziowie. Następnie autor obliczył wartości współczynnika zgodności sędziów W-Kendalla dla 4 sędziów łącznie oraz współczynnika korelacji rang  $r_s$ -Spearmana dla każdej pary sędziów. Wreszcie, aby porównać obie skale pod względem rozkładów ocen w poszczególnych testach obliczono wartość statystyki chi-kwadrat.

Tabela 5

Wyniki badań nad rzetelnością konsensusyjną trzech testów WISC i WISC-R przy zastosowaniu współczynnika W-Kendalla

Testy	Wartości		Procent wariacji tłumaczonej W	Wartość chi-kwadrat	P
	W	$r_s$			
<b>WISC</b>					
Rozumienie	0,866	0,821	67,5	211,3	0,0001
Podobieństwa	0,975	0,967	93,4	237,9	0,0001
Słownik	0,980	0,973	94,7	239,1	0,0001
<b>WISC-R</b>					
Rozumienie	0,983	0,977	95,5	239,8	0,0001
Podobieństwa	0,992	0,989	97,9	242,0	0,0001
Słownik	0,991	0,988	97,6	241,8	0,0001

Dla testów skali WISC wartości współczynnika W wskazują na wysoką i bardzo wysoką zgodność sędziów, co do kryteriów, którymi się posługiwali przy ocenianiu.

W przypadku skali WISC-R otrzymane wartości współczynnika W są bardzo wysokie dla wszystkich ocenianych testów. Uzyskane współczynniki W tłumaczą od 95,5% (Rozumienie) do 97,9% (Podobieństwa) ogólnej wariacji. Resztę, czyli od 2,1% (Podobieństwa) do 4,5% (Rozumienie), wariacji wyjaśniają różnice indywidualne w sposobie oceny badanych osób przez sędziów.



W każdym ze sprawdzanych testów skali WISC-R uzyskano więcej całkowicie zgodnych lub minimalnie tylko niezgodnych ocen niż w odpowiednich testach skali WISC.

Podobne prawidłowości, na które zwrócono uwagę wyżej, występują również, gdy za kryterium porównawcze trzech testów obu skal przyjmiemy obliczone współczynniki  $r_s$ -Spearmana.

Ponieważ na podstawie przedstawionych wyników można było sformułować hipotezę, że testy skali WISC-R charakteryzuje wyższa konsensyjność niż testy skali WISC, autor postanowił zweryfikować to przypuszczenie stosując test chi-kwadrat. Okazało się, że między rozkładami wyników ocen we wszystkich trzech testach występują bardzo istotne statystycznie różnice.

Wydaje się, iż o wyższej konsensyjności skali WISC-R decyduje lepiej sprecyzowany system ocen w omawianych testach skali albo też trafniej dobrane przykłady odpowiedzi na pytania testowe, zamieszczone w instrukcji do skali WISC-R.

Cała skala WISC-R, jak i tworzące ją testy wykazują wysoce zadowalającą konsensyjność, o czym świadczą bardzo wysokie współczynniki zgodności, stosunkowo dużo ocen jednomyślnych lub minimalnie różniących się. Ujawnione różnice (od 2 do 3 punktów przeliczonych) mają niewielkie znaczenie dla praktyki. Natomiast testy skali WISC, mimo długoletniego stosowania skali w praktyce psychologicznej i istnienia obszernej instrukcji, która dość szczegółowo opisuje sposoby oceny odpowiedzi, obarczone są większymi błędami pomiaru. Dotyczy to w szczególności testu Rozumienie.

### **Trafność skal WISC i WISC-R w świetle wyników własnych badań**

#### **Trafność diagnostyczna skali WISC-R**

Jako kryterium zewnętrzne dla skali WISC-R można było wykorzystać skalę WISC. Z dotychczasowych badań wynika, że skale te w dużej części mierzą te same zdolności, a ich konstrukcja oparta jest na tej samej teorii inteligencji (Wechsler 1974; Sattler 1988). Skorelowano więc wyniki poszczególnych testów oraz skal WISC i WISC-R. Uzyskane współczynniki zawiera tab. 6.

Testy skal słownych i całe skale werbalne WISC-R i WISC wyżej korelują ze sobą niż testy skal bezsłownych i całe skale wykonawcze.

