

IV. DOBÓR PRZY PRZYJMOWANIU PERSONELU.

I. ZWYKŁE SPOSOBY PRZYJMOWANIA

Drugie ważne zadanie badania zdolności do zawodu dla celów doboru polega na tem, by ustalić, czy dana osoba nadaje się do danego zawodu, t. j. czy posiada właściwości, pożądane dla danego zawodu.

Celem odpowiedniego wyboru robotnika, czy pracownika używa się obecnie rozmaitych sposobów. Uwzględnimy tu pokrótce tylko niektóre z nich, aby podkreślić wyższość metod naukowych.

1. *Polecenia*. Najbardziej rozpowszechnioną i najzwyczajniejszą procedurą jest zatrudnianie ludzi na podstawie polecenia. Wprawdzie powszechnie wiadomo, że polecenia takie wydaje się tylko w tym celu, by poprzeć swego krewnego czy znajomego, (często jest to „przysługa za przysługę”) dalej, że idzie tu o to, by poleconemu w jakikolwiek sposób dopomóc. Wydolność jego jest rzeczą uboczną, a potęgą sugestji polecającego przyzwycięża wątpliwości. Ponadto zatrudnianie na podstawie poleceń często jest aktem wdzięczności, albo chęci zobowiązania sobie drugich. Na podstawie poleceń trafnego wyboru pracowników przeprowadzić nie można.

2. *Swiadectwa*. Nie można im przypisać żadnej większej wartości, ponieważ są zazwyczaj zbyt subiektywne.

Są pracodawcy którzy swoim pracownikom odchodzącym wystawiają świadectwa dobre, jednak niezgodne z prawdą „z dobrego serca” lub z chęci „nieszkodzenia”. Inni wystawiają najlepsze świadectwa, byle tylko pozbyć się pracowników, którzy okazali się złymi (pochwałami wyrzucają z zajęcia), albo żeby uniknąć przykrego zajścia przy pożegnaniu. Gdyby z drugiej strony dobry, obowiązkowy pracownik chciał zmienić miejsce, to wystawia mu się często świadectwo we wstrzemięźliwych, zimnych słowach, by utrudnić mu odejście. Prócz tych subiektywnych wpływów działa także okoliczność, że wielu przełożonych albo nie zna rzeczywistych właściwości pracownika i dlatego ich nie podkreśla, albo wprost nie jest zdolna do wystawienia świadectwa i ujęcia go inaczej niż w stereotypowe, nic nie mówiące frazesy. Dlatego też często świadectwo o właściwościach poszukującego pracy niczego nie wyjaśnia.

Szefowie wymagają często określonych właściwości kandydata, które jednak opierają się na całkiem fałszywej interpretacji faktów. Dlatego też naprz. w świadectwie szczególniejszą zwraca się uwagę na to, jak długo pozostawał pracownik w służbie. Nowe badania przyczyn zmiany zawodu pozwoliły wszakże stwierdzić znamieny fakt, że stosunek między zdolnością a zmianą zawodu nie jest jednoznaczny: zmianę zawodu w wielu zawodach uprawiają najmniej zdolni, w innych właśnie najzdolniejsi, którzy poszukują stanowisk jeszcze lepiej dostosowanych do swoich zdolności. Jak donosi James B. Miner, według jego badań na pracownikach wielkich sklepów, korelacja między zdolnościami a trwaniem służby wynosiła — 0.45, t. zn., że im zdolniejsza jest odnośna osoba, tym krócej zajmuje swoje stanowisko. W innym sklepie

zmieniało stanowiska w przeciągu sześciu miesięcy 40 — 50% dzielnych i nieudolnych pracowniczek a tylko 20 — 25% pracowników średniej miary¹⁾. Podobnie stwierdził również A. W. Kornhauser: Według dwóch badań zmiana zawodu (podczas sześciu miesięcy dla pracowników o małej inteligencji) wynosiła w pierwszym badaniu 30 — 45%, w drugim 30 — 60%, najrzadziej zmiana występowała u pracowników średniej inteligencji, najczęściej przy najwyższym stopniu inteligencji²⁾.

Widzimy z tego, jak ostrożnie należy przyjmować nawet te dane w świadectwach, które powszechnie uważa się za „przychylne”.

3. *Wywiady.* To samo co do świadectw, odnosi się również i do wywiadów osobistych z zastrzeżeniem, że w wyrażeniu słownym jeszcze większy upust mogą znaleźć afekty i uprzedzenia, niż w stwierdzaniu piśmiennem.

4. *Wrażenia osobiste.* Wielu przełożonych, majstrów i t. d. chwali się, że umieją ocenić kandydata według wrażenia osobistego: „Pierwsze spojrzenie wystarcza mi w zupełności”. Wypadki jednak, w których to się udaje, należą do rzadkości. Przy ocenianiu osoby mimowoli odgrywają rolę uczucia sympatji i antypatji. Jakiś szczegół w ubraniu, w wyglądzie zewnętrznym albo zachowaniu może uczuciowo odpychać, podczas gdy antypatycznie oddziałująca osoba może być najbardziej zdolnym pracownikiem. Ważny czynnik należy uwzględnić w podobnym osobistem przedstawieniu: są kandydaci, którzy przy tym

¹⁾ James B. Miner: Standardising tests for vocational guidance. School and Society 1921 VI 4, Cytuję według sprawozdania w ZangPs. T. 71. Str. 366.

²⁾ A. W. Kornhauser: Some business applications of a mental alertness test. JPeRes. T. I, zesz. I 1922.

akcie, tak ważnym dla ich losu, zachowują się bardzo „nerwowo”, niespokojnie, odpowiadają na pytania lękliwie i niewystarczająco, a w ten sposób wydają się „głupszy niż są w rzeczywistości” (pewna odmiana znanego lęku egzaminacyjnego). Z drugiej strony wielu jest takich, których takie przedstawianie się podnieca, a przez to okazują się ruchliwsi i wydolniejsi, aniżeli później przy pracy, gdy odpadnie podniecenie. Na pytanie czy można wnioskować o dzielności, potrzebnej w zawodzie, z osobistego wrażenia, w większości wypadków odpowiedź wypadnie przecząco.

H. Link¹⁾ spróbował przeprowadzić porównanie między pierwszym wrażeniem z rzeczywistą sprawnością pracownicy, zbadaną później. Wrażenie notował według kilku punktów; np.

Ogólna inteligencja: dobra

Rytm: dobry

Uwaga: dobra

Cechy cielesne: (smukła, zdrowa)

Powierzchność: (zręczna, miła)

Ogólne uwagi: (ambitna, zręczne ruchy, inteligentnie wygląda i t. d.)

Na podstawie takich sądów podzielił pewną liczbę dziewcząt na trzy grupy: na dobre, niezadowolające i nijakie. Ale kiedy przydzieliło się potem dziewczęta do dwojakiej pracy: do kalibrowania i kontroli, okazało się, że te które dobrze kalibrowały były złymi kontrolerkami. Proste wrażenie, — a więc metoda obserwacji — nawet u wyćwiczonego psychologa nie wystarczała, aby stwierdzić różnicę zdadności dziewcząt do rozmaitej pracy.

¹⁾ Link: l. c. Str. 14 i 122 (nast.).

Ścisłe związane z kwestją pierwszego wrażenia jest zagadnienie, czy fizjognomja może dać pewne wskazówki o zdolnościach człowieka. Stare teorie frenologiczne Gall'a i Lavatera pokutują podziś dzień i są popularne szczególnie w Ameryce, chociaż nie posiadamy dziś zupełnie jakiegoś bardziej obiektywnego potwierdzenia związku między rysami twarzy a właściwościami i zdolnościami, aniżeli przed stu laty. Zwracamy tu uwagę na teorię Blackforda i Newcombe¹⁾, która znalazła silny oddźwięk w Ameryce, a zajmowała się właściwościami „blondynów” i „brunetów”. „Normalni blondyni” ich zdaniem mają być pozytywni, dynamiczni, pełni inicjatywy, agresywni, opanowujący, aktywni, szybcy, optymiści i t. d. podczas gdy „normalni bruneci” mają posiadać właściwości przeciwne. „Teorię” tę uznali Paterson i Ludgate na podstawie badań za fałszywą²⁾. Obecnie w Brukseli „Instytut Nyssens” podkreśla ważność wiadomości fizjonomicznych dla ludzi interesu i uprawia w tym kierunku skrzętną propagandę zapomocą broszury L. A. Vauht'a: „Lecture pratique du Caractère”. Pozytywnych zdobyczy jednak mamy niewiele. Słuszne jest, że wielu przełożonych przez specjalny talent i wielkie doświadczenie osiąga jakiś „dar oceniania według fizjonomji, lecz jest to dar czysto osobisty. Większość ludzi widzi w człowieku to, co chce w nim widzieć, albo osądza go według analogji (np. X jest dzielny, Y wykazuje znaczne podobieństwo z nim, przeto Y również jest dzielny³⁾).

¹⁾ Blackford and Newcombe: The job, the man, the boss, New York, Doubleday 1919.

²⁾ Paterson D. G. i Ludgate K. E. Blond and Brunette Traits. A quantitative study. JPeRes, I., 1922, 122 — 128.

³⁾ Slawson John: The reliability of judgement of personal traits JApplPs. 6 (2) 1922, p. 161 — 171.

W wyższym jeszcze stopniu odnosi się to, co powiedziałam, do osądzania według fotografii, których obecnie prawie z reguły żąda się razem ze świadectwami. Poniżamy już to, że momentalny obraz współczesnej mocno retuszowanej (na korzyść dla modelu) fotografii nie oddaje prawdopodobnie nigdy cech charakterystycznych fizjonomji. Fotografia jeszcze więcej, niż żywe oblicze, daje sposobność do błędnych sądów, ponieważ przedewszystkiem brak w niej mimiki. W pewnych zawodach jednak, w których bardzo zależy na wyglądzie zewnętrznym (np. sprzedawczyni) można już według nadesłanych fotografii wydać sąd o zewnętrznej zdatności kandydatki i stosownie do tego przeprowadzić wybór.

Giese wymyślił bardzo ciekawą *próbę dla kandydatów na posady*, aby zbadać, czy można według fotografii i świadectw dokonywać trafnego wyboru. Próba polega na tem, że osobie badanej daje się 10 kart, z których każda zawiera fotografię wraz z dobrym wyciągiem ze świadectwa, z całkiem ogólnemi opinjami o pracy (zadowolający, pilny, bardzo dzielny). Osoba badana winna teraz na podstawie tych kart wybrać damę do salonu mód, która będzie prezentować na sobie eleganckie suknie.

Karty te zawierają następujące personalja:

1) Fotografia ukazuje zdrową, miłą twarz. Świadectwa, zdrowie znakomite. Wysokość 1,45.

2) Miła powierzchowność i dobre świadectwa. Przy podobiznie jednak umieszczono uwagę, że kandydatka była wychowywana w przytułku i uprawiała prostytutkę.

3) Wszystkie dane umiarkowane. Z powodu krótkowzroczności nosi obrzydliwe okulary rógowe.

4) Choć świadectwa dobre, świadectwo zdrowia zwraca uwagę na napady padaczki.

5) Widzimy nieco dziwną postać z włosami krótko ściętymi. Świadectwa i t. d. dobre.

6) Świadectwa i zdrowie zadowolające. Była karana za usiłowane morderstwo na własnem nieślubnem dziecku.

7) Wszystko mierne. Jedno ramię silniej spadające.

8) Cechy normalne. Wiek 47 lat.

9) Twarz piękna. Świadectwa zadowolające, z których jednak wynika, że kandydatka lubi zmianę miejsca.

10) Świadectwa dobre. Dama jednak bywa, jak wynika z papierów, dość często chora.

Najzdatniejsza więc jest kandydatka Nr. 5; powinna przeto być wybraną przez osobę badaną na podstawie karty. Zapomocą sekundnika mierzymy czas, jakiego potrzebuje osoba badana, ażeby zdecydować się. Prędkie rozstrzygnięcie i poprawny wybór są dowodem znajomości ludzi.

5. *Podania.* Pewnej orientacji o kandydacie dostarczają również piśmienne podania, które wystosowują starający się albo jako odpowiedź na anonsy, albo jako prośby do kierownictwa przedsiębiorstw w sprawie wolnej posady. Jeżeli podania wpłynęły na skutek anonsu, dokonanie wśród nich poprawnego wyboru nie powinno właściwie nastęrczać trudu; tylko ci mogą być wzięci pod uwagę, którzy najlepiej odpowiadają postawionym wymaganiom¹⁾.

W Ameryce, gdzie zmiana miejsc wogóle jest częstszą niż w Europie, a podania wpływają w wielkiej liczbie, wy-

¹⁾ S. Poffenberger A. T.: The letter of application in vocational selection. JApplPs 6 (1) 1922 S. 74 — 80.

konywano próby w tym celu, aby oceniać podania o posady. A więc podania, wnoszone o jakie zajęcie, szereguje się według właściwości które służą za podstawę oceny (t. zw. order of merit method). Jeżeli liczba jest zbyt wielka, to dzieli się podania na grupy (po 5 lub 7); grupę do której potem należą prawdopodobnie najodpowiedniejsi kandydaci ocenia się 1; drugą — 2 i t. d. przy obsadzaniu uwzględnia się potem rozumie się osoby z najlepszymi stopniami (1, 2).

Wielką liczbę skomplikowanych metod oceniania, które nie nadadzą się zupełnie w stosunkach europejskich, znajdujemy w Poffenbergera A. F. i Vartiana: „The letter of application in Vocational Selection”¹⁾. Dowodzą one jak silne jest dążenie do wynalezienia środków, by oceniać poprawnie podania i uczynić je godnymi zaufania. Dla tego ostatniego celu podany jest np. następujący sposób: 25 podań o posadę buchaltera zostało ułożone według czterech cech — inteligencja, czystość, rzetelność, takt, przez 50-ciu zawodowców i ludzi interesu. Po miesiącu proszono owych sędziów, by według tych samych cech ułożyli te same podania na nowo. Potem przeprowadzono porównanie między pierwszym a drugim sądem każdego sędziego w odniesieniu do każdej cechy. Przy silnych odchyleniach oceny sposób ten oczywiście nie może uchodzić za dopuszczalny. Próba podobna wykonana przez Waltona wykazała, że pewne cechy wykazują silniejszą zgodność, aniżeli inne; największą zgodność osiągnęła ocena taktu (taktowne zachowanie) najmniejszą — czystość²⁾.

¹⁾ JappePs 6 1922 S. 74 — 81.

²⁾ Cyt. według Griffitsa: Fundamentals of vocational psychology 1924. Str. 102.

Złe doświadczenia z podaniami należy przypisać w wielkiej mierze tej okoliczności, że ogłoszenia o wolnych posadach nie są układane w sposób celowy. Zakłady, które zawiadamiają o wolnych posadach, podają o nich całkiem skąpe wiadomości (o wynagrodzeniu nie wspomina się prawie całkiem, często opuszcza się nawet nazwę firmy, żeby kandydat nie zasięgał o niej wiadomości i in.) Z drugiej strony także kandydaci w swych piśmiennych zgłoszeniach nie podają zbyt wielu szczegółów o swej osobie, a pozostawiają to do osobistego przedstawienia się w firmie. Griffiths¹⁾ podaje dwie tablice ze statystyką wolnych posad na podstawie 123 ogłoszeń w gazetach amerykańskich wraz z wykazem wymagań, które stawia się kandydatom i danych, których się od nich żąda.

Z tablic tych bierzemy następujące liczby:

Ogłoszenia, które zawierają kilka danych o posadach do objęcia (biuro, pracownik bankowy, kupiec i t. d.)

Adres firmy	18
Wymagany wiek	15
Wymagane doświadczenie	46
Wykształcenie	6
Wysokość wynagrodzenia	5
Inne szczegóły dane	80

Te same ogłoszenia (123) żądały następujących wiadomości w podaniach kandydatów:

Wiek	42
Doświadczenie	62
Wykształcenie	10
Wymagania co do wynagrodzenia	54
Referencyj	28

(1) l. c. Str. 96 — 97.

Pisma ręcznego	3
Fotografij	0
Zadnych szczegółów	3

Widzimy więc, jak ubogie są dane o posadach, przez co utrudnione jest zgłaszanie się odpowiednich osób; również skąpe jest żądanie danych w podaniach, przez co utrudniony jest poprawny wybór.

To samo odnosi się także do Europy.

Na jednej stronicy anonsów „Berliner Tageblatt'u” (4 lutego 1925) zanalizowałam z tego punktu widzenia ogłoszenia firm o wolnych posadach.

Znajdowało się tam 47 anonsów (z częściowo podwójnemi posadami zaofiarowanemi jak np.: dekorator i sprzedawczyni); z tego było 10, a więc mniej niż czwarta część, takich, których sprecyzowano jako tako wymagania. No. „Inżyniera-kupca” z inicjatywą i skłonnościami organizacyjnymi do korespondencji i prac w zarządzie poszukuje berliński wielki zakład przemysłowy”. Prawie $\frac{1}{5}$ składa się jednak z ogólnych wyrażen, które prawie z reguły powtarzają się. Np. „Poszukujemy od 1 marca kilka *dzielnych* ekspedjentek i kilku *dzielnych* dekoratorów”. „*Wprawną* rysowniczką”, „*Obrotną* ekspedjentką”, „*Wybitnie* uzdolnioną damą do pomocy dla szefa”, „*Pierwszorzędnym* komiwojażerem i przedstawicielem”, „*Doskonałą* sprzedawczynią” i t. d. Dziwne, że zawsze szuka się „doskonałości”, a potem skarży się na marne wyniki pracy.