Tabela 6

Współczynniki korelacji między wynikami badań WISC i WISC-R oraz poziomy istotności i procenty tłumaczonej przez nie wariancji wspólnej

Testy i skale	Wartość r	Poziom istotności P	Procent wspólnej wariancji
Wiadomości	0,877	0,0001	76,9
Rozumienie	0,497	0,0001	24,7
Arytmetyka	0,596	0,0001	35,5
Podobieństwa	0,630	0,0001	39,7
Słownik	0,741	0,0001	51,0
Powtarzanie Cyfr	0,564	0,0001	31,8
Uzupełnianie Obrazków	0,586	0,0001	34,3
Porządkowanie Obrazków	0,310	0,007	9,6
Wzory z Klocków	0,697	0,0001	48,6
Układanki	0,424	0,0001	18,0
Kodowanie	0,589	0,0001	34,7
Labirynty	0,312	0,007	9,7
Skala słowna	0,887	0,0001	78,7
Skala bezsłowna	0,648	0,0001	42,0
Skala pełna	0,841	0,0001	70,7

Trzeba jednak podkreślić, że zadowalające korelacje uzyskano tylko dla testów Wiadomości, Słownika i Wzorów z Klocków oraz dla całych skal werbalnych i pełnych. Pozostałe współczynniki korelacji, głównie testów: Porządkowania Obrazków, Labiryntów, Układanek i Rozumienia są zdecydowanie za niskie – jak dla technik opartych na tych samych założeniach teoretycznych, o identycznej konstrukcji i mających w swym założeniu mierzyć podobne zdolności.

Spowodowane to jest prawdopodobnie zmianami w materiale testowym oraz zmianami w systemie punktowania ww. testów wprowadzonymi w polskiej wersji skali. Najbardziej kontrowersyjny wydaje się system oceny odpowiedzi w Układankach. Mimo wprowadzonych poprawek, nadal ma on niewiele wspólnego z oryginalnym systemem twórcy skali, gdyż za rozwiązania najłatwiejsze Wechsler proponuje mniejszą możliwą do wykazania liczbę punktów niż za rozwiązania najtrudniejsze. W polskim systemie ocen jest dokładnie odwrotnie.

### Trafność prognostyczna WISC i WISC-R

W celu wykazania stopnia zależności między wynikami w WISC i WISC-R a wynikami w testach osiągnięć szkolnych oraz ocenami szkolnymi policzono współczynnik korelacji wielokrotnej, zaś aby określić, które ze zmiennych

użytych jako kryteria zewnętrzne omawianych skal można w największym stopniu przewidzieć na podstawie wyników w obu skalach, obliczono równania regresji wielokrotnej.

Aneks zawiera tabele, w których zestawiono wartości współczynników korelacji skal WISC i WISC-R z testami: osiągnięć szkolnych w zakresie jęz. polskiego i z ocenami z tego przedmiotu; wiadomości i umiejętności matematycznych i ocenami szkolnymi z matematyki oraz z ocenami szkolnymi z innych przedmiotów; a także wartości współczynników korelacji wielokrotnej skal WISC i WISC-R z testami osiągnięć szkolnych i ocenami szkolnymi.

Analizując wyniki zestawione w tabelach Aneksu, można wysnuć wniosek, że porównawcze skale WISC i WISC-R są trafnymi narzędziami pomiaru inteligencji – na podstawie wyników badań tymi skalami można z dużym prawdopodobieństwem przewidzieć (prognozować) sukcesy w zakresie jęz. polskiego, matematyki i innych przedmiotów. Do takiej konkluzji upoważniają zbadane związki między wynikami badań obu skalami a rezultatami testów mierzących umiejętność cichego czytania ze zrozumieniem, poprawnego ortograficznie pisanie, znajomość materiału programowego z matematyki z zakresu klas V lub VI oraz z ocenami szkolnymi z jęz. polskiego, matematyki, geografii, biologii, historii, jęz. angielskiego i pracy–techniki.

Na podstawie wyników w skali pełnej WISC można z dużym prawdopodobieństwem prognozować rezultaty wszystkich trzech stosowanych testów osiągnięć szkolnych; natomiast na podstawie wyników w skali WISC-R można przewidywać sukcesy w opanowywaniu umiejętności cichego czytania ze zrozumieniem i przyswajania sobie wiadomości i umiejętności matematycznych. Poza tym, opierając się na wynikach w skali słownej WISC, można przewidzieć sukcesy lub niepowodzenie w testach Czytanie i Ortografia, natomiast wykorzystując wyniki w skali słownej WISC-R, można prognozować powodzenie w uczeniu się umiejętności cichego czytania ze zrozumieniem oraz nabywaniu wiedzy matematycznej.