Na tej samej stronicy znajdowało się 20 ogłoszeń o posady. Z pomiędzy nich tylko 6 utrzymanych jest w tonie ogólnych zdań np. „smukła dama poszukuje stanowiska modelki w salonie mód”; prawie $\frac{3}{4}$ składało się jednak z mniej lub więcej wyczerpujących informacji o posiadanych zdolnościach, utrzymanych również w superlaty-

wach: „Nadreńczyk, 30 letni, który prowadził w ciągu 3 lat handel gazet z przeszło 30 dziennikami i czasopismami, dobry organizator, umiejący po francusku i po angielsku, pozostający dotychczas na dawnym stanowisku jako kierownik interesu, zamieni swą posadę na przedstawicielstwo wielkiego dziennika”.

Gdyby udało się przeprowadzić dobrą organizację anonsów i podań oraz ustalić wytyczne dla redagowania ich, to przyczyniłoby się znacznie do poprawnego wyboru pracowników. Dotychczas, jak to widać z przytoczonych przykładów, sprawa znajduje się na bardzo niskim poziomie.

6) *Wypełnianie formularzy podaniowych.* Specjalny rodzaj podań stanowią formularze, które muszą być wypełnione przez ubiegających się o posadę. Przy obsadzaniu posad używa się dwóch rodzajów formularzy. Jeden polega na przedłożeniu kilku pytań (które często wypełnia się w poczekalni firmy), na podstawie których pracodawca zyskuje pewną orientację o kandydacie; to ułatwia mu ustne załatwienie sprawy. Inny rodzaj formularzy wypełnia się dopiero na stanowisku i zawiera wszelkie możliwe wiadomości o przeszłości, przebiegu życia i t. d. (coś w rodzaju policyjnej kontroli). My uwzględnimy tylko pierwszy rodzaj, który jako środek wyboru nadaje się lepiej, aniżeli podania, ponieważ wszystkim kandydatom stawia się jednakowe pytania tak, że odpowiedzi dają od razu przegląd co do tych kandydatów, którzy najlepiej odpowiadają postawionym wymaganiom.

Przy tych formularzach wiele zależy od ułożenia pytań. Muszą one tak być dobrane, żeby wydobywały cechy najważniejsze dla zawodu, a nie powinny zbyt znacznie rozciągać się na rzeczy uboczne, osłabiające tem samem zain-

interesowanie odpowiadającego. Bardzo rzadko jednak zostaje to utrzymane. Nawet kwestionariusz, który Edison przedłożył do odpowiedzi starającym się o stanowiska inspektorów warsztatowych (jedną jego część ogłosił jeden z kandydatów Charles Hansen z pamięci w „New York Times” z 11 maja 1921¹⁾) dowodzi, o ile rzeczy zbędnych i bez znaczenia wypytują się tam.

Nie powinno się pominąć tutaj faktu, że Amerykanka Miss D. B. Goldsmith²⁾ zużytkowała daty formularzy jako środek do stwierdzenia dzielności agentów pewnego towarzystwa ubezpieczeń. Każdą cechę, wzgl. każdą odpowiedź na pytania formularza (wiek, wykształcenie, stanowisko, doświadczenia i t. p.) oceniała w punktach 0 — 4 (np. wiek: lata 28 — 29 oznaczony został przez + 2, lata 30 — 40 przez + 3 i t. d.), a potem sumę punktów, otrzymanych w ten sposób, porównała z późniejszymi wyczynami, (wzgl. z wysokością wynagrodzenia, które w Ameryce stosuje się do wyczynu). Według wiadomości, podanych przez autorkę, agenci najwyżej ocenieni w ten sposób, w praktyce osiągałi również najwyższe wyczyny.

Praca ta (której dokładniej nie mogłam poznać) może stanowić dowód, że poprawne stawianie pytań w formularzach dostarcza również poprawnych wiadomości o przyszłych zdolnościach kandydata. Ale od takiej rozbudowy formularzy jesteśmy jeszcze bardzo dalecy.

7) *Charakter pisma*. Wobec rozwoju, jaki osiągnęła grafologia dzięki podstawowej pracy K l a g e s a²⁾, w ko-

¹⁾ Goldsmith: The Use of the Personal History Blank as a Selesmanship Test. Jappls 6 1922.

²⁾ Dr. K l a g e s: Handschrift und Charakter. (Pismo i charakter). Leipzig Barth -0 Aufl., 1927.

łach przemysłowych odgrywa coraz to większą rolę charakter pisma jako środek poznania człowieka. Żąda się „własnoręcznie” pisanych ofert, podań, życiorysów, nawet kilku stron prób pisma (szczególnie na wyższych stanowiskach) a grafolog, zatrudniony w firmie, albo ktoś, kto zajmuje się tem „zawodowo”, „określa”. Jest to o tyle uzasadnione, że pismo wyjawia *właściwości charakteru* jak wytrwałość, gorliwość, rozrzutność, lekkomyślność, bojaźliwość, aktywność, skromność, trzeźwość, sumienność i t. d. które to właściwości w wykonywaniu zawodu mają ogromne znaczenie. Jeżeli się chce mieć np. spokojnego, skromnego człowieka jako sekretarza, a pismo starającego się o posadę wykazuje krnąbrność, wielką ambicję, nie należy wtedy doradzać zatrudniania go u siebie. Ale, jak fizjonomja, tak i pismo dotychczas (nie przesadzamy tego, co może być w przyszłości) nie zdołają udzielić nam wiadomości o zdolnościach *intelektualnych*, pożądanych dla zawodu. Nie uzyskujemy z pisma żadnej wiadomości o rodzaju uwagi, pamięci, sprawności zmysłów, zmyśle mięśniowym i t. p. właściwościach, które konieczne są do wykonywania pewnych zawodów. Pismo pozostanie więc zawsze tylko *jednym* ze środków pomocniczych w stwierdzaniu zdatności do zawodu. Ważne również jest i to, żeby oceny pisma dokonywała osoba dokładnie obznajmiona z tą sztuką, oraz wyszkolona w praktycznem jej stosowaniu; wobec rzadkości takiego należytego wykształcenia grozi niebezpieczeństwo, że powierzy się losy ludzi dyletanckiemu szarlatanowi.

8. *Rozmowa.* Z nienaukowych sposobów do wybierania najlepszych kandydatów na posady ma rozmowa największe znaczenie i bywa powszechnie używana. Wszystkie omawiane dotychczas sposoby postępowania są w za-

sadzie tylko środkami pomocniczymi dla niej, gdyż zarówno podanie, jak pismo ręczne, polecenie i fotografia prowadzą ostatecznie do rozmowy, która ma wszystkie dotychczasowe wrażenia zebrać razem i skontrolować.

Rozmowa ma właściwie dwa cele: umożliwić wrażenie powierzchowności (wzrost, wyraz twarzy, zachowanie się, manieri), a ponadto dać pewien wgląd w charakter, inteligencję i wiadomości kandydata. Pierwszy omówiliśmy już powyżej, a obecnie naszkicujemy tylko cel drugi.

Pierwsze pytania, postawione kandydatowi, odnoszą się w pierwszej linii do jego wydolności. „Czy zna Pan pracę? Czy potrafi Pan ją dobrze wykonać? Czy już spełniał Pan podobne obowiązki?” Zwracamy się do *uczciwości* kandydata, do jego szczerości. O jej rozmiarach możemy się jednak przekonać dopiero wtedy, gdy po zatrudnieniu porównamy jego obietnice z rzeczywistymi wynikami pracy. Drugą kategorię pytań stanowią mniej lub więcej ukryte egzaminy, dotyczące wiadomości, które musi posiadać kandydat. Jako trzecie stawia się pytania, które w każdej rozmowie uzależnione są od odpowiedzi kandydata, krótko mówiąc, pytania czysto ludzkiej natury (o położeniu obecnem i dawniejszem pracownika, o stosunkach rodzinnych i t. p.). W ostatnim czasie w Ameryce przy zatrudnianiu pracowników stawia się często pytania, odnoszące się wprost do zawodu, który chce objąć kandydat. Np.: „Jakie doświadczenie zdobył Pan jako monter karabinów?” „Niech mi Pan wymieni najważniejsze urządzenia w karabinie?” „Dlaczego, według Pańskiego zdania, odtylcowość karabinów jest ważna?” i t. p.¹⁾.

¹⁾ Link, o, c. Str. 118.

Bez wątpienia są tacy przełożeni i szefowie, którzy przez talent, lub doświadczenie umieją zinterpretować poprawnie każdą odpowiedź w połączeniu z grą mimiczną (którą uważnie obserwują) i dlatego przy zatrudnianiu robotników na podstawie takiej rozmowy nie popełniają nigdy błędów. Iluż jednak jest takich? Prowadzenie rozmowy jest szczególniejszą sztuką. Rzadko spotyka się ludzi, którzy nie zbaczają od tematu, częściej spotykamy takich, którzy schodzą na drobnostki zamiast trzymać się rzeczy najważniejszych, traktują sprawy osobiście zamiast ograniczyć się do rzeczowości, a co najważniejsze... fałszywie tłumaczą sobie odpowiedzi. Nie baczą zupełnie na to, albo i przeocząją zgoła, że kandydat w chwili rozmowy znajduje się w stanie mniej lub więcej silnego podniecenia, które go stosownie do jego usposobienia, albo ożywia (ekscytuje) tak, że wydaje się lepszym niż jest w życiu codziennem, albo przeciwnie tak go kępuje, że najlepsze jego strony całkiem nie występują. Nieświadomość, czy otrzyma posadę, a przez to uwolni się od troski o byt, czyni go niepewnym siebie. Stawiający pytania musi więc być dobrym obserwatorem, a ponadto musi posiadać zdolność wczuwania się w psychikę drugiego, z jednej strony, aby stawiać kandydatowi właściwe pytania, jeżeli potrzeba, w tonie ośmielającym — z drugiej strony, by odpowiedzi i wiadomości, dostarczane przez kandydata o sobie samym, tłumaczyć poprawnie. Ponadto nie może być podatny na wpływy, gdyż niektórzy kandydaci narzucają silne sugestje, którym niekiedy może łatwo ulec.

W kilku amerykańskich zakładach przemysłowych spróbowano zapomocą drukowanego kwestionariusza ułatwić pytania przełożonemu. (Kandydat nie wypełnia kwe-

stjonarjusza (Link¹⁾) przytacza w swojej książce podobne zestawienia, które tutaj reprodukujemy:

- Gotowość do pracy: dawniejsze świadectwa dobre
- Czy chce wykonywać ten rodzaj pracy
- Bierny
- Próbował innych rodzajów pracy
- Niezadowolający
- Znajomość pracy: długoletnie doświadczenie w tej pracy
- dobrze wykształcony
- zadowolający
- zupełnie nie
- Wierność: Mówi dobrze o wszystkich swoich dawnych przełożonych
- niebardzo dobrze
- złe
- Trzeźwość: Abstynent
- umiarkowany
- pijak okresowy
- pijak nałogowy
- Dążenie do dalszego wykształcenia: szkoły dokształcające
- inne studia
- ambicja
- Stałość w pracy: najdłuższy czas zajęcia
- ustawiczne zmiany.

Jak Link zaznacza, nie wszystkie pytania są celowe; albowiem na pytanie: „Czy pan jest pijak? Każdy skwapliwie odpowie: „nie”²⁾. Również zbędne są pytania dotyczące dawnych przełożonych. Ale, niewątpliwie, poprawny wybór z tych pytań odda dobre usługi przełożonemu.

W bardzo wielkich zakładach przemysłowych, gdzie personel często się zmienia (np. w domach towarowych!),

¹⁾ l. c. Str. 114

²⁾ Zresztą pytanie to z powodu prohibicji w Ameryce stało się niepotrzebne.

znajdują się specjaliści „referenci” od przyjmowania pracowników, gdzieindziej sam szef, kierownik oddziału lub wermistrz i t. p. przyjmują do zajęcia personel, który potem im podlega. Czy bardziej celowy jest system pierwszy czy drugi — to nie jest jeszcze rozstrzygnięte. Walter Dill Scott (według danych dostarczonych przez Griffitsa) dał przyczynek do tego zagadnienia. W „American Tobacco Company” ośmiu nadzorcom i jednemu kierownikowi oddziału postawił kolejno to samo zadanie: z pomiędzy 29 ubiegających się wybrać 15 i każdemu przydzielić odpowiednią rangę. Okazało się, że jeden z kandydatów od jednego z nadzorców otrzymał miejsce 1, od drugiego miejsce 19, od trzeciego miejsce 22. Później okazało się w praktyce po kilku tygodniach, że żaden z praktycznie najdzielniejszych pracowników nie został przez żadnego z sędziów oceniony nawet jako „dobry”¹⁾. Okazuje się więc, jak niepewna jest (wyłączywszy wypadki osobistego talentu) także i rozmowa przy zatrudnianiu pracowników.

9. *Praca próbna i czas próby.* Najpewniejszym sposobem stwierdzenia zdolności kandydata jest t. zw. praca próbna t. j. efektywny wyczyn, który wykonać musi kandydat, aby udowodnić swoją zdolność i sprawność; tak np. nowozgłaszającej się stenografistce każe się odczytać stenogram, zgłaszającemu się przykrawaczowi przykroić jakąś część ubrania i t. d. Ale często wyczyny zawodowe są tego rodzaju, że nie można ich uchwycić w jednej pracy próbnej (tak np. kupiec, agent handlowy, sekretarz, mąż zaufania); wtedy wprowadza się czas próby t. zn. zatrudnienie na niewielki przeciąg czasu, (kilka dni do kilku ty-

¹⁾ Według Griffitsa l. c. Str. 129 i nast.

godni). Niewątpliwie taki czas próby, podczas którego nowozatrudnionego pracownika obserwuje się bardzo bystro, da dokładne pojęcie o jego zdolnościach. Jeżeli jednak ów krótkotrwale zatrudniony nie odpowiada wymaganiom i musi się skutkiem tego brać nowego na próbę — to jest to, naturalnie, znaczną stratą czasu. Czasowa praca „na próbę” nie spełnia również swego zadania o tyle, — jak to zauważono — że pracownik, przyjęty na próbę, zadaje sobie w tym czasie jaknajwięcej trudu, żeby uzyskać posadę na stałe, a stąd stwarza pewne pozory, że istotnie pracuje wydajnie.

Najkorzystniej przedstawia się wypadek wtedy, gdy na czas próby przyjmuje się pewną liczbę reflektantów, a zatrzymuje tego, który się okaże najlepszym. W takim postępowaniu jednak moment pieniężny odgrywa pewną rolę, a ponadto w małych zakładach przemysłowych także niemożność dania odpowiedniego zajęcia odnośnym osobom.

10. *Szkolenie zawodowe.* Niezawsze przyjmuje się na próbę takie osoby, które w pewnym zawodzie są już wykształcone, lecz często idzie o takie, które dopiero muszą wyszkolić się w pewnym określonym zajęciu. Za typowy przykład podobnych okresów próbno-kształcących uważamy szkoły zawodowe, których uczniowie, nie wykazujący w pewnym określonym czasie wykształcenia pewnych określonych zdolności, zostają usunięci np. szkoły telefonistek i telegrafistek i t. p. Należy tu jednak uwzględnić znaczne koszty, szczególnie, jeżeli niekiedy tylko niewielki procent z wielkiej masy kształcących się osiągnie rezultaty. Właśnie pewne towarzystwo telegraficzne (Bell Telephone Co) musiało zwolnić z pracy znaczną część dziewcząt, które ukończyły szkołę przygotowawczą i wtedy

zwróciło się do Münsterberga z zapytaniem, czy nie dałoby się zgóry ocenić psychicznej zdatności telefonistek. Także w Niemczech psychotechniczne badania zdatności powstały najpierw w jednej ze szkół zawodowych (np. w fabryce A. E. G. w Berlinie). Jest to dowodem, że i szkolenie nie jest zadowalającym środkiem do przeprowadzania selekcji.

W ostatnich latach traktuje się i wprowadza nauczanie i szkolenie z całkiem nowych punktów widzenia. Gastew¹⁾ w Rosji, Friedrich²⁾ w Niemczech stoją na stanowisku, że *racjonalny* sposób nauczania doprowadziłby do zniesienia różnic poziomów i wykształciłby zarówno dobrze *wszystkich* kandydatów do tych samych wyczynów. Przeto poprzednie, wybieranie byłoby całkiem zbyteczne. Ale praktyka nie wypowiedziała o tem rozstrzygającej opinii, podobnie jak i nauka nie rozwiązała ważkiego dla kształcenia problemu ćwiczenia t. j. wyćwiczalności rozmaitych zdolności. I dlatego pozostaje w użyciu obecnie wszędzie pospolita metoda szkolenia jako niezupełnie wystarczający środek do przeprowadzenia doboru przyszłych pracowników.

Należy obecnie zebrać razem rezultaty pospolitych metod doboru pracowników:

Świadectwa, wystawiane ustnie czy piśmiennie, osobiste wrażenia, odpowiedzi reflektantów udzielane w kwestjonariuszach, małe okolicznościowe badania (rękopisów),

¹⁾ Por. F. Baumgarten: Arbeitswissenschaft und -Psychotechnik in Russland (Wiedza o pracy i psychotechnice w Rosji), München, Oldenburg 1914.

²⁾ A. Friedrich: Das Anlernen auf psychotechnischer Grundlage (Szkolenie na psychotechnicznych podstawach), PrakPs. 4 (I) 1922.

próbne prace, czasy próby, szkolenie nie wykazały się jako wystarczające środki do wynalezienia wydolnego pracownika. Dążono nadal do wynalezienia takiego postępowania, któreby w najkrótszym czasie w obiektywny jednoznaczny sposób z najmniejszym wysiłkiem i najmniejszymi kosztami pozwoliło na wydanie pewnego sądu o człowieku zupełnie dotychczas nieznanym. Za takie postępowanie uważa się obecnie *metodę testów*.

2. POSTĘPOWANIE NAUKOWE. METODA TESTÓW.

A. Definicja.

Test (ang. próba) jest to ściśle określone zadanie, które w najprostszy sposób w najkrótszym czasie na dowolnem miejscu (t. j. poza jakimś laboratorium) może stwierdzić pewne określone zdolności osobnika. Jeżeli się chce np. przewidzieć, czy zgłaszający się podoła ciężkim zadaniom pamięciowym, wymagany przez jego przyszłe stanowisko, to daje mu się test pamięciowy: po jednorazowym odczytaniu 12 szeregów po 3 pojęcia, pozostające ze sobą w związku np. choroba — lekarz — polepszenie; zima — bieda — mróz; każe mu się je powtórzyć. Jeżeli powtórzenie odbywa się odrazu bez błędu, to wnioskujemy z tego, że reflektant także w skomplikowanych sytuacjach pod względem pamięci dorasta do zadań. Zapomocą testu wydobywamy od badanego odpowiedzi, z których wnioskujemy o jego wydolności. Test przeto jest pewnego rodzaju dorywczą próbą do badania zdolności. Ponieważ próbę tę przeprowadza się w określonych warunkach eksperymentu, więc można porównywać wyniki badania większej liczby osobników.

W literaturze naukowej zalecił test: pewnie po raz pierwszy założyciel nauki o eugenicie Francis Galton¹⁾ w swojej książce: „Inquiries into human faculty”. (1883) Amerykanin Cattell użył pierwszy małej serii testów do badania inteligencji ludzkiej²⁾ w Niemczech C. Rieger równocześnie z Cattellem napisał szkic o badaniu testami³⁾. W tym samym czasie we Włoszech Guicciardi i Ferrari, a potem także Lombroso używali tej metody do stwierdzenia właściwości psychicznych u przestępców⁴⁾. Ale te początkowo używane testy

¹⁾ Cyril Burt: *Englisch Contributions to the Study of Mental Tests* w: *Segona Conferencia Internacional de Psicotécnica Aplicada a l'Orientacio Professional*. Barcelona 1922. Str. 322.

²⁾ Cattell, Mc Keen: *Mental Tests and Measurements*. Mind 1890. Str. 377 — 381.

³⁾ Rieger C.: *Beschreibung der Intelligenzstörungen* infolge einer Hirnverletzung nebst einem Entwurf zu einer allg. anwendbaren Methode der Intelligenzprüfung. (Opis zaburzeń inteligencji skutkiem zranienia mózgu wraz z przyczynkiem do ogólnie używanej metody badania inteligencji). *Verhandlungen der phys. med. Gesellschaft in Würzburg* N. F. 25, Nr. 4, 1899).

⁴⁾ Test, jako środek do badania człowieka pod względem mądrości i dzielności właściwie nie jest wynalazkiem nowożytności. Węzeł gordyjski, którego Aleksander Wielki nie rozwiązał, lecz rozciął, jest testem, a jego rozwiązanie charakteryzuje wielkiego męża pod względem przystępowania do trudności. Zagadka, którą Sfinks zadawał Tebanom, a której rozwiązanie miało oddać panowanie nad Tebami i rękę królowej Jokasty, jest również testem. Sędziowie Wschodu posługiwali się testami, ażeby w trudnych sporach, w których brak było świadków wydać sprawiedliwe sądy. Tak jest w sławnym Salomonowym sądzie: z dwóch matek, przyznających się do jednego dziecka, rozpoznanie prawdziwej matki przez propozycję zabicia dziecka, przez co uczucia macierzyńskie ujawniły się u prawdziwej matki. Toż samo jest i z prośbą sędziego perskiego, który miał rozstrzygnąć, który zdwu mężów jest panem, a który nie-

odnosiły się do badania pojedynczych wyczynów zmysłowych, ruchów i elementarnych przebiegów psychicznych. (Np. zadaniem Cattella było badanie maksymalnej szybkości ruchu ramion, stwierdzenie liczby zapamiętanych głosek po jednorazowym przeczytaniu, przepoławianie kreski na oko). Rozstrzygający krok naprzód uczynili w r. 1896 Binet i Simon przyczem Binet zwrócił uwa-

wolnikiem. Rozkazał katowi spuścić miecz na głowy obydwoh; jeden usunął wtedy głowę, drugiemu nie drgnął ani jeden muskuł i ten właśnie został uznany za prawego pana. W historii średniowiecza spotykamy wiele przykładów postępowania sądowego, które posiadają formę „testu”, np. t. zw. „sąd boży” — rzucano kobietę oskarżoną o czary do wody: jeżeli pływała po powierzchni, była niewinna, jeżeli tonęła, była czarownica. W „Żywotach” Vasari'ego znajdujemy ciekawy przykład testu, wymyślonego spontanicznie przez artystę. Poseł papieża Benedykta XI przybył do Giotto i prosił go o mały rysunek, który chciał pokazać papieżowi. Giotto chwycił kawałek papieru i przycisnąwszy mocno ramię do tułowia tak, że stanowiło ono jakby cyrkiel, i zatoczywszy ręką koło, wyrysował pędzlem umaczanym w czerwonej farbie, doskonałe koło wraz ze średnicą. Dworzanin myślał, że Giotto uważa go za głupca. Tymczasem papież poznał z tej próby, o ile Giotto przewyższał wszystkich innych malarzy swego czasu doskonałością rysunku (Giorgio Vasari. Die Lebensbeschreibungen der berühmtesten Architekten, Bildhauer und Maler). (Żywoty najsławniejszych architektów, rzeźbiarzy i malarzy). Tłum. niem. M. Wackernagel. Strausburg, Heitz 1916. Str. 170 — 172). Jest to już test taki, jakiego używa współczesna psychologia. W „Contes philosophiques” Voltaire'a znajdujemy również znaczną ilość takich testów. Współczesny test odróżnia się od tych prób przez swoje opracowanie naukowe, one jednak przewyższają go doskonałą znajomością człowieka. O testach w poc. w XIX pisze Kammel: „Beitrag zu einer Geschichte der Testforschung im 18 und 19 Jahrhundert (Przyczynek do historii testu w XVIII i XIX w.). Jahrbuch des Vereins für christliche Erziehungswissenschaft 17, 1926.

gę na to, że indywidua różnią się między sobą o wiele więcej przez swoje t. zw. wyższe funkcje psychiczne jak uwaga, fantazja, inteligencja i t. p. aniżeli przez czynności zmysłowe; obydwaj więc skierowali testy ku badaniu tych funkcji u dzieci. Badania dokonali na setkach dzieci, przy czem stworzyli testy odpowiednie dla różnych stopni wieku (system szczeblowy badania inteligencji). W ten sposób udało im się wynaleść normy psychiczne, służące do charakterystyki rozwoju duchowego t. j. ustanowić *wyczyny normalne*. Przez to stało się znowu możliwe użytkowanie testów do badania anormalnych duchowo (upośledzonych) a w ten sposób wprowadzona została metoda testów także do pedagogiki i medycyny.

Jako dalszy etap rozwoju metody testów można uważać próbę Münsterberga, by użytkować metodę testów do stwierdzania zdolności do zawodu. Metoda zyskała przez to olbrzymie pole zastosowania w przemyśle, handlu i rzemiośle; a dążąc do służenia zadaniom praktycznym uczyniła wielki krok naprzód w doskonaleniu się i zyskiwaniu coraz to większej pewności.

B. Rodzaje testów.

Testów czyli dorywczych prób badania zdolności duchowych, istnieje obecnie bardzo wiele. Według pewnej już przestarzałej i bardzo niezupełnej bibliografji jest ich kilka tysięcy; tymczasem już liczba ich wzrosła. Podanie dokładnej ich liczby jest zresztą utrudnione z tego powodu, że są testy, wynalezione jedynie dla pewnego określo-

¹⁾ Mitchell D. and Ruger G. J. Psychological Tests. Revised and classified. New York. Bureau of d. Experiments, Bull 9, 116 S 1918.

nego celu, a są i takie, które miały inne przeznaczenie, a używane są dla innego, jakiegoś nowego celu. Testu F b b i n g h a u s a (wypełnianie luk w niekompletnym teście) używano początkowo jako testu inteligencji, później także jako testu do badania sprawności mowy i zdolności kombinacyjnej. Testu B o u r d o n a (wykreślanie pewnych liter w tekście) używa się jako testu uwagi, koncentracji i znużenia. Liczba testów zmienia się zależnie od tego, czy przytacza się sam test jako taki, czy też i wszystkie możliwości zużytkowania go.

Zależnie od punktu widzenia można testy rozmaicie podzielić¹⁾. Są testy:

1) *nieme* (test *muet* albo *non* — *verbal*), jeżeli zadanie musi być wykonane zapomocą ręki np. podział kąta prostego na połowy, ustawienie pewnej figury z klocków, wyginanie z drutu według wzoru (w Ameryce gdy idzie o takie testy zręczności) nazywają się one „*performance test*”, Nadają się one z jednej strony, dla analfabetów i mało wykształconych, z drugiej strony służą do badania zdolności, które wyrażają się w ruchach (np. zręczność);

2) *testy werbalne* jeżeli badany musi dać na pytanie odpowiedź ustną i

3) *testy piśmienne*, jeżeli badany musi dać odpowiedź na piśmie. Specjalną grupę stanowią testy t. zw. *aparaturowe* (które szczególnie rozpowszechnione są w Niem-

¹⁾ Nie podajemy tu żadnych „systemów” klasyfikacji testów tak, jak je sporządzili Claparède, Cyril Burt, Joteyko, Kornhauser i Kingsbury, gdyż chodzi nam jedynie o czysto-praktyczne punkty widzenia. Ale dotychczas nie ma jednności zarówno w nazwach, jak i w podziale testów; jest to zadanie, któremu ma się poświęcić w niedalekiej przyszłości Międzynarodowa Komisja Psychotechniczna.

czech); wtedy badanie zdolności duchowych przeprowadza się na mniej lub więcej skomplikowanych przyrządach i aparatach (aparat do badania pamięci, koncentracji i w. in.).

Każdy z tych rodzajów testów można zużytkować z korzyścią w określonych celach. Badanie zapomocą aparatów posiada określone zalety: 1) aparaty zezwalają na dokonanie zarówno badania, jak i oceny wyczynu w zawsze tych samych warunkach, 2) każde badanie może być w każdej chwili w tych samych warunkach *powtórzone*, a w ten sposób staje się możliwa *obiektywna kontrola*, 3) można uzyskać precyzję w obliczaniu rezultatów, 4) można zmniejszyć, albo nawet wyeliminować w zupełności wpływ osobowości badającego, 5) aparaty zezwalają na lepsze obserwowanie badanego, ponieważ badający uwolniony jest od posługiwania się środkami pomocniczymi, 6) nadają się dla pracowni o dużym ruchu i przy niedobrze personelu, ponieważ zaoszczędzają wiele czasu, 7) mimo wielkich kosztów, jakie pociąga za sobą jednorazowy zakup aparatów, są one w rezultacie tańsze, jeżeli chodzi o większą liczbę badań, ponieważ zużywa się przy nich znacznie mniej materiałów i utensyliów pisarskich.

Testy aparatuowe mają z drugiej strony tę niższość, że nie dają ściśle indywidualnego obrazu wyczynu człowieka badanego, co ma miejsce w testach piśmiennych. Ponadto, ponieważ dzieci mało są obeznane z aparatami, są więc często niemi onieśmiałe i zaskoczone, co wpływa obniżająco na wynik. Wreszcie mają aparaty i tę ujemną stronę, że nie pozostawiają *trwałego* obrazu odpowiedzi badanego, co ma miejsce np. w odpowiedziach piśmien-

nych, które zezwalają na zademonstrowanie wyczynu w każdym czasie¹⁾.

Według rodzaju *żądanej odpowiedzi* dzielią się testy na 1) *alternatywne*, jeżeli idzie tylko o ściśle określoną odpowiedź wzgl. określone wykonanie ruchu, jak np. ustawić na aparacie kąt *prosty* lub *orzec*, czy sąd „wszyscy ludzie są śmiertelni” jest prawdziwy, czy fałszywy, t. j. może tutaj iść tylko o: „albo — albo”, 2) *wolne niewyznaczone* testy, w których odpowiedź pozostawiona jest wolnemu sądowi albo swobodnemu wykonaniu tak, że badany może wydać dowolny sąd, 3) *testy wyznaczone*. W Amerykańskich testach piśmiennych używa się jeszcze pewnego szczególniejszego rodzaju odpowiedzi; mianowicie kandydat musi podkreślić z wielu, przedłożonych mu sądów ten, który wydaje mu się właściwy; np. w zdaniu „Volt jest miarą elektryczności, siły wiatru, opadów wodnych, opadów śniegowych, sił wody” — ma być podkreślone: „elektryczności”. Jest to tak zwana metoda szeregów wartości (*scale of values method*).

¹⁾ Dotychczasowe wielkie rozpowszechnienie się aparatów spowodowane zostało rozmaitemi względami, które nie mają nic wspólnego z nauką. Z jednej strony aparaty pozwalają na większy zysk materialny dla pewnej wymyślonej metody drogą patentów ochronnych, powtórę odgrywają ważną rolę w rozpowszechnianiu ruchu psychotechnicznego. Aparaty są bowiem widzialnym znakiem badań psychologicznych; wywierają one wielkie wrażenie na laiku, dla którego aparat jest zawsze symbolem dokładności i naukowości. Dlatego to już zgóry jest się lepiej uprzedzonym do badania z pomocą aparatów, niż do skromnego papieru i ołówka. Rozwiązania, umożliwione przez aparaty, niejednego laika przekonały do badań.

Ponadto, im bardziej niepewnym czuje się ktoś co do swoich zdolności jako badający, tem chętniej chwyta się aparatu. Widzi się to wyraźnie przy zakupie aparatów przez obcokrajowców w Niemczech.

Według *ogólnego celu badania* rozróżniamy 1) *testy pojedyncze*, które badamy poszczególne duchowe zdolności (jak uwagę, pamięć). Testy uzyskują wtedy nazwę według swego celu: testy pamięciowe; testy uwagi. Jeżeli używamy jakiegoś testu w formie stopniowanej, aby zmierzyć stopień zdolności np. w badaniu wrażliwości dotykowej zapomocą szeregu płytek, których powierzchnia jest coraz to gładziej wypolerowana, to taki test nazywa się *testem stopniowym*. 2) *Szeregi testów* (ang. test groups) t. zn. mniejsza lub większa liczba testów, których używa się, by zbadać pewną całość zdolności. Jeżeli np. chce się stwierdzić duchowy poziom robotnika, to używa się testów, które wypróbowują rozmaite zdolności i wiadomości. 3) *Serje testów* albo *systemy testów* (Test battery) są to większe ilości testów usystematyzowanych i ułożonych z pewnego określonego punktu widzenia dla pewnego określonego celu tak, że zachodzi między nimi związek psychologiczny, np. testy do badania uzdolnień technicznych, zdadności do zawodu telefonistki i t. p.

Jeżeli serja testów równocześnie jest ustopniowana dla rozmaitych stopni wieku albo rozmaitych stopni zdolności, to stanowi ona system szczeblowy (skala), jak np. metoda Binet-Simona do badania inteligencji dzieci. Aby szereg badań wykonać możliwie ekonomicznie (w najkrótszym czasie i w najprostszy sposób), łączy się rozmaite testy w t. zw. „Omnibus - testy”. Wartości poszczególnych testów „Omnibus testów” ściąga się w jedną ogólną notę. Przykładem takich „Omnibusów” są „Army-Tests”, „National Intelligence Tests” i wogóle znaczna ilość testów amerykańskich do badania inteligencji. Wśród „omnibus-testów” rozróżnia się testy koliste i spiralne (cycle omnibus test, spiral omnibus test); pierwszy rodzaj

składa się z szeregu testów równie trudnych, w drugim stopniują się trudności w regularnych odstępach¹⁾.

Rozróżnia się ponadto testy *jakościowe* i *ilościowe*. Jeżeli się bada jedną czy więcej osób pod względem ich właściwości (pamięć, dar kombinacji, zmysł techniczny i t. d.), to kontroluje się rozmaite *jakości*; jeżeli natomiast bada się dwie osoby, aby przekonać się, która z nich ma *lepszą* pamięć, albo pojemniejszą uwagę, to chodzi nam o test *ilościowego* rodzaju.

Tam, gdzie mamy do czynienia nie z dorosłymi, lecz z młodocianymi, (jak to jest zawsze z terminatorami), testy ilościowe mogą dostarczyć wiadomości także o stopniu rozwoju zdolności w pewnym określonym wieku. Są to wtedy testy *rozwaju* czyli *testy wieku* (tests d'âge, tests de développement).