Te same testy WISC i WISC-R w różnym stopniu umożliwiają przewidywanie opanowania przez badane dzieci umiejętności cichego czytania ze zrozumieniem, poprawnego ortograficznie pisanie oraz przyswajania materiału programowego z zakresu matematyki.

Wartości współczynników R są nieco wyższe dla poszczególnych skal WISC i WISC-R. Otrzymane dane są w przeważającej mierze zgodne z wynikami badań innych autorów.

### Trafność czynnikowa WISC i WISC-R

Analizę czynnikową wykonano wykorzystując metodę głównych składowych z rotacją Varimax. W skali WISC udało się wyodrębnić dwa podstawowe czynniki: wykonawczy (tworzą go testy: Klocków, Układanek,

Braków na Obrazkach, Symboli Figur i Cyfr, Porządkowania Obrazków, Labiryntów; Rozumowania Arytmetycznego; skale: wykonawcza i pełna są również nasycone tym czynnikiem) oraz werbalny (składają się nań testy: Słownika, Wiadomości, Podobieństw, Powtarzania Cyfr oraz skale: słowna i pełna). Wymienione czynniki tłumaczą 65,4% ogólnej wariancji.

Skala WISC-R mierzy trzy czynniki: werbalny (tworzą go testy: Wiadomości, Słownika, Arytmetyki, Podobieństw, Rozumienia i Powtarzania Cyfr oraz skale: słowna i pełna; czynnik wykonawczy (składają się nań testy: Porządkowania Obrazków, Układanek, Uzupełniania Obrazków, Wzorów z Klocków, Labiryntów; cała skala bezsłowna i cała pełna są również nasycone tym czynnikiem) oraz czynnik koncentracji uwagi i uczenia się wzrokowo-ruchowego (reprezentowany przez test Kodowania). Czynnik pierwszy wyjaśnia 52%, drugi 11,1%, trzeci 7,1% wariancji.

## Wnioski końcowe

Sformułowane poniżej wnioski potwierdzają hipotezy zweryfikowane przez autora:

1. System oceny wyników w testach: Podobieństwa, Rozumienie i Słownik okazał się wysoce obiektywny, zarówno w skali WISC-R, jak i WISC. Skala WISC-R odznacza się wyższą konsensyjnością (największa różnica dotyczy testu Rozumienia). Jak wynika z badań, 4 sędziowie kompetentni dokonują ocen różniących się o 1–2 P.P. w testach skali WISC-R i o 2–3 P.P. w testach skali WISC.

2. Analiza czynnikowa wykazała, że skala WISC-R mierzy trzy, a WISC dwa podstawowe czynniki: słowny, wykonawczy i koncentracji uwagi i uczenia się wzrokowo-ruchowego w skali WISC-R oraz werbalny i wykonawczy w WISC. Czynniki wyodrębnione w wyniku analizy wskaźników pomiaru za pomocą skali WISC-R są w dużej mierze zgodne z czynnikami wyróżnionymi przez A. S. Kaufmana (1975) i A. B. Silvers-teina (1986) w USA oraz A. Matczak, A. Piotrowską, W. Ciarkowską (1991) w Polsce. Natomiast wyniki analizy czynnikowej skali WISC, potwierdzają rezultaty wcześniej przeprowadzonych badań M. Choynowskiego (1969), K. Lowella i S. Shieldsa (1967), A. Maxwella (1959) i S. Jones (1962).

3. Wyniki skali słownej i pełnej obu technik korelują ze sobą w stopniu wysokim, a wyniki skali wykonawczej w stopniu umiarkowanym. Zadawalające korelacje uzyskano między wynikami pomiaru testami Wiadomości, Słownika i Klocków. Natomiast niepokojąco niskie współczynniki korelacji

odnotowano między wynikami testów: Labirynt i Porządkowanie Obrazków (współczynniki wskazujące na zbieżność wyraźną, lecz małą) oraz Układanki i Rozumienie (wskazujące na zbieżność istotną). Można zatem przyjąć, że testy te mierzą tylko w niewielkim stopniu te same zdolności. Podsumowując, skala WISC-R jest trafnym narzędziem pomiaru inteligencji, lecz na podstawie teorii należało oczekiwać większej zbieżności wyników w obu skalach.