Zapomocą testów bada się 1) wiadomości²⁾ np. wiadomości szkolne i zawodowe (t. zw. Schulleistungsteste, Scholastic test); *zdolności* i *właściwości* psycho-fizyczne (inteligencji, charakteru, aktywności — Intelligence - tests, Charakter - tests), 3) *common sense* — zdrowy rozsądek

¹⁾ Fréyd M. Measurment in Vocational Seletion JPeRes 2 (6) 1923, str. 249.

²⁾ Istnieje rozbieżność zdań co do zagadnienia, w jakim stopniu testy mogą badać wiadomości szkolne i zawodowe. Meumann uważał badanie wyczynów szkolnych za jedno z zadań testów (Vorlesungen zur Einführung in die experimentelle Pädagogik, Leipzig 1913 T. 2 N. 324). Tęgo samego mniemania jest Cyryl Burt. W. Stern natomiast chętnie wyłączałby wiadomości szkolne z zakresu celu badań, ponieważ test służy przede wszystkim do badania zdolności. Wiadomości i zdolności nie wolno z sobą mieszać (Intelligenzprüfung der Kinder und Jugendlichen Rozdz. VI). Rozwój metodyki testów w Ameryce przyznał słusność pierwszemu stanowisku.

ludzki (badania, używane często przez Amerykanów): np. Dlaczego buty sporządza się ze skóry? — Co pan robi, jeżeli Pański ładunek węgla wpadnie w glinę? 4) *wyczyny i sprawności*. Zapomocą testów dowiadujemy się o stopniu wykształcenia technicznego, doświadczenia, ćwiczenia (t. zw. tests for technique).

Należy tutaj podkreślić ogólny charakter badania testami: test może zbadać, czego człowiek *nie może* dokonać; nie może on jednak zbadać, czego człowiek jeszcze może dokonać. To ograniczenie jest z jednej strony siłą, z drugiej strony słabością badania testami.

Jeżeli badania testami wykonywane są w specjalnym celu, aby rozpoznać duchową właściwość badanego, to testy, służące takiemu badaniu inteligencji nazywa się w Niemczech „*Intelligenztests*” w Ameryce „*mental tests*”. Jeżeli jednak testów używa się w tym celu, aby stwierdzić zdatność do pewnego określonego zawodu, to testy te nazywa się testami zdatności, albo *testami zawodowymi*. (Były nawet próby, aby odróżnić test, służący do wybierania dobrego robotnika z całego szeregu kandydatów — od testu, który ma udzielić wiadomości o zdatności do zawodu (Tests de sélection et Tests d'orientation) chociaż jest to raczej teoretyczna subtelność, a nie rozróżnienie rzeczowe. Testy używane w szkołach amerykańskich do badania wyczynów zbiorowych w rozmaitych przedmiotach i sprawnościach nazywają się „*Achievement-Tests*” — testy wyczynów zbiorowych).

Test inteligencji można również użytkować jako test zawodowy np. badanie sprawności zmysłów, lecz naodwrot nigdy test zawodowy nie może stać się testem inteligencji. W Ameryce odróżnia się przy testach zawodowych także t. zw. *Trade tests*, które odnoszą się do pewnych prac rze-

mieślnicznych i kilku zajęć dokonywanych według stałego schematu, jak stenografia, pisanie na maszynie.

Przy organizowaniu armji amerykańskiej podczas wojny wybierano żołnierzy do rozmaitych prac jak mechanicy, piekarze, kucharze, stajenni, telegrafiści. Testy które służyły do tej selekcji, nazwano Trade-tests. Są to według definicji Kornhausera i Kingsbury'ego testy, które mierzą nietylko zdolność do szkolenia się w zawodzie, ale zarazem i zdobytą sprawność i wiadomości, które zawiera zawód. (The trade-tests is a test designed to measure not merely the aptitude for acquiring a trade, but the actually learned skill and knowledge involved in that trade¹).

Testy zawodowe mogą być *analityczne*, jeżeli idzie o to, by zbadać poszczególne potrzebne do zawodu właściwości i zdolności (np. rozróżnianie barw, zręczność rąk i t. d.) albo *syntetyczne*, jeżeli test ułożony jest w ten sposób, że w pewnym ogólnym akcie bada cały szereg właściwości. Tak Kronfeld zbudował do badania lotników aparat, zapomocą którego bada się równocześnie poszczególne zdolności, konieczne dla lotnika.

Ponadto testy zawodowe (zarówno jak i testy inteligencji mogą być: 1) *abstrakcyjne* jeżeli test dla zbadania kompletnej wydolności zawodowej mierzy proste funkcje duchowe w tem założeniu, że między skomplikowanemi a elementarnemi funkcjami duchowemi zachodzą pewne określone stosunki. Podobnym przykładem jest próba Münsterberga, by zbadać przytomność umysłu oficerów okrętowych w niebezpieczeństwie zapomocą testu, w którym należało 24 kartonów z wydrukowanemi litera-

¹) Psychological Tests in Business, Chicago 1924, str. 23.

mi rozdzielić jak najszybciej na cztery serie, według częstości występujących na nich samogłosek. 2) Testy mogą odtwarzać rzeczywistość t. zw. Wirklichkeitstests). Istnieje kilka ich rodzajów: a) test może przedstawić pracę zawodową w zmniejszonym rozmiarze (t. zw. „replicas” albo „miniature job tests”). Lipmann naprz. zużytkował do badania drukarzy kasztę zecerską, zawierającą tylko 9-liter); b) Test może być wiernym powtórzeniem przebiegu pracy zawodowej (test posiada cechy „pełnej żywkości”) kiedy np. zdolność równoczesnej pracy oburącz bada się na suppcie (t. zw. testy na modelach, „sample Tests”). c) Test może dawać schematyczne przedstawienie rzeczywistych stosunków (t. zw. „analogus” Tests) jak np. w próbie Münsterberga, by motorniczych badać na urządzeniu, które posiada schematyczne podobieństwo z czynnością motorniczego na przedniej platformie. Münsterberg mówi o tym rodzaju badania: „Im bardziej schematyczna jest zewnętrzna strona aparatu, im więcej z niej wyłączymy wszelki pierwiastek asocjatywny, w tem czystszej postaci da się poznać sam wyczyn²⁾, ³⁾).

Testy te nazywa się niekiedy niesłusznie testami „symbolicznymi”.

W badaniach używa się także (wprawdzie bardzo rzadko) 1) testów *trickowych*, których rozwiązywanie wy-

¹⁾ Meyers Ch. S. Industrial Psychology in Great Britain. 1926, str. 114.

²⁾ Münsterberg: Psychologie und Wirtschaftsleben, str. 48.

³⁾ Należy zaznaczyć, że głównie w Ameryce mówi się o testach „życiowych” i „schematycznych”. W Europie określa się je krótko jako „badania”, ponieważ brak im zasadniczej cechy testu: prostoty.

mała pewnego *sztucznego podejścia* np. w teście do badania kombinacyjności technicznej, w którym chodzi o otworzenie lub rozłożenie na części skrzyneczki; 2) *testów „z tuszem”*... jeżeli w szeregu regularnych prób reakcji, wprowadza się nagle podniecie silną, wprowadzie zapowiedzianą wcześniej, ale nieoczekiwaną¹⁾; *pseudotestów*, które pozornie służą celom innym, niż te które zapowiedział badający. W ten sposób Giese dał test, służący pozornie do kontroli uwagi, który jednak w rzeczywistości miał zbadać inklinację erotyczną²⁾.

Testy równoważne t. j. takie, których celem jest stwierdzenie dokładnie tych samych zdolności lub wyczynów, a różnią się jedynie formą zewnętrzną, nazywają się „równoległemi”. Używa się ich w badaniach powtarzających się często (w szkołach i zakładach przemysłowych), ażeby wykluczyć znajomość rozwiązania testów u kandydatów.

Badanie testami przeprowadza się, albo na pojedynczych osobach (t. zw. *badanie indywidualne*), albo równocześnie na określonej grupie osób (klasie szkolnej lub pewnej liczbie starających się o posadę — t. zw. *badanie grupowe*), albo także na całej masie osób równocześnie (t. zw. *badanie zbiorowe*; w amerykańskich badaniach wojskowych badano równocześnie po 500 osób). Jeżeli grupę osób bada się pojedynczo, badanie takie nazywa się *kolektywnem*³⁾. O ile przy badaniach pojedynczych można zażyt-

¹⁾ Piorkowski: *Psychologische Methodologie* etc. I wyd. str. 30.

²⁾ Badanie inklinacji erotycznej w „Atti de la 3 Conferenza Internazionale di Psichotechnica applicata” Milano 1923, str. 138.

³⁾ Niektórzy psychologowie nazywają jednak również badanie grupowe, badaniem kolektywnym, nie używają natomiast żadnej nazwy dla pojedynczego badania określonej grupy osób.

kowywać wszystkie rodzaje testów (nieme, słowne, piśmienne, aparatowe) — do badania grupowego i zbiorowego nadają się najlepiej testy piśmienne.

Badania indywidualne podejmuje się tam, gdzie chodzi o dokładniejsze i staranniejsze zbadanie osoby; zabierają one ze zrozumiałych powodów wiele czasu i wysiłków, czasem po kilka godzin w ciągu kilku dni z kolei, podczas gdy badaniem zbiorowym jednego dnia zbadać można równocześnie większą liczbę ludzi. (Amerykańska metoda badań w armji wymagała 45minut czasu dla grupy). Od celu i zadań badania zależy więc, którą metodę zastosujemy.

Należy tu jeszcze stwierdzić wielką różnicę, zachodzącą między badaniami jednostkowymi a zbiorowymi. Z obu tych rodzajów badania każdy ma swoje zalety praktyczne i wady naukowe.

Badanie zbiorowe zezwala zbadać równocześnie wielką liczbę ludzi tym samym testem, przez co zaoszczędza się wiele czasu, a poza tem stwarza się dla wszystkich badań te same obiektywne warunki: dokładnie tę samą porę dnia badania dla wszystkich kandydatów w przeciwieństwie do badań indywidualnych, gdy bada się kolejno, a więc jednego rano, drugiego w południe, kiedy już jest zmęczony i t. d. Badający zwraca się *w ten sam sposób* do wszystkich kandydatów, nie zajmuje więc zupełnie jakiegóś życzliwszego stanowiska wobec jednych niż wobec drugich, albowiem badanie masowe nie pozwala na odróżnianie osób poszczególnych a przez to samo na występowanie uczuć sympatji czy antypatji. Ale takie badanie masowe, prócz wymienionych zalet, posiada i wielkie braki, które prowadzić mogą do fałszywego oceniania zadań. A mianowicie w badaniach masowych *nie można stwier-*

dzić czasu, w którym badany rozwiązuje zadanie, co jednak dla oceny wyczynu ma niezmiernie znaczenie. Ponadto, dystans który występuje w badaniach zbiorowych między kierownikiem a badanym, może wpłynąć na ujmowanie kandydata, jako czegoś nieosobowego prawie, jakiegoś abstractum, tak, że staje się niemożliwy dokładniejszy wgląd w cały proces rozwiązywania zadań, a osądza się tylko rezultaty końcowe.

3. WARTOŚĆ TESTÓW.

A. *Symptomatyczna i diagnostyczna wartość testów.*

Zadanie testu polega na postawieniu *diagnozy*, czy i w jakim stopniu osobnik badany posiada pewną zdolność. W badaniu zdolności do zawodu chodzi w takiej diagnozie zarazem o stwierdzenie, czy osobnik badany także w przyszłości dzięki swym zdolnościom okaże się dzielny t. j. chodzi również o *prognozę co do jego wydolności*.

Ważne tedy jest pytanie, o ile na podstawie tak krótkiego, prostego badania można wyciągać wnioski tak bardzo doniosłe. Czy test rzeczywiście może stanowić możliwie jednoznaczną wskazówkę co do poszukiwanej właściwości psychicznej badanego kandydata? Czy na podstawie zdolności do szybkiego sortowania kart można wyciągać wnioski o zdolności do szybkiego decydowania się w chwili niebezpieczeństwa? Czy proste, krótkie badanie posiada rzeczywiście taką *wartość symptomatyczną*?

Pytanie to jest zaprawdę najsłabszym punktem całego badania testami, gdyż od odpowiedzi na nie zależy wogóle prawomocność testów. Odpowiadano na nie w rozmaity sposób.

Są psychologowie i psychiatrzy, którzy twierdzą, że każda, nawet byle niepozorna czynność czy odpowiedź ma znaczenie symptomatyczne dla danego człowieka. Bleuler mówi: „W pewnem znaczeniu każdy proces psychiczny, każdy ruch wykonany w taki sposób, w jaki właśnie się odbywa, możliwy jest tylko u pewnego określonego człowieka o określonej przeszłości. Każda pojedyncza czynność reprezentuje całego człowieka¹⁾. Stosownie do tego każda odpowiedź, względnie każde rozwiązanie testu może dać wyjaśnienie w sprawie badanego osobnika”. Jeżeli uznamy za prawdziwy ten sąd ogólny, to musimy jednak przyjąć jako założenie, że odpowiedzi i czynności wyjaśnia się poprawnie t. j. że *umie* się poprawnie je wyjaśniać. I tu musimy wobec dzisiejszego stanu naszej wiedzy o ludziach, otwarcie przyznać się do tego, że nasza sztuka wyjaśniania jest jeszcze bardzo prymitywna. Symptomatologia naukowa pozostaje jeszcze, ogólnie biorąc, w zawiązkach, szczególnie w psychologii, gdzie symptomy (mowa, pismo, ruchy mięśni) są natury funkcjonalnej i dlatego zależą od warunków przypadkowych (stan usposobienia i zdrowia). Właśnie w psychologii dalecy jesteśmy od tego, by móc jakiś zewnętrzny wyraz chwilowy złączyć z pewną chroniczną dyspozycją psychiczną. Nie łatwem jest wyjaśnić jakiś zewnętrzny znak, czy wyraz t. j. rozpoznać jego znaczenie symptomatologiczne, a jeszcze trudniej jest z tych samych powodów stworzyć symptom funkcji psychicznej, zdolności, którym właśnie powinien być test. W jaki sposób wogóle znajdujemy test? Psychologowie nie wypowie-

¹⁾ E. Bleuler: Ueber die Bedeutung von Assoziationsversuchen. W „Diagnostische Assoziationsstudien“ wydane przez C. G. Jung'a, Leipzig, Barth 1911, str. 4.

dzieli się prawie zupełnie w tej sprawie. Według skąpych wiadomości sądząc, wydaje się, że nie tworzy się go na podstawie intuicji, nieświadomego natchnienia, lecz raczej rozważania prowadzą do utworzenia testu. U Münsterberga znajdujemy w tej sprawie bardzo znamienne miejsce. Zadaniem jego było ustalić zapomocą testu pewność i szybkość decyzji w chwili niebezpieczeństwa; musiał więc test wynaleść. W tym celu przeprowadził on cały szereg rozumowań. „Wyszedłem mówi on — z przekonania, że tak skomplikowany akt decyzji musi stać w określonych stosunkach do szeregu prostszych funkcji duchowych i że zmierzenie według zrutowanej metody tych prostych procesów, jak np. procesu rozróżniania, procesu asocjacji, procesu sugestji i t. d. wystarczy, by wyciągnąć wnioski, dotyczące skomplikowanego procesu decyzji”¹⁾ Utało się już takie postępowanie, że wynajdując test, szuka się stosunków między złożonymi a prostymi funkcjami. Bardzo często są to tylko analogje, a nawet powierzchowne podobieństwa i według nich konstruuje się testy”.

Kiedy dla uproszczenia badania pamięci badano ją zapomocą powtarzania szeregu bezsensownych zgłosek, uzasadniało się to śmiało przypuszczeniem, że pomiędzy zatrzymywaniem w pamięci wrażeń w życiu praktycznym, co w najważniejszym stopniu zawisło od naszego zainteresowania — a zapamiętaniem całkiem nieinteresujących zgłosek, przebiegają dość silne nici. Chciano w ten sposób na podstawie wykuwania na pamięć wysnuwać wnioski o zdolności do zatrzymywania w umyśle wyobrażeń rzeczy ze świata otaczającego.

¹⁾ Münsterberg: Psychologie und Wirtschaftsleben, Leipzig, str. 56.

Co uprawnia do tego śmiałego twierdzenia? Obserwacja odkryła lub stwierdziła stosunki tego rodzaju co najwyżej jako przypadkowe, a następnie spróbowano stwarzać je systematycznie, ażeby je wyzyskać dla celów praktycznych. Ta więc droga do wynalezienia testu przez hipotetyczne przypuszczenia, na których gruntuje się jego zastosowanie, pozwala wysunąć żądanie aby każdy test, choćby się nawet opierał na najbardziej podstawowych i szczegółowych rozważaniach, podlegał sprawdzeniu, czy rzeczywiście bada tę zdolność, którą ma badać, czy posiada on — innemi słowy — *wartość symptomatyczną*. A szczególnie w psychotechnicznych badaniach zdolności, w których chodzi nietylko o rozpoznanie zdolności, lecz także o postawienie prognozy i decydowanie o losie ludzkim, ma to jak najwyższe znaczenie, by *każdy* test był sprawdzony co do swojej wartości symptomatycznej.

Już Binet, któremu zawdzięczamy używanie testów w praktyce, słusznie zauważył: „un test mental n'a de valeur que si l'on sait ce qu'il signifie”¹⁾ zgadza się to z przekonaniem, wyrażonem przez L. L. Thurstone'a na Kongresie Oksfordzkim w r. 1923: „A test is never by itself good or bad. It is always good or bad with reference to the criterion, that the test is supposed to diagnose”²⁾. U Bineta znajdujemy jeszcze przestrożę, która ma ogromne znaczenie: Należy nietylko ten sam eksperyment kilka razy powtórzyć z tą samą osobą, ażeby dobrze zrozumieć to, co się czyni, ale należy także zanalizować kil'ka

¹⁾ De l'intelligence, Paris, Alcan, 1922, str. 304.

²⁾ L. Thurstone: The Principles of Vocational Guidance. Proceedings of the VII Intern. Congress of Psychology. Cambridge, Univers. Press 1925.

innych, nieco odmiennych eksperymentów, które mogą wyjaśnić tamte pierwsze. „Rezultat jednej metody eksperymentalnej jest często dwuznaczny, ponieważ zezwala na kilka wyjaśnień, między którymi trudno dobrze wybrać... Pewien rezultat można również dobrze wyjaśnić żywością myślenia, jak i brakiem zastanowienia, inny znowu zarówno zdolnością obserwacji jak i ubóstwem myślenia“¹⁾ Z tą przestrogą nie liczono się jednak prawie wcale. Ciężko grzeszyli dotychczas psychotechnicy w sprawie określenia symptomatycznej wartości testów. Należałoby raczej powiedzieć, że wogóle nie określano tej wartości zgóry, lecz stosowano testy według swego widzimisie i oczekiwano od praktyki dowodów ich dobroci. W ten sposób doszło do tego, że obecnie są w użyciu tysiące testów, co do których jeszcze ciągle brak dowodów, że rzeczywiście badają te zdolności, które rzekomo mają badać. Whipple mówił już w r. 1910: *We too often do not know what we are measuring*²⁾ — i niewiele zmieniło się od tego czasu. Nawet i te testy, którym zawdzięczamy używanie testów, nie są wolne od tych braków, ponieważ i tam tylko rzadko wprowadzono psychologiczną analizę testu. Skutkiem tego zachodzi często taki wypadek, że musi się test, już wielokrotnie używany dodatkowo zbadać, aby się upewnić czy spełnia on swe zadania.

Tak zdarzyło mi się, gdy chciałam użyć raz bardzo rozpowszechnionego testu *Bourdon'a*. Przekonałam się przy sprawdzaniu testu, że dobre wykonanie próby osiąga się nie tylko przez zdolność koncentracji, ale także przez

¹⁾ Binet: *De l'intelligence*, str. 299.

²⁾ Whipple: *Manual of Mental and Physical Tests*, 1910, str. 3.

inne zdolności, czysto wzrokowej natury; wobec tego test ten nie bada dobrze koncentracji¹⁾. To samo skonstatawałam również w bardzo popularnym teście analogji, w którym okazało się, że *poprawne* jego rozwiązanie nie zawsze osiąga się przez poprawny bieg myśli, lecz że w teście tym trzeba zawsze żądać *uzasadnienia* odpowiedzi, by upewnić się, że rozwiązanie odpowiada również psychicznemu procesów i sądu, opartego na analogji²⁾.

Jeszcze w wyższym stopniu odnoszą się powyższe uwagi do testów aparatowych. Przyjęło się ogólne przekonanie, że testy aparatowe dają większą pewność w badaniach, niż testy nieaparatowe, ponieważ pozwalają dokładnie zmierzyć wyczyn. Ale już K ü l p e zauważył całkiem słusznie, że aparat może być pod względem technicznym skonstruowany bez zarzutu, a mimo to rezultaty, za pomocą niego osiągnięte, mogą zupełnie nie dawać dokładnego obrazu badanej właściwości. Robimy tu mianowicie ten błąd, że pojęcie dokładności, które cechuje aparaty fizykalne, przenosimy na badania psychologiczne. Tymczasem w rzeczywistości jedno z drugim niewiele ma wspólnego. Weźmy np. aparaty, których celem jest badanie wrażeń stawowych (t. zw. Gelenkprüfer): Badany winen najpierw korbę aparatu kręcić tak długo, aż napotka pewien określony opór. Potem kręci po raz wtóry korbę, póki nie uzyska tego samego oporu. Dokładność każdorazowego położenia korby kontroluje się zapomocą skali. Ale kiedy tę próbę przeprowadziłam kilka razy na sobie samej, nie mogłam, mimo intensywnej samoobserwacji

¹⁾ Baumgarten: Eine Konzentrationsprobe. PraktPs. 2, (11) 1921.

²⁾ Baumgarten F. Der Analogietest. PraktPs. 4. (7) 1923.

odczuć żadnych napięć stawów, ścięgien i mięśni, natomiast moja mniej lub więcej poprawna odpowiedź polegała wyłącznie na pamięci czasu. Stale miałam wrażenie: „tak długo trwało to po raz pierwszy”. Ponieważ nie przeprowadzono żadnych badań i nie stwierdzono, który z unerwionych stawów działa głównie podczas kręcenia korbą, więc nie wie się także w gruncie rzeczy, co właściwie badało się zapomocą aparatu. W każdym razie badano zapomocą niego, jak zapomocą większości aparatów, nie zdolność, lecz pewien całkiem określony wyczyn.

Dotychczas nie mamy prawie zupełnie psychologicznej analizy testów. I dlatego należy tu położyć jak największy nacisk na twierdzenie, że nie dojdziemy do *naukowo* ugruntowanej teorii testów, dopóki nie zanalizujemy każdego testu, używanego w praktyce. Analizę należałoby właściwie przeprowadzić w ten sposób: psychotechnik stosuje każdy nowy test do siebie i przez samoobserwację orientuje się, jakich funkcji psychicznych wymaga rozwiązanie testu wzgl. jakie zjawiska psychiczne zachodzą w nim podczas rozwiązywania; w ten sposób na sobie samym bada cały proces psychiczny, zachodzący podczas rozwiązywania zadania; następnie musi od badanych, na których przeprowadza badania próbne, zebrać wypowiedzenia się o tem, w jaki sposób oni wykonali test, aby z jednej strony sprawdzić przebiegi psychiczne u siebie i innych, a z drugiej, aby stwierdzić rozmaite rodzaje sposobów rozwiązywania; dopiero na podstawie tego materiału psychologicznego może wydać sąd o symptomatycznej wartości testu.

Symptomatyczna wartość testu jest tak bardzo związana z jego wartością diagnostyczną, że prawie zawsze uważa się je za synonimy. Ale to nie jest słuszne. Symp-

tom jest to *znak* pewnej właściwości lub zdolności; diagnoza testowa jest to ilościowe i jakościowe wypowiedzenie się o zdolności: Wartość symptomatyczna określa, czy test rzeczywiście bada pewną zdolność, wartość diagnostyczna — w jakim stopniu i w jakim zakresie to spełnia. Wartość symptomatyczna jest przeto jedną z podstaw dla diagnozy testowej i to podstawą najważniejszą. Należy również ściśle rozróżnić wartość diagnostyczną od wartości prognostycznej, albowiem test może dobrze badać właściwości, istniejące dotychczas, a nie potrafi nic powiedzieć o ich przyszłym rozwoju t. zn. powiedzieć, czy terminator będzie dobrym robotnikiem, czy robotnik, dziś uznany za niezdolnego, zawsze niezdolnym pozostanie. W dotychczasowej praktyce nie utrzymuje się rozróżnienie tych trzech wartości, co wyszło na jaw przy sprawdzaniu testów czyli t. zw. kontroli, jak to zaraz zobaczymy. Niedostateczne wstępne sprawdzanie psychologiczne testów zastępuje się metodą, posługującą się wynikami badań testowych. Mianowicie skoro obrało się dla pewnego celu pewien test — czyto na podstawie rozważenia czy na podstawie analogji — przeprowadza się próby ustalenia jego wartości praktycznej przez następczą „kontrolę”. Używa się rozmaitych sposobów takiej kontroli.

Spearman pierwszy przeprowadził taką kontrolę przez swoje *porównania testów z testami*. Stosował te same testy (testy do badania inteligencji) do tych samych osób i stwierdził zapomocą obliczeń korelacji, które z uwzględnionych testów posiadają wyższy wskaźnik korelacji. Testy takie z wyższym wskaźnikiem korelacji (nazywa je „principal tests”) są, jego zdaniem, wyrazem ogólnej zdolności psychicznej (nazywa ją „ogólną inteligencją”), a stąd mogą one dać wyobrażenie o inteligencji danego osobnika, posiadają wartość reprezentacyjną.

Te porównania jednych testów z drugimi t. j. ustalanie jaka korelacja zachodzi między szeregami rang, nadanych wynikom różnych testów, jest jednak metodą rzadko używaną do kontrolowania wartości testów wzgl. do stwierdzania, czy wyniki testów sprawdzają się.

Pewna odmiana tej metody polega na tem, że nowy test, o którego wartości chce się wyciągnąć wnioski, korelujemy z testem o uznanej wysokiej wartości. Jeżeli przytem stwierdzi się wysoką korelację, to uznaje się nowy test również za pełnowartościowy. W Ameryce, gdzie w czasie wojny poddano kontroli wielką ilość testów, jak np. Army Alpha Intelligence Test, nowy test bada się zwykle przez korelowanie z Army Test. L. L. Thurstone, który posiada wielkie doświadczenie w przeprowadzeniu testów, ostrzega wyraźnie przed takim postępowaniem: The fact, that a new test correlates well with some test which in turn correlates well with a certain criterion does not prove that the new test can be used for diagnosing that criterion¹⁾.

Ostrzeżenie to jest usprawiedliwione, albowiem wobec niedostatecznej znajomości struktury psychicznej można przeoczyć, że wysoka korelacja dwóch właściwości, badanych przez test, zależy od przypadkowych czynników. Często również wypowiada się przekonanie (i postępuje się stosownie do niego), że kontroli wymagają tylko te testy, których używa się po raz pierwszy. Natomiast testy, których już używano dla pewnych celów, a obecnie ma się zastosować dla innych nie potrzebują tej kontroli. Tak np. w Niemczech stosowali test kombinacji (ułożyć z trzech wyrazów możliwie najwięcej zdań) bez żadnych zmian:

¹⁾ L. L. Thurstone (zob. wyżej).

Klutke w badaniu telefonistek, Moll i Piorkowski w badaniu pracowników więziennych zarówno jak i w badaniu berlińskich 12 — 13 letnich dzieci szkolnych, jakkolwiek inne wymagania trzeba postawić co do inteligencji młodocianych, inne co do inteligencji urzędników więziennych. Nie można się tedy zgodzić na podobne poglądy. Test może się zupełnie dobrze nadawać do jednego określonego badania, natomiast zawieść zupełnie przy badaniu pozornie pokrewnej funkcji; np. *test do badania uwagi* nie nadaje się jako test do badania pamięci, jakkolwiek uwaga i pamięć są funkcjami pokrewnymi. Ponadto, ponieważ testy zawodowe w większej części badają wyczyny a nie zdolności, zdarza się często, jak to udowodnił Muscio, na zdolnościach motorycznych, że dobre rozwiązanie testu sprawności motorycznych nie daje żadnego dowodu ani żadnej rękojmi, że rozwiązany również zostanie inny test z tej samej dziedziny¹⁾.

Ponadto tenże sam test może służyć do badania całkiem rozmaitych zdolności: Tapping-test może służyć do badania zręczności, a także do badania przebiegów woli i t. d. Przewidywanie, do czego można używać pewnego testu, może opierać się mniej lub więcej na prawdopodobieństwie; ale zawsze, jeżeli chce się danych pozytywnych, musi się test zbadać pod względem jego symptomatycznej i diagnostycznej wartości.

Druga metoda stwierdzenia wartości symptomatycznej testu opiera się również na zasadzie porównania; lecz nie porównywa się tutaj jednego testu z drugim, *lecz test z wyczynem w praktyce*. Np. daje się pewien test do roz-

¹⁾ Muscio: Motor capacity with special reference to vocational guidance. Brit. Journ. of Psychology 13 (2), 1923.

wiązania pewnej liczbie robotników w jakimś zakładzie przemysłowym. Jeżeli dobrze rozwiązują go robotnicy, których dobre wyczyny już są znane, a źle robotnicy źli, uważamy go za test, zapomocą którego można stawiać diagnozę co do określonej zdolności zawodowej. Ponieważ w zakładzie, gdzie jest większa liczba robotników, podział na „dobrych” i „złych” jako zbyt ogólny nie zadowala, a konieczne jest stopniowanie wyczynów według jakości, więc postępuje się w sposób następujący: wermistrz lub przełożony szereguje pracowników według ich wyczynów w szereg rang, psychotechnik układa ich w ten sam sposób na podstawie badania, a obydwie listy, sporządzone od siebie niezależnie, porównywa się ze sobą. Jeżeli w większości wypadków wykazują zgodność, uważa się test za „dobry”.

Kierownicy badań tacy jak np. Stern, Schackwitz, Kornhauser i in. zauważyli niejednokrotnie, że robotnicy, zatrudnieni w warsztatach przez czas dłuższy, często zawodzili podczas badań. Wyjaśniono to w ten sposób, że robotnicy długo praktykujący zautomatyzowali swoje ruchy, a w sztucznej lub nawet całkiem nowej sytuacji czuli się źle¹⁾. Również Cimbali²⁾ podaje podobne wiadomości o doświadczeniach, dokonanych przez siebie

¹⁾ Kornhauser i Kingsbury podają jeszcze inne wyjaśnienie, a mianowicie, że inteligentniejsi opuszczają zawód, a nie-inteligentni pozostają w zawodzie; ci naturalnie rozwiązują test źle (op. cit. zob. str. 134). To jednak stosują się tylko do warunków amerykańskich z ich silną wymianą robotników. Giese twierdzi, że gorliwość kandydatów do zawodu stanowi główny warunek ich dobrego rozwiązywania testów.

²⁾ W. Cimbali: Die Erziehung zur Tüchtigkeit in Schule und Beruf. Hamburg 1919, str. 40.

na pracownikach w „wyższych” zawodach. Psychjatra, znany ze swoich doświadczeń klinicznych oraz z dziedziny anatomji mózgu, które stanowią dowód jego zdolności obserwacyjnej, wykazywał przy aparacie do badania zdolności ujmowania wielkie wahania obserwacji: 30/1000 — $\frac{1}{2}$ sekundy. Inny znany lekarz i operator niestrudzony w swej pracy zawodowej wykazał przy tym samym aparacie: powolność, umiarkowaną pewność i prędkie znięczenie. Badanie psychologiczne właściwości, uznanej za konieczną w zawodzie, nie dopisywało u bardzo dobrych praktyków. Takie fakty, choć obecnie jeszcze niezbyt szeroko znane (prawdopodobnie badający nie zauważyli ich), ostrzegają jednak, że trzeba być ostrożnym przy sprawdzaniu testów w powyższy sposób. Z drugiej strony, wobec tego, że robotnicy, zajęci w warsztacie przez ćwiczenie nabyli już pewnej sprawności, przeto z ich rozwiązywania testów nie można bez zastrzeżeń wnioskować, jaki w tem udział mają zdolności, a jakie nabyte sprawności. Ponieważ test chce badać tylko zdolności, więc wyczyn zawsze da o niej tylko niedokładne wyobrażenie, i nie będzie przeto oznaką zdolności.

Ważniejsza jeszcze jest okoliczność ta, że, gdy chodzi o sprawdzenie poszczególnych testów (a właśnie takie sprawdzanie mamy teraz na względzie), to majster czy przełożony bardzo rzadko potrafi wydać słuszny sąd o poszczególnych zdolnościach. Jeżeli zdoła wydać trafny sąd o zręczności albo o zmyślności technicznej, to trudność sprawi mu ocena jakości i rodzaju pamięci uczniów. Na jego „znajomości ludzi” w takim wypadku polegać nie można¹⁾.

¹⁾ O tem bardziej wyczerpująco w części II, rozdz. II-e.

Inna odmiana kontroli testów jest często używana w ostatnim czasie. Bada się kandydatów (lub terminatorów do pewnego zawodu) zapomocą pewnych testów i oddaje się do zawodu. Jeżeli w praktyce okażą się dobrymi albo szybko wyuczą się (co uważa się za objaw zdolności) ci, którzy, dobrze rozwiązyali test, to jest to dowodem poprawnego doboru testu; czyli test taki posiada wartość symptomatyczną.

Metoda ta ma tę wadę, że nie przyjmuje się do zawodu kandydatów, którzy testu nie rozwiązyali; niema przeto przeciwdowodu, że ci, właśnie nie okażą się dobrymi w praktyce. Ponadto, ponieważ używało się więcej testów, trzeba wpierrw starannie rozróżnić, które z nich właściwie posiadają wartość symptomatyczną, co wymaga uciążliwego obliczania korelacyj.

Jakkolwiek nie można pominąć *porównywan testów z praktyką*, jako że ostatecznie test winien stawiać prognozę wyczynów w praktyce, to jednak powinno się pamiętać o powyższych zastrzeżeniach, żeby zbytnio nie przeceniać wartości tej metody.