4. Wyniki pomiaru z wykorzystaniem omawianych skal pozwalają w pewnym stopniu na prognozowanie późniejszych osiągnięć w nauce szkolnej w zakresie jęz. polskiego, matematyki, a także innych przedmiotów. Uzyskane współczynniki korelacji między wynikami pomiarów za pomocą obu skal i testów cichego czytania ze zrozumieniem, poprawnego ortograficznie pisania i wiadomości matematycznych oraz między ocenami szkolnymi należą najczęściej do rzędu umiarkowanych. Najlepszym predyktorem osiągnięć w nauce szkolnej jest skala pełna, naj słabszym – skala wykonawcza. Wartości współczynników korelacji wielokrotnej również wskazują na umiarkowaną zależność między wynikami w skalach: słownej, wykonawczej i pełnej porównywanych narzędzi, a uwzględnionymi łącznie wynikami z testów osiągnięć szkolnych oraz ocenami szkolnymi. Spośród ocen z wybranych siedmiu przedmiotów, na podstawie wyników uzyskiwanych w skalach słownej, wykonawczej i pełnej WISC i WISC-R, najpewniej można przewidzieć późniejsze oceny z matematyki. Wyniki badań skalą WISC-R pozwalają z największym prawdopodobieństwem przewidzieć wyniki uzyskiwane w teście cichego czytania ze zrozumieniem oraz w teście matematyki, a badania skalą WISC także wyniki w teście Ortografia.

5. Wyniki otrzymywane w skali WISC-R są istotnie niższe od wyników w skali WISC, przy czym największa różnica dotyczy skali słownej, największa – wykonawczej. Znaczące różnice występują również między odpowiadającymi sobie testami: Słownika, Podobieństw, Rozumienia, Porządkowania Obrazków, Układanek, Wzorów z Klocków, Powtarzania Cyfr.

6. Własności psychometryczne obu skal są porównywalne z własnościami ich oryginalnych, amerykańskich wersji (może poza nieco za niską – jak na zastosowane kryterium – trafnością diagnostyczną skali WISC-R) i innych uznanych na świecie narzędzi pomiaru inteligencji.

7. Przedstawione w pracy wyniki badań wykazują zbliżoną użyteczność obu skal jako narzędzi pomiaru inteligencji. O przewadze skali WISC-R nad WISC nie tyle świadczy jej znacząco wyższa rzetelność lub trafność, ile fakt, iż skala WISC-R jest skalą znormalizowaną na reprezentatywnej próbie populacji dzieci polskich. Zaletą skali WISC-R są wprowadzone modyfikacje do Słownika, Wiadomości, Arytmetyki, Powtarzania Cyfr,

Klocków. Skala ta nie jest jednak pozbawiona wad. Jej słabą stroną jest na pewno test Labiryntów, słabo różnicujący badane dzieci, a także test Układanek, w których wprowadzony przez autorki polskiej wersji skali system punktowania odpowiedzi jest na pewno dyskusyjny. Jak się również okazało, do trzech testów skali WISC-R (Słownika, Wiadomości i Rozumienia) dołączono niezgodne z empirią wskazówki dotyczące reguł przerywania testu.

Zaprezentowane wyżej wyniki badań pozwalają na stwierdzenie, że Skala Inteligencji D. Wechslera dla Dzieci – Wersja Zmodyfikowana (WISC-R) jest rzetelnym i trafnym narzędziem pomiaru inteligencji dzieci w wieku 11,0–11,11 lat (to samo można napisać o jej poprzedniczce – skali WISC).

## Bibliografia

- Chrzan-Feluch B., Kicińska M., Olecka J., Radziwiłł H. (1986), *Test „Matematyka kl. V” – podręcznik*, COMPW-Z MEN, Warszawa
- Flynn J. R. (1987), *Massive IQ Gains in 14 Nations: What IQ Test Really Measure*, „Psychological Bulletin”, 2, 171–191
- Kostrzewski J. (1970), *Wprowadzenie do Skali Inteligencji D. Wechslera dla Dzieci od 5 do 15 lat*, PTHP, Warszawa
- Matczak A., Piotrowska A., Ciarkowska W. (1991), *Skala Inteligencji D. Wechslera dla Dzieci – wersja zmodyfikowana – podręcznik*, Pracownia Testów Psychologicznych PTP
- Sattler J. M. (1988), *Assessment of Children Third Edition*, J. M. Sattler Publisher, San Diego
- Sobołewska M. (1986), *Podręcznik do testu „Czytanie” M. Grzywak-Kaczyńskiej*, COMPW-Z MEN, Warszawa
- Wechsler D. (1974), *Manual for the Wechsler Intelligence Scale for Children-Revised*, D. Wechsler Psychological Corporation, New York