Wspomina jeszcze Claparède¹⁾, że metoda, którą można nazwać metodą monograficzną, może również dać podstawę do wniosków o diagnostycznej wartości testu: polega ona na obserwacji pewnego osobnika w ciągu całego jego życia., odnośnie do zdolności, które wykazuje jak i odnośnie do prac, które z dobrym wynikiem wykonywa. Jeżeli się tę osobę podda badaniu zapomocą szeregu testów, to można te testy, które rozwiązuje ona dobrze, uznać za „symptomatycznie” dobre. Metody tej nie za-

¹⁾ Claparède: Comment diagnostiquer les aptitudes chez les écoliers. Paris. 1924, str. 8.

słusowano jeszcze dotychczas nigdy, ale metodologicznie byłoby korzystne wejść na tę drogę, aby stwierdzić, w których testach dobrzy praktycy zaznaczają się jako dobrzy

B) *Zdolność różnicująca testów.*

(Rozsianie i wycechowanie testu).

Ustalenie symptomatycznej wartości testu stanowi tylko jedną stronę, coprawda, najważniejszą, jego praktycznej używalności; albowiem chodzi tu tylko o zagadnienie, co bada ten test. Jest jednak i drugie ważne zagadnienie, *jak* bada test t. zn. czy nadaje się on także jako *instrument do pomiaru*.

Każdy test musi dawać wyniki zróżnicowane, czy to jakościowego, czy ilościowego rodzaju, albowiem jedno z zadań badania zapomocą testu, polega na tem, by stwierdzić różnice wśród badanych. Im bardziej przeto zróżnicowane są wyniki badania testem, czyli, jak się to nazywa — im większe jest rozsianie, tem wyższa jest diagnostyczna i praktyczna wartość testu. Test „rozsięwa” dobrze, jeżeli jego wyniki badania dadzą się przedstawić zapomocą krzywej częstości (G a u s s a: krzywa dyspersji¹⁾). Wykreślenie krzywej G a u s s a na podstawie badania testami większej liczby osób jest przeto środkiem do stwierdzenia wartości rozsiania testu.

Zamiast sporządzania krzywej rozsiania zadowala się często w testach, przeznaczonych dla określonego wieku, ustaleniem liczby procentowej badanych, którzy w badaniach wstępnych test rozwiązali. Jeżeli jest to 75%, test uważa się za nadający się do użytku²⁾.

¹⁾ Bliższe szczegóły powyżej zob. str.

²⁾ Liczba ta została uzyskana empirycznie. Po raz pierwszy

Zróznicowanie wyników testu byłoby jednak bezcelowe, gdyby ich nie można było odpowiednio zinterpretować wzgl. ocenić. Cóż oznacza pewien określony wyczyn badanego np. zapamiętanie 3 obrazków z siedmiu małych obrazków geometrycznych, pokazanych badanemu w przeciągu 1 sekundy? Czy jest to już dobry, czy jeszcze zły wyczyn? O wartości jakiegoś rozwiązania testu nie możemy zyskać jasnego pojęcia, jeżeli nie wiemy, w jakim stosunku stoi ten wyczyn do wyczynów innych osób, zatem dopóki nie przeprowadzimy porównania. Stąd trzeba koniecznie *wycechować* test¹⁾ t. j. ustalić stałe oceny wszystkich możliwych jego rozwiązań. Podobnie, jak cechujemy instrumenty fizykalne, np. termometry, znacząc punkty wrzenia i marznięcia, a potem odległość całą dzielimy dowolnie na części (na 100 lub 80 stopni), podobnie możemy empirycznie ustalić przy większej grupie indywiduów punkt najniższy i najwyższy, najgorsze i najlepsze rozwiązanie pewnego testu, a resztę odpowiedzi ustopniować tak, by każda dowolną odpowiedź jakiejś osoby określała od razu stopień wyczynu. Wyniki badań ułożone według krzywej Gaussa określa się jako „dobre“, „złe“, „średnie“ nie subiektywnie i dowolnie, lecz te oceny wyłaniają się jako konieczność stosownie do częstości ich występowania.

Robertag, potem i Goddard stwierdzili przy użyciu systemu Bineta na młodzieży, że z grupy osobników 75% test rozwiązywało mniej więcej dobrze, podczas, gdy nie dorastało do odpowiedzi 25%.

1) Pojęcie „wycechowanie“ nie jest wśród psychotechników jednoznaczne. Jedni rozumieją przez nie „ustalenie wartości przeciętnych“, „ustalenie norm“, podczas, gdy inni (np. Giese) nazywają wycechowaniem wszechstronne zbadanie testu i jego używalności. Byłoby lepiej zatrzymać właściwe znaczenie słowa „wycechowanie“ i używać go tylko w pierwszym znaczeniu.

Rzut oka na krzywą poucza natychmiast, jak należy oceniać jakiś wyczyn. Cała metoda cechowania dokładniej omawiana będzie poniżej¹⁾; tu pragnę tylko zwrócić uwagę na konieczność takiego cechowania.

Ażeby móc stwierdzić rozsianie testu i dokonać wycelowania, trzeba test sprawdzić. Musi się test poddać badaniu „testowemu”. Musi się więc przeprowadzić jeszcze powtórna kontrolę. Trzeba przytem uwzględnić pewne kwestje metodologiczne:

1) Na jakiej liczbie osób trzeba test sprawdzić, aby go uznać za używalny? Niełatwo jest dać na to jakąś jedną odpowiedź. W Ameryce operuje się wielkimi liczbami od stu do kilku tysięcy, w Europie ogranicza się do znacznie mniejszych liczb. Pierwsze porównania badań terminatorów przemysłowych z wynikami w praktyce, które opublikował poraz pierwszy M o e d e, ograniczyły się do 20 uczniów²⁾. Należy to już do przeszłości, ale dotychczas jeszcze nie ustalono właściwie minimum. G i e s e mówi o statystykach prób w psychotechnice: „Jako miarę normalną zbadać należy conajmniej 200 osób, przyjmując za założenie populację homogeniczną”³⁾. W innym miejscu mówi: „Jeżeli się bada próbą pojedynczą, to po zbadaniu 500 osób ledwo możemy uzyskać pewność”. Żądanie to, choćby nawet było słuszne, nie da się jednak przeprowadzić skutkiem zewnętrznych warunków — poprostu nie zawsze rozporządzamy tak wielu osobnikami do zbadania. Ale 100 osobników wystarczy do badania wstępnego, które

¹⁾ Zob. rozdz. „Ustalanie norm”.

²⁾ M o e d e, Die Psychotechn. Eignungsprüfungen des industriellen Lehrlings, PrakPs 1 (3), str. 79.

³⁾ Handbuch psychotechn. Eignungsprüfungen, str. 647.

⁴⁾ Psychotechnisches Praktikum, str. 74.

dobrze zorientuje co do wartości testu. Pozostaje tylko warunek, by badający używał testu ze szczególniejszą starannością.

Wśród jakiej kategorii osób należy przeprowadzić badanie kontrolne? Ponieważ rezultaty testu wykazują wielkie różnice, zależnie od poziomu duchowego, wieku, środowiska badanego, stało się palącą potrzebą, ażeby badanie testem przeprowadzić na osobach z tych sfer, dla których jest on przeznaczony. A więc testy dla młodzieńców sprawdzać należy na młodzieży, testy dla robotników na robotnikach. Kornhauser i Kingsbury twierdzą, że nawet testy dla tego samego zawodu, ale dla rozmaitych zakładów przemysłowych, powinny być sprawdzane na robotnikach obydwóch zakładów, albowiem w każdym zakładzie przemysłowym wykonywa się pracę w inny sposób, a inne ukształtowanie pracy wymaga innych testów¹⁾.

4. OCENIANIE WYCZYNÓW UZYSKANYCH ZAPOMOCA TESTÓW.

A. *Ocenianie punktami.*

Rozwiązanie testu można bardzo łatwo oceniać w zadaniach alternatywnych. Jeżeli zadanie zostało rozwiązane, uważa się odpowiedź za *dobrą*, jeżeli nie zostało rozwiązane za *złą*. Amerykańskie Army Mental Tests, które w znacznej części składały się z testów alternatywnych, używały dwóch sposobów takiego oceniania. Pierwszy polegał na tem, że w jednym teście, który składał się z poszczególnych zadań alternatywnych, liczyło się tylko nie-

¹⁾ Psychological Test in Business, str. 64.

rozwiązane zadania, tak, że w ten sposób można było natychmiast stwierdzić liczbę dobrze rozwiązanych zadań. Techniczną nazwą takiego oceniania jest „Number Right” (Liczba poprawnych rozwiązań). Ale taki sposób oceniania faworyzował bardzo wysoki procent rozwiązań dobrych, żeby więc ten rezultat poprawić, posługiwano się innym sposobem: zestawieniem dobrych i złych rozwiązań i odejmowaniem drugich od pierwszych; jest to t. zw. ocenianie „Right minus Wrong”. (poprawne rozwiązania minus błędy¹⁾).

Przy testach niealternatywnych, na które można dawać rozmaite odpowiedzi, powstaje pytanie: jak oceniać taką odpowiedź, to znaczy jak uznać jedną odpowiedź za lepszą od drugiej, nie rozstrzygając przytem według subiektywnego widzimisie, lecz na podstawach obiektywnych. Jest to jedno z najważniejszych pytań w zakresie metodyki testów i poświęcono mu w kołach psychologów i psychotechników wiele uwagi. I my przeto potraktujemy je nieco obszerniej.

Przedewszystkiem trzeba rozróżnić dwa rodzaje testów: 1) testy, które pozwalają na ocenianie ilościowe wzgl. testy mierzące i 2) testy, które zezwalają tylko na ocenianie jakościowe. Zajmijmy się najpierw testami mierzącymi, a jako wstęp omówimy wogóle problem mierzenia w psychologii.

Te osoby, prowadzące badania, które są z zawodu inżynierami, pojmują mierzenie w psychologii jako mierzenie w zwykłym znaczeniu. Niema dość silnych słów, by ostrzec przed takim pojmowaniem sprawy. Treści psy-

¹⁾ C. S. Yoakum and M. Yerkes: *Army Mental Tests*, New York, Holtz, Co 1920, str. 7, 66.

chiczne (wrażenia, przedstawienia, uczucia i t. d.) nie da-
dzą się mierzyć w ten sposób, co fizyczne lub fizykalne
wielkości, które można wyrazić przez liczbę jednostek
tego samego rodzaju, a więc długość przez jednostki dłu-
gości, objętość przez jednostki objętości. Wrażenie, którego
nie można dokładnie reprodukowac, nie może służyć jako
miara do mierzenia wrażeń. To, co w tym kierunku osią-
gnięto w psychologii, jest (przy założeniu, że różnym czu-
ciom odpowiadają różne podniety) mierzeniem podnieć,
wywołujących wrażenia (podnieć wzrokowych, słuchowych,
cieplnych i t. d.) a w ten pośredni sposób określanie siły
wrażeń. Ponieważ nasze wrażenia zależą nietylko od pod-
nieć, lecz także od ogólnego stanu duchowego: zmęczenia,
zainteresowania i t. d. więc liczby wymiaru trzeba brać,
naturalnie, z wielkimi zastrzeżeniami. Jeżeli przeto mó-
wimy: „badamy” pewną zdolność (naprz. pamięć) to
w rzeczywistości nie badamy funkcji przypominania jako
treści świadomości, jako subiektywnego przeżycia danej
osoby, lecz jako odpowiedź na podnieć obiektywną wy-
wołaną przez nas (naprzykład powtórzyć szereg głosek).
przyczem robimy założenie, że czynna jest funkcja,
o którą nam chodzi. Nie możemy jednak twierdzić, że
przytem nie działają także inne funkcje (uwaga, wyobraź-
nia wzrokowa i słuchowa i in.) i nie wiemy jaką rolę one
odgrywają. Mamy jednak podstawy, by twierdzić, że im
bardziej złożone jest zadanie, tem większe grozi niebez-
pieczeństwo, że badana zdolność nie wystąpi wyraźnie.
W psychotechnicznych badaniach muszą być uwzględnione
dwa szczególne wypadki tego podstawowego faktu: jeżeli
dajemy kandydatowi zadania np. przepołowienie kreski,
i przepołowienie koła, a twierdzimy przytem (na podsta-
wie częstości rozwiązywania), że drugie zadanie jest trzy

razy trudniejsze od pierwszego, to nie wolno sobie wyobrażać, że psychicznie w przeżyciu badanego przy rozwiązywaniu drugiego zadania wystąpił trzy razy większy wysiłek, aniżeli przy rozwiązywaniu pierwszego. Wprawdzie możemy wyczuć, że w drugim wypadku zrobiliśmy wysiłek większy, niż w pierwszym. Czy jednak stosunek ten wyraża się jako dwu — pięcio — czy sześciokrotny, tego nie możemy skonstatować. Już F e c h n e r, twórca psychofizycznych metod mierzenia, podkreślał z całym naciskiem: „najpierw i bezpośrednio posiadamy tylko sąd, czy coś jest większe, mniejsze lub równe we wszystkich tego rodzaju stosunkach — nigdy jednak nie wiemy „ile razy“, co przecież jest koniecznie potrzebne do prawdziwego mierzenia“¹⁾. I chociaż zatrzymujemy z praktycznych względów takie ilościowe ocenianie, to jednak każdy praktyk musi zdawać sobie sprawę z faktycznego stanu rzeczy, ażeby nie trzymał się liczb niewolniczo.

Tosamo odnosi się i do drugiego wypadku. Gdy rozdzielamy dla poszczególnych wyczynów noty (cenzury 1, 2, 3, 4, 5) to arytmetycznie pozostaje między notami zawsze ten sam odstęp. Jeżeli natomiast stopnie zdolności psychicznych określamy notami 1, 2, 3, 4, 5 to jeszcze z tego nie wiemy, czy psychicznie nie są one równowartościowe, to jest: czy ten, który otrzymał notę 2 nie wykonał swego zadania z takim samym wysiłkiem i z tym samym udziałem swoich zdolności, jak ktoś inny, którego rozwiązanie zostało ocenione na 3 lub 4. Te same noty można uzyskać dzięki różnym środkom psychicznym, natomiast różne noty można uzyskać na podstawie tego samego stopnia wysiłku i zdolności. Dwaj badani, którzy uzyskali te

¹⁾ Elemente der Psychophysik T. I, str. 541.

same noty, mogą posiadać dwie różne konstytucje psychiczne. W psychotechnicznych badaniach, w których przecież chodzi tylko o wyczyn: czy i w jaki sposób został on wykonany, albo dlaczego się nie udał, nie uwzględnia się tego zupełnie. Ale psychotechnik powinien sobie zdawać sprawę z tego niepsychologicznego postępowania.

Zajmijmy się teraz ilościową oceną wyczynów, używanych zapomocą testów. Jeżeli się na przykład w grupie badań X. Y. Z. na przeciąg tego samego czasu da do wyuczenia się na pamięć 36 słów, i jeśli się okaże, że X wyuczył się 18, Y — 30, a Z 36 słów to *ilościowa* ocena tego wyczynu jest możliwa. Ocenia się na przykład zapamiętanie jednego wyrazu jako *jeden punkt*). Stosownie do tego X uzyskuje liczbę punktów 18, Y — 30, Z — 36, najgorszy wyczyn niezapamiętania ani jednego słowa otrzymuje 0, najlepszy wyczyn 36 punktów. Takie ocenianie jest całkiem dowolnem, przyjętem dla uproszczenia. Możnaaby z równym prawem operować dwoma, trzema punktami albo ułamkami. Próbowano uzasadnić ocenianie zapomocą punktów; wychodziło się przytem ze stwierdzenia faktu, że żadna grupa osób nie daje nigdy z jednej strony najlepszego, a z drugiej najgorszego rozwiązania testów, że wyczyny jej zbliżają się mniej lub więcej do wartości krańcowych w obydwu kierunkach. Rzadko kto przy próbie (zostanmy przy powyższym przykładzie) zapamiętywa wszystkie 36 słów, niema prawie takiego, ktoby nie zapamiętał ani jednego słowa. Najczęściej zdarza się, jako najgorszy wyczyn, zapamiętanie 3-ch słów i jako najlepszy wyczyn zapamiętanie 27 słów. Rozpiętość wyczynu wynosi

¹⁾ Ocenianie w punktach wprowadził poraz pierwszy do psychologii Yerkes.

zatem $27 - 3 = 24$ słowa. Ocenia się wyczyn najwyższy pewną określoną notą — ze względu na wygodę przyjmuje się notę 10, a każde zapamiętane słowo ocenia się w zakresie rozpiętości rozsiania jako $10/24 = 0,4$. W ten sposób liczby punktów zastępujemy liczbami, wyrażającymi wartość — notami. Liczby, wyrażające wartość, pozwalają wyraźniej ująć wyczyny osób wewnątrz pewnej grupy i przeprowadzić porównania. Nie należy jednak zapominać, że choć takie ocenianie wyczynów nie jest tak dowolne, jak w pierwszym wypadku, (ocenianie każdego wyrazu przez jeden punkt), jednak jest ono względne, a mianowicie uzyskane empirycznie na podstawie właściwości pewnej danej grupy.

Testy jednak niezawsze są takie, by można każdą dobrą odpowiedź ocenić jako jeden punkt. Są i takie, w których każda odpowiedź częściowa zamyka w sobie już 2 zadania. Tak np. w teście pamięciowym chodzi nie tylko o to, żeby zapamiętać poprawnie słowa, lecz żeby wymienić je w odpowiednim następstwie. A więc zupełnie dobra odpowiedź, to jest poprawne wypowiedzenie słowa w odpowiednim miejscu, powinna być oceniona dwoma punktami. Poprawna odpowiedź częściowa, na przykład właściwe powtórzenie słów, ale w niewłaściwym porządku, albo samo utrzymanie porządku słów, przedstawia wartość jednego punktu. Ocena testu musi się więc stosować do natury testu. Są także testy, których rozwiązanie czasem polega na stu odpowiedziach pojedynczych. Tak np. w teście uzupełniania luk badany musi uzupełnić poszczególnymi zgłoskami, albo nawet całymi wyrazami opuszczenia w teście. Należałoby więc dać 100 punktów za dobre jego rozwiązanie. Aby nie operować zbyt wielkimi liczbami przy-

muje się za jeden punkt 5 dobrych uzupełnień. 100 uzupełnień uzyskuje przeto liczbę 20 punktów.

Niezawsze ocenia się pojedyncze odpowiedzi w jednakowy sposób. Zdarza się, że test nie przedstawia jedno-
litego zadania; części jego są to łatwiejsze, to trudniejsze. Ażeby ocenę zrobić jaknajmniej skomplikowaną często przeocza się tę różnicę; a jednak przy dokładnych ocenach powinno się brać pod uwagę stan faktyczny. Np. test pamięci może zawierać szereg łatwych lub trudnych do zapamiętania wyrazów: deszcz, pogoda, zimno — lub Guadaluquivir, przeciwprostokątnia, kooperacja. Nierównym jest przeto wyczynem zapamiętać trzy pierwsze lub trzy drugie słowa. Uwzględniając ten stosunek, ocenia się szereg drugi jako 3 razy trudniejszy, a więc ocenia się pierwszy — trzema punktami, drugi jako $3 \times 3 = 9$ punktów. Takie postępowanie nazywa się wprowadzeniem do oceny „współczynników trudności”.

Aparat również może dostarczać zadań o rozmaitej trudności. Nie jest równoznacznem podzielić w optometrze linję i półkole na połowy; okaże się znowu koniecznem wyżej wspomniane obliczanie zapomocą „współczynników trudności”. Przytem każdy specjalny rodzaj aparatu wymaga specjalnych metod obliczania, ażeby usunąć błędy. Pierwszą pracą, która uwzględniła rozmaite sposoby oceniania w rozmaitych aparatach była Herwiga: Auswertungsverfahren bei der psychotechnischen Eignungsprüfung „PraktPsy” 2 (2). Uwzględnił on jednak w niej tylko aparaty W. Moedego.

Nietylko części jednego testu, lecz także poszczególne testy wykonane przez jednego badanego określa się zapomocą współczynników trudności. Pospolicie oznacza się właściwości konieczne do spełniania jakiegoś zawodu jako

100% „ważności“, a następnie przyznaje się każdej poszczególniej właściwości pewną stopę procentową ważności. I tak w badaniach wybitnie uzdolnionych miasta Berlina wyrażano ważność wszystkich badanych właściwości liczbą 100%. Pamięci logicznej przyznawano 15%, próbie uczenia 15%, testowi Ebbinghausa 10%, testowi analogji 5% i t. d. Badając ucznia, można w ten sposób stwierdzić, czy niedopisał pod względem mniej, czy więcej ważnej właściwości. Powstaje pytanie, na jakiej podstawie opierają się podobne podziały według ważności.

Wspomniane współczynniki można zdobyć empirycznie. Jeżeli jakiś test dajemy pewnej liczbie osób, to niemal z reguły powtórzy się zjawisko, że pewne części testu dobrze rozwiąże wiele, a inne części testu tylko niewiele osób, czasem zaledwie jedna jedyna osoba. Te rzadko rozwiązywane części testu sprawiają zapewne większe trudności i uprawniają do wniosku, że stopień trudności odpowiada częstości rozwiązania.

Wysokość współczynników poszczególnych części testu jest zatem proporcjonalna do częstości rozwiązań. Ale niewiele jest takich dociekań, które zajmują się empirykiem uzasadnieniem współczynników, w większej części wprowadza się je na podstawie rozumowania, przypuszczenia — że tak powiem — na podstawie „duchowej miary w oku“ i w ten sposób stoją one otworem dla wszelkiej dowolności.

Obserwacja w praktyce wskazuje wprawdzie na to, że np. przy *piłowaniu pilnikiem* silniej zajęta jest uwaga niż czułość palców, lecz czy rzeczywiście uwaga zajęta jest dwa razy więcej niż wrażliwość dotykowa (a nie 3 albo $4\frac{1}{2}$ razy więcej) to możemy tylko przypuszczać. Wprawdzie jest możliwe, że właśnie intuicja trafia tutaj w sedno rzeczy, lecz pozostaje ona zawsze czysto subiek-

tywnym środkiem, na którym nie wolno się oprzeć, jeżeli chce się traktować rzecz naukowo. To trzeba mieć wyrażnie na względzie. Musiałoby się wprzód dokładniej zbadać każdy zawód (i lepszymi metodami niż te, którymi rozporządzamy obecnie), ażeby osiągnąć jakiekolwiek liczby, uzasadnione obiektywnie.

Jak już wspomnieliśmy, punkty, osiągnięte w badaniach, podlegają jeszcze ocenie wzgl. nadaje się im stopnie. Czyni się to wtedy, gdy ocenianie w punktach wykazuje wysokie liczby (np. 200 punktów), a wyczyny stają się z powodu tych wysokich ocen nieprzejrzyste. Bardzo często nadaje się stopnie bez uprzedniego oceniania w punktach, jeżeli się wyczyn ocenia natychmiast. Ma to miejsce np. przy testach jakościowych.

W nadawaniu stopni niema obecnie żadnej zgodności. Każdy nadaje stopnie, że tak powiem — według własnego systemu. Najczęściej używa się not 1 — 10 i 1 — 5. Ten ostatni sposób oceniania przy którym 1 = bardzo dobry, 2 = dobry, 3 = dostateczny, 4 = słaby, 5 = bardzo słaby jest sposobem używanym tradycyjnie w szkołach (niemieckich) a posiada i w psychotechnice wielu zwolenników. Spotykamy się z twierdzeniem, że ocenianie według 5 stopni postępuje prędzej i pewniej, aniżeli ocenianie według 10, że zezwala też na subtelne zróżnicowanie ocen, co można osiągnąć przez wstawienie połówek lub znaków + (plus) i — (minus).

Różnica między temi skalami stopni niema jednak znaczenia, ponieważ wyczyny, ocenione według 1 — 5 lub 1 — 10 można bez trudności ze sobą porównać. Stosunkowo o wiele więcej trudności nastęrcza *zasada nadania stopni*, którym wyczynom należy nadać oceny od 1 do 5. Większość psychotechników dzieli poprostu cały zakres

rozsiania pomiarów (np. 2 — 29 zapamiętanych wyrazów w próbie pamięci) na 5 różnych części, i każdej części nadaje się stopnie od 1 do 5. Inni twierdzą, że postępowanie takie jest niesprawiedliwe. Błędy, występujące najczęściej, nie są to błędy najcięższe, muszą one być bliskie badanemu z jakichkolwiek powodów, przeto powinno się je traktować łagodniej. Stopnie powinny się więc stosować do statystyki częstości. Wtedy wyczyn ocenia się według zasady częstości (1 : 2 : 4 : 2 : 1) tak, że mniej więcej 10 najlepszych badanych uzyskuje notę 1, dwudziestu dalszych notę 2, 40 następnych notę 3, 20-u notę 4, a ostatnich 10 notę 5. Rupp, który wśród innych także tego sposobu używa, ogranicza go, twierdząc że jedynka w jednym badaniu nie odpowiada jedynce w innym badaniu, albowiem najlepsze rozwiązania przy jednym badaniu mogą rzeczywiście przedstawiać dobre wyczyny, które się wyraziście odróżniają od przeciętnych, podczas, gdy przy innym badaniu wypadek ten może być bardzo rzadki. W takich wypadkach Rupp nie udziela stopni 1 — 5 lecz tylko 2, 3, 4. Że sąd subiektywny odgrywa znaczną rolę tam, gdzie ocenia się wyczyny, to jest widoczne. Choć w zasadzie postępowanie takie może być nawet słuszne, to jednak może ono ustrzec od błędów tylko wprawnego praktyka.

Zbadaliśmy — jak dotychczas — test tylko z jednej strony (lub w jednym wymiarze), a mianowicie tylko pod względem jakości. Lecz w rozwiązywaniu testu chodzi jeszcze o co innego — o ilość czasu, przeznaczanego na odpowiedź. Dlatego też w teście ocenia się nie tylko jakość, lecz ponadto mierzy się również czas wykonania. Często przypuszczano, że istnieje związek przyczynowy między poprawnością wzgl. błędnością rozwiązania a wielkością

zużytego czasu w tem znaczeniu, że mówi się o prędkim ale fałszywym lub o powolnem, ale dobrem rozwiązaniu. Tymczasem według statystycznego badania, które wykonała p. dr. B a g a n z w Charlottenburskim Instytucie psychotechnicznym na 230 terminatorach, okazało się, że ci, „którzy pracowali bez błędów są również szybcy w pracy, u powolnych zaś zarazem spotyka się najwięcej błędów¹⁾”). Naturalnie, stwierdzono również istnienie wyjątków. Obserwacje innych psychotechników (R u p p) wskazywałyby raczej na to, że szybkość należy traktować jako właściwość istniejącą w osobniku, ponieważ są osoby, które wszystko robią prędko (prędko a dobrze lub prędko a źle). Dla praktyki, ponieważ w wielu zawodach idzie o szybkie wykonywanie ruchu (zwłaszcza przy obsłudze maszyn), stwierdzenie czasu, naturalnie, odgrywa ogromną rolę. (Zresztą z przyzwyczajenia uważa się — i to na podstawie empirji) — że szybkie a dobre działanie jest najbardziej celowe, najbardziej ekonomiczne; dla wielu jest ono zarazem dowodem wyższej inteligencji). Jeżeli więc chcemy stwierdzić wielkość czasu rozwiązania, to musimy określić empirycznie, ile czasu wymaga rozwiązanie pewnego testu.

I tu także osiąga się drogą empiryczną wartości średnie w punktach ustala się np. najkrótszy i najdłuższy czas, którego wymagało rozwiązanie pewnego testu przy większej liczbie badanych (np. 13' i 40' zakres rozsiania wynosi przeto 27'). Jeżeli dla rozwiązania najszybszego ustanowi się 10 punktów, dla najwolniejszego 1 punkt, to oceny leżą między 10 a 1 czyli 1 punkt wynosi $27'/9 = 3'$.

¹⁾ B. Herwig: Auswertungsverfahren bei nichtapparativen psychologischen Proben". PrakPs. 3 (4), str. 121.

W wypadku, gdy dajemy kolejno po sobie dwa testy, przy których również mamy ocenić czas, to musimy również wprowadzić rozmaite współczynniki czasu dla rozmaitych zadań, gdyż trudniejsze zadanie, wymaga naturalnie więcej czasu, niż łatwiejsze. Krótszy czas wyraża się przez niższą liczbę punktów i otrzymujemy w ten sposób rozwiązanie testu, którego jakość oceniamy rosnącą liczbą punktów, a czas malejącą liczbą punktów. Takie podwójne ocenianie jest, naturalnie, mało przejrzyste, a ponieważ w doborze bardzo zależy na przejrzystości wyników badania, więc to podwójne ocenianie w praktyce jest niebardzo celowe. To samo odnosi się do wypadków, w których musimy w teście ocenić nie jakość i czas, lecz jakość i ilość, (zob. w teście z lukami: ile luk wypełniono, ile z nich wypełniono dobrze).

Münsterberg¹⁾ pierwszy spróbował przez odpowiednie postępowanie sprowadzić te dwa pomiary rozwiązania testów do jedności; np. mnożył liczbę błędów przez liczbę sekund i wynik ten przyjmował za miarę wyczynu. Jasne, że można również zastosować tu dodawanie lub odejmowanie. Np. w teście, w którym chodziło o liczbę obliczonych wierszy, a zarazem o liczbę opuszczonych błędów, zastosowałam następujące obliczenie: Każdy wiersz przerobiony uzyskał wartość 1, każdy błąd wartość $\frac{1}{2}$ wiersza, którą odejmowałam od rezultatu. Jeżeli badany wykonał 20 wierszy a zrobił przytem 5 błędów, to otrzymywał notę: $20 - 2\frac{1}{2} = 17\frac{1}{2}$ i t. d.²⁾. Niezawsze da się

¹⁾ Münsterberg: Psychologie und Wirtschaftsleben, str. 60.

²⁾ Baumgarten: „Der Konzentrationstest“, PrakPs. 2 (1) 1921.

przeprowadzić tak przejrzyste obliczenie: dlatego to już Whipple, a w ostatnich czasach przede wszystkim Giese wystąpili z żądaniem, by zawsze uwzględniać tylko jedną wartość, tylko jedną *składową*, i dawać jej postać nadającą się do porównań. W tym sensie mówi Whipple o *time-limit method* i *work-limit method*. W pierwszej metodzie stwierdza się rozmaitość wyczynów w określonym czasie, w drugiej rozmaitość czasów zużytych na wykonanie tego samego wyczynu. Whipple uznaje za lepszą drugą metodę¹⁾. Giese uważa za rzecz obojętną, którą ze składowych uznamy za miernik — czas czy jakość, główną rzeczą jest, ażeby badany musiał wykonać zadanie, spełniając warunki zgóry narzucone bądź pod względem czasu, bądź pod względem jakości, przyczem obliczamy również tylko jeden składnik wyczynu²⁾. (Nazywa on to: Einkomponentenrechnung — rachunek jednej składowej, a nazwa ta się przyjęła³⁾). Wiele z współczesnych aparatów i testów wykazuje tendencję do zastosowania się do tego „rachunku jednej składowej”.

Rzeczywiście „rachunek jednej składowej” jest sposobem uzyskiwania możliwie poprawnych obliczeń wyczynu, ale trzeba pamiętać o tem, że każdy wyczyn w życiu rzeczywistym przedstawia wartości różnokierunkowe. Pewna praca może być wykonana np. w najkrótszym czasie, w pożądanym zakresie i należyście dobrze, czyli charakteryzuje ją czas, jakość i ilość. Ograniczenie psychologicznych metod obliczania stanowi przeszkodę do wszechstron-

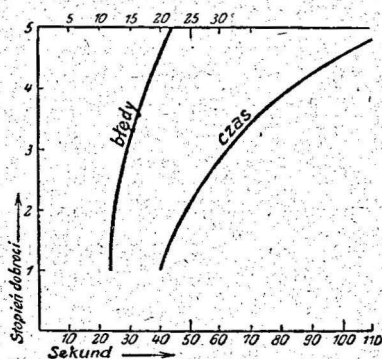
¹⁾ Whipple: *Manual of Mental and Physical Tests* 1910, str. 6.

Giese: *Psychotechnisches Praktikum*. Halle 1923, str. 70.

³⁾ Giese: *Handbuch psychotechnischer Eignungsprüfungen*, Halle 1925, str. 646 — 647.

nej oceny wyczynu; ocena staje się ograniczona, przybiera postać sztuczną. Do dziś nie badano jeszcze, o ile ograniczenie oceny testu do jednej tylko wartości wpływa na obniżenie wartości symptomatycznej testu, o ile fałszuje psychologiczne ukształtowanie badania wyczynu. Np. gdzie idzie o ustalenie typu pracownika — czy jest szybki a dokładny, czy powolny i niedokładny — nie daje „rachunek jednej składowej” prawdziwego obrazu rzeczywistego stanu rzeczy.

Dlatego też tak doświadczony psychotechnik jak Poppelreuter¹⁾ zwalczał ostro „rachunek jednej składowej”, a w praktyce szuka się jeszcze ciągle postępowania, któreby zezwoliło na odosobnione podkreślenie jakości i czasu. I tak pewien zakład w Niemczech środkówowych stwierdza dla każdego zadania liczbę błędów i zużyty czas w sekundach i przenosi uzyskane wartości na wykres czasu i błędów (zob. rys. 3).



Rys. 3.

1) Poppelreuter: Richtlinien, str. 146 i nast.

Dla każdego wykresu czasu i błędów ustanawia się potem klucz do oceny odpowiadający przeważającej ważności czasu lub jakości. Ocenę liczbową dla rozwiązania zadania oblicza się jako średnią arytmetyczną obydwóch stopni z wykresu czasu i błędów. Jasne jest, że takie postępowanie wymaga wiele czasu, a przez to w praktyce jest bardzo „nieporęczne”.

Spór: „Rachunek jednej składowej” czy „rachunek wielu składowych” możnaby rozstrzygnąć przez badanie zgodności jednego i drugiego postępowania z praktyką. H. Kellner¹⁾ podjął właśnie próby tego rodzaju, że ustalał oceny według czterech rozmaitych sposobów postępowania. Raz przy ściśle określonej ilości pracy i obliczonej dobroci pracy nie uwzględnił czasu pracy, drugi raz mierzył czas, trzeci raz ustalał czas metodą rachunkową, czwarty raz metodą empiryczno-graficzną. Rezultaty wykazały, że za najlepszą uważać należy metodę empirycznego uwzględniania czasu próby, ale w tym najpomyślniejszym przypadku okazała się znowu tylko średnia zgodność z praktyką. Zagadnienie więc oczekuje na dalsze badania.