PAWEŁ NOWAK

## Reliability and validity of WISC-R scale in the light of the research of 11 years old children

Psychometric characteristics of the Wechsler's Intelligence Scale for Children-Revised, (which was published in Polish version in 1991) were insufficient. Prognostic and consensual validity were not estimated, diagnostic validity was not fully estimated, also there was no researches comparing WISC and WISC-R. Wishing to complete that defects and trying to assess the Polish version of WISC-R author of this work examined a group of 62 children in age of 11,0–11,11. Children were tested with Wechsler's scales and after a period of one year, author studied children's school achievements, using tests of ability and knowledge and also school marks of seven school subjects.

The results of that research are as follows:

The WISC and WISC-R scales are highly reliable. The weakest test in both compared scales is Comprehension. WISC-R tests have higher coefficients of judge conformity than WISC tests.

Both compared scales allow estimating school achievements in Polish, mathematics and other subjects, but the most predicative result concerning school success or failures is the result in full scale of WISC and WISC-R.

WISC scale measures two factors (verbal and performance) and WISC-R scale three factors (verbal, performance and ability of concentration and visually-motorial learning).

Convergence of the results in both WISC and WISC-R scales is high (full and verbal scale) or moderate (performance scale). Low convergence of the results in the following tests: Composition, Labyrinth, Ordering pictures and Comprehension, indicates that those tests in both techniques measure different abilities.

The advantages of the WISC-R scale are: modification and modernization of some tasks in the tests, Polish normalization, assessment criterions of answers to the test's tasks. The disadvantages are: system of awarding points in the test Composition, weakly differing test Labyrinth, inconsistent with empiricism hints concerning the way of stopping examination in the tests Comprehension, Information and Vocabulary.

**Key words:** WISC-R Scale, reliability, validity.

## Aneks

Tabela 1

Współczynniki korelacji skal WISC z testami osiągnięć szkolnych w zakresie jęz. polskiego i z ocenami szkolnymi, poziomy istotności uzyskanych współczynników i procenty tłumaczonej przez nie wariacji

Skale	Miary	Czytanie		Ortografia		Oceny szkolne
		liczba				
		punktów	błędów	punktów	błędów	
Słowna	r	0,636	-0,358	0,632	-0,388	0,563
	p	0,0001	0,002	0,0001	0,001	0,0001
	r <sup>2</sup> · 100%	40,4%	12,8%	39,9%	15,0%	31,7%
Wyko- nawcza	r	0,603	-0,462	0,555	-0,313	0,402
	p	0,0001	0,0001	0,0001	0,007	0,001
	r <sup>2</sup> · 100%	36,4%	21,3%	30,8%	9,8%	16,2%
Pełna	r	0,704	-0,466	0,670	-0,394	0,542
	p	0,0001	0,0001	0,0001	0,001	0,0001
	r <sup>2</sup> · 100%	49,6%	21,7%	44,9%	15,5%	29,4%

Tabela 2

Współczynniki korelacji skal WISC-R z testami osiągnięć szkolnych w zakresie jęz. polskiego i z ocenami szkolnymi, poziomy istotności uzyskanych współczynników i procenty tłumaczonej przez nie wariancji

Skale	Miary	Czytanie		Ortografia		Oceny szkolne
		liczba				
		punktów	błędów	punktów	błędów	
Słowna	r	0,700	-0,409	0,612	-0,369	0,594
	p	0,0001	0,0001	0,0001	0,002	0,0001
	r <sup>2</sup> · 100%	49%	16,7%	37,4%	13,6%	35,3%
Wyko- nawcza	r	0,630	-0,363	0,567	-0,215	0,393
	p	0,0001	0,002	0,0001	0,047	0,001
	r <sup>2</sup> · 100%	39,7%	13,1%	32,1%	4,6%	15,4%
Pełna	r	0,746	-0,435	0,658	-0,329	0,552
	p	0,0001	0,0001	0,0001	0,005	0,0001
	r <sup>2</sup> · 100%	55,6%	18,9%	43,3%	10,8%	30,5%