B. Ocenianie testów aparatowych.

Ocenianie testów aparatowych nastrocza jeszcze niektóre szczególne trudności. Pozornie aparat jest idealnym środkiem oceniania, ponieważ może podać natychmiast wyczyn liczbowo, ale niestałość testu, którą omówiliśmy w poprzednim rozdziale, jak i konstrukcja aparatów nie pozwalają na zdobycie liczby bez błędu. Jeżeli badanego podda się kilkakrotnej próbie przy aparacie, je-

¹⁾ Kellner H.: Die Urteilsbildung bei den psychotechnischen Prüfungen IndPste 2 (10) 1925, str. 303 — 315.

żeli np. każe mu się przepołowić linję, to następuje podział więcej lub mniej dobry, który zbliża się mniej lub więcej do obiektywnie poprawnej wartości, czyli jak się to wyraża do obiektywnego punktu zerowego (N). Wychyn przeto wykazuje mniejsze lub większe wahania. Ażeby uzyskać obraz prawdziwego wychynu badanego, czyli aby uniknąć o ile możności wahań, wprowadza się powszechnie 5-o albo nawet 8-o krotne powtórzenie każdego wychynu przy aparacie. W ten sposób uzyskuje się 5 runiej lub więcej różnych wychynów. Jeżeli się obliczy średnią arytmetyczną tych indywidualnych wartości jako przeciętny wychyn badanego, to otrzymuje się t. zw. wartość średnią, która charakteryzuje subiektywny punkt zerowy (NS); potem ustala się, jak wielkie są odchylenia poszczególnych wychynów od tego wychynu przeciętnego. Suma tych odchyłeń podzielona przez liczbę prób daje nam pojęcie o tem, jak wielki jest stopień zmienności wychynu u osobnika badanego. Ta wartość nazywa się *roz-sianiem*, *średnią zmiennością* albo *średniem odchyleniem* (m. V.). Jeżeli wychyny poszczególne wyrażają się np. w wartościach: 2, 8, 4, 6, 4 to ich wartość średnia = średniej arytmetycznej = $2 + 8 + 4 + 6 + 4 = 24$, $24/5 = 4,8$. Odchylenia poszczególnych prób od wartości średniej wynoszą więc:

(4,8 — 2), (4,8 — 8), (4,8 — 4), (4,8 — 6), (4,8 — 4).

Średnia zmienność jest to suma absolutnych wartości tych odchyłeń podzielona przez liczbę prób więc:

$$m V = \frac{|4,8-2| + |4,8-8| + |4,8-4| + |4,8-6| + |4,8-4|}{5} =$$

$$= \frac{2,8 + 3,2 + 0,8 + 1,2 + 0,8}{5} = \frac{8,8}{5} = 1,76$$

Ważność tego obliczenia polega na tem, że na jednej liczbie unaocznia się, czy poszczególne wyczyny są skupione, czy rozproszone około wyczynu przeciętnego. Im bardziej są skupione to zn.: im mniejsza jest średnia zmienność, tem lepszy jest wyczyn badanego. Przy ocenianiu wyczynów na aparacie ważne jest również stwierdzenie, czy u badanego nie zjawiała się tendencja do popełniania błędów w tym samym „kierunku”; w naszym przykładzie, czy nie zjawiała się tendencja do przepoławiania na prawo albo na lewo obiektywnego punktu środkowego, czy więc występowała częściej skłonność do przeceniania, czy niedoceniania odcinków. W takich wypadkach mówi się o „*błędach stałych*”. Błąd stały oblicza się z odchylenia subiektywnego punktu zerowego od obiektywnego punktu zerowego: $K F = N_s - N_o$. Może on wypaść dodatnio przy niedocenianiu, albo ujemnie przy przecenianiu). W podanym przykładzie błąd stały jest dodatni. Jeżeli $N_o = 0$ to $K F = N_s$.

Prócz błędów stałych oblicza się jeszcze „średni błąd absolutny”, jeżeli się ustala wartość średnią odchyłeń między obiektywnym punktem zerowym, a poszczególnymi wyczynami bez uwzględniania znaków algebraicznych.

Jeżeli $N_o = 10$ a wyczyny poszczególne = 2, 8, 4, 6, 4, to średni błąd absolutny wynosi:

$$\begin{aligned} m F &= \frac{|10-2| + |10-8| + |10-4| + |10-6| + |10-4|}{5} = \\ &= \frac{26}{5} = 5,2 \end{aligned}$$

Jeżeli chce się ustalić stosunek odchyłeń do siebie, to uzyskuje się tak zwaną ew. średnią zmienność wewnątrz-

osobniczą ($\bar{I}$ m). Oblicza się ją z sumy odchyień od siebie nawzajem poszczególnych prób z uwzględnieniem ich porządku, a dzieli się przez liczbę prób.

$$I_m = \frac{|2-8| + |8-4| + |4-6| + |6-4|}{5} = \frac{14}{5} = 2,8$$

Którą metodę oceniania należy więc wybrać. Metodę błędów stałych, średniego błędu absolutnego, średniego punktu zerowego, czy wreszcie średniej zmienności? Wybór pozostawiano dotychczas osobistym upodobaniom psychotechnika. Naogół uważa się metodę średniej zmienności za metodę, która umożliwia dokładniejszą ocenę. Ale rozstrzygnięcie powinna dać właściwie praktyka, to zn. specjalne badanie. Blumenfeld na podstawie swoich doświadczeń podniósł zarzuty przeciw ocenianiu wyczynu według subiektywnego punktu zerowego. Twierdzi on, że ocena wyczynów według średniego błędu absolutnego jest o wiele ważniejsza. Wynika to całkiem jasno z przykładu, zacytowanego przez niego. Pięciu kandydatów ma przepławić odcinek, każdy z nich wykonywa zadanie 5 razy.

Kandydat badany	P R Ó B Y	Ns	mV	mF
A	+1+2+3+4+5	+3,0	1,2	3,0
B	—1—2—3+4+5	+0,6	3,1	3,0
C	—1—2—3—4+5	—1	2,4	3,0
D	—2—3—5+1+4	—1	2,8	3,0
E	—1—2—4+3+5	+0,2	3,0	3,0

Z tabeli tej wynika, że kandydat A jest najlepszy. A jednak według obliczenia Ns jest on najgorszy. Według

mF wszyscy kandydaci są równi, zróżnicowanie wartości uzyskujemy przy mV¹⁾). Marta Moers poraz pierwszy dla miary w oku (dzielenie odcinka na połowy) badała: „który sposób oceniania przy tem badaniu byłby najlepszy”²⁾). Z jej przesłanek wynika, że zapomocą średniej zmienności można lepiej przeprowadzić selekcję, aniżeli obliczając błąd stały. Ponieważ mV okazuje się bardziej stała niż KF, więc nadaje się lepiej jako podstawa obliczeń.

Z tego dociekania, które dotyczyło całkiem specjalnego badania wynika, że właściwie należałoby sposób oceniania oprzeć na uprzednich badaniach wstępnych i że większość prac, które bez różnicy posługują się różnym sposobem oceniania według swego widzimisię, (bez uzasadnienia) świadczy o wielkim niedbalstwie.

C. Jakościowa ocena wyczynów, uzyskanych zapomocą testów.

Jeżeli nie można ocenić testu według pewnej miary, względnie określić ilościowych, w takim razie oceniamy go jakościowo. Jeżeli chcemy np. zbadać zdolność kombinacji pewnej grupy osób i wyznaczymy jako zadanie: utworzyć zdanie z trzech słów: zimno, wilgoć, wypadek, to trudno jest oceniać na podstawie jakiejś „miary” takie rozwiązania j. np. „Zimno i wilgoć spowodowały wypadek”. „Aby zapobiec wypadkowi z powodu zimna i wilgoci, trze-

¹⁾ Blumenfeld: Eignungsprüfungen an Lehrlingen der Metallindustrie im Psychotechnischen Institut an der technischen Hochschule Dresden. Maschinenbau — Betrieb, 5 (22) (1923).

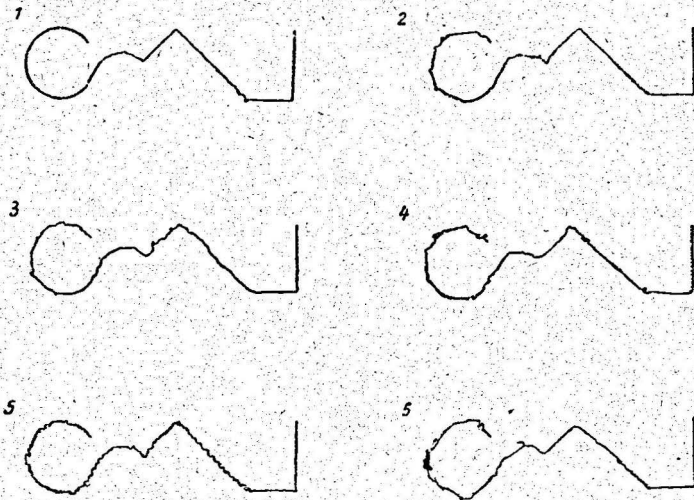
²⁾ Marta Moers: Ein Beitrag zur Untersuchung der Augenmassprüfung Zang Ps. 23 (5 — 6) 1924, str. 238 i nast.

ba chodzić ostrożnie". Takie odpowiedzi można ocenić tylko jakościowo. Ocena jakościowa pozostanie zawsze subiektywna i nie pozwala na znaczniejsze zróżnicowanie wyczynów.

Posiadamy niewiele stopni oceny. Bardzo dobry, dobry, dostateczny, niedostateczny i zły. Dlatego też w badaniach zbiorowych często wogóle rezygnuje się z takich testów, a używa się ich tylko dla lepszego zorientowania się co do badanego, nie przypisując mu takiego znaczenia, co testem dającym się mierzyć. Kilka testów aparatowych nie nadaje się również do mierzenia wyczynów, tak np. (t. zw. po niemiecku *Zweihandprüfer*), aparat, na którym bada się zdolność do współdziałania obu rąk. Zarówno w konstrukcji aparatu *Moedego* jak i aparatu *Ruppa* chodzi o to, ażeby dokładnie przejść ołówkiem po wyrysowanej krzywej linii; w pierwszym aparacie, kręcąc jednocześnie dwiema rękami korbami, poruszamy płytę pod stale przytwierdzonym ołówkiem, w drugim aparacie posuwamy ołówek zapomocą równoczesnego poruszania dwu dźwigni). To odrysowanie bywa mniej więcej poprawne. Należałoby więc stwierdzić wszelkie odchylenia od wzoru i ocenić je. Tej oceny dokonywa się wprawdzie drogą empiryczną, ale w sposób subiektywny. W badaniu próbnym (dokonanem na mniejszej lub większej liczbie kandydatów, dzieli się otrzymane krzywe według ich jakości. Z każdej grupy wybiera się charakterystyczną krzywą jako „krzywą wzorową” (zob. rys. 4). W późniejszych badaniach porównywa się wyczyn kandydata z temi „krzywami-wzorcami” i ocenia się je według podobieństwa z jednym z typów.

A więc i na aparacie ocenianie może być dowolne i subiektywne: po pierwsze przy ustalaniu krzywych wzor-

cowych i podziale ich na 5 albo mniej grup, powtórę przy zaliczaniu wyczynu do jednej z tych grup. Od subiektywnego sądu badacza zależy w zupełności, czy wyczyn kandydata oceni właściwie, czy za nisko lub za wysoko. Od zdolności badającego też zawisło, czy potrafi ocenić podobieństwa, czy kandydat zostanie zaliczony do właściwej kategorii, czy też przeceniony, czy niedoceniony. Ocena tkwi nie w aparacie, lecz w *sądzie badającego*. Nie wolno tego pominąć w czasach, kiedy to „przysięga” się na każdy aparat, kiedy od każdego aparatu oczekuje się bezwzględnie poprawnej oceny wyczynu.



Rys. 4.

D. *Tworzenie oceny ogólnej na podstawie metody testów.*

Powyżej omówiliśmy ocenianie pojedynczego testu, obecnie możemy zająć się drugim podstawowym zagadnie-

niem badania testami: w jaki sposób można porównać ze sobą i odpowiednio ocenić rozwiązania testów większej ilości osób.

Przy wyborze pracowników chodzi o to, aby stwierdzić, który z kandydatów podołał najlepiej całemu szeregowi badań. W jakiż sposób najłatwiej i najprędzej dojść do oceny ogólnej. Jeżeli chodzi o badanie dwu lub trzech osób, to wystarczy proste porównanie ze sobą poszczególnych wyczynów. Przy 100 osobach jednak porównanie takie wymagałoby zbyt wiele czasu, a nawet jest niemożliwe.

O ile badanie obejmuje tylko jedną próbę to postępujemy tak: porządkujemy wyczyny na podstawie oceny w punktach w szereg, tak iż przydzielamy pierwsze miejsce najwyższej liczbie punktów, a na następne miejsce przeznaczamy stopniowo mniejsze liczby punktów. Przypuśćmy, że liczba punktów wynosi dla A — 7, dla B — 20, dla C — 14, E — 15, F — 8, G — 2, H — 26, J — 25, K — 25, K — 33, to uzyskujemy następującą listę rang.

Osoby	Wyczyn w punktach	Ranga
D	36	1
K	33	2
H	26	3
J	20	4
B	20	5
E	15	6
C	14	7
F	8	8
A	7	9
G	2	10

(Naturalnie, że szereg można uporządkować rosnąco, najczęściej jednak używa się szeregu malejącego).

Miejsce każdej osoby w tym szeregu nazywa się *rangą*.

Czasem zdarza się, że 2 osoby wykazują ten sam wyczyn. Np. J i B osiągnęli po 20 punktów. W takim razie jednemu przypada ranga 5-a, drugiemu 4. Prowadzi to do niesprawiedliwości, zwłaszcza, gdy w praktyce o jedno miejsce ubiegają się 4 osoby a 2 z nich na podstawie swych wyczynów wykazują takie same kwalifikacje do otrzymania tej posady. Przypadek (albo dowolność) decyduje wtedy o wyższej randze jednego z nich. Ażeby być sprawiedliwym ustanawia się zamiast dwu rang 4 i 5 jedną rangę $4\frac{1}{2}$ a lista rang wygląda wtedy następująco:

Osoby	Wartość w punktach	Ranga
D	36	1
K	33	2
H	26	3
J	20	$4\frac{1}{2}$
B	20	
E	15	
		5

(Wtedy o wyborze między J i B decydują jeszcze inne względy).

Jeżeli w jakiejś większej liczbie kandydatów (powiedzmy 100) np. 6 osób wykazuje ten sam wyczyn, to w podobny sposób tworzy się dla nich jedną rangę wspólną. Jeżeli np. 6 osób otrzymało w próbie pamięci 20 punktów, to zamiast rang 4, 5, 6, 7, 8, 9, wyznacza się im rangę wspólną, mianowicie:

$$\frac{4+5+6+7+8+9}{6} = \frac{39}{6} = 6\frac{1}{2}$$

Aby tego rachunku uniknąć używa się formuły skła-

dającej się ze sumy skrajnych rang podzielonej przez 2.

$$\frac{4+9}{2} = \frac{13}{2} = 6\frac{1}{2}$$

Jeżeli badanie składa się z większej ilości prób, to dla wszystkich testów tworzy się takie szeregi rang, że ustawia się t. zw. listę ogólną i dodaje się do siebie poszczególne liczby, wyrażające miejsce w szeregu, które otrzymał badany za każdy test. Suma ta oznacza w tym wypadku wyczyn ogólny kandydata; z sum w ten sposób utworzonych zestawia się znowu listę rang. Np.

L I S T A R A N G						
Osoby	Pamięci	Uwagi	Kombinacje	Zręczności	Suma	Ogólna lista rang
A	1	1	1	2	= 5	1
B	3	2	4	4	= 13	3
C	4	5	2	5	= 16	4
D	5	4	5	3	= 17	5
E	2	3	3	1	= 9	2

Osoby ułożone według rangi na liście rang:

- A — 1
- E — 2
- B — 3
- C — 4
- D — 5

Tak postępuje się, jeżeli wszystkie testy uważamy za jednakowo ważne. Zdarzają się jednak wypadki, że niektóre zdolności np. uwaga i pamięć dla niektórych zawodów ocenia się znacznie wyżej np. nawet trzykrotnie wyżej,

aniżeli zręczność, ponieważ są one dla tego zawodu konieczne. Dla każdej przeto badanej właściwości wyznacza się jakąś specjalną wartość (indeks właściwości) i wyznacza się odpowiedni współczynnik; przytem, dla ułatwienia rachunku, określa się wyczyn zbiorowy jako 100% a każdemu poszczególnemu testowi przyznaje się pewną określoną liczbę procentową: np. pamięć 35%, uwaga 30%, kombinacja 25%, zręczność 10%; więc $35 : 30 : 25 : 10 = 7 : 6 : 5 : 2$. Poszczególne rangi mnoży się przez te współczynniki ważności, a potem postępuje się jak wyżej; wtedy nowa lista ogólna przedstawia dla każdego badanego obraz właściwości, pożądaných w pewnym zawodzie.

Osoby	Lista rang				Suma	Ogólna lista rang
	pamięci	uwagi	kombinacji	zręczności		
A	(1.7) 7	(1.6) 6	(1.5) 5	(2.2) 4	= 22	1
B	(3.7) 21	(2.6) 12	(4.5) 20	(2.4) 8	= 61	3
C	(4.7) 28	(5.6) 30	(2.5) 10	(2.5) 10	= 78	4
D	(5.7) 35	(4.6) 24	(5.5) 25	(2.3) 6	= 90	5
E	(2.7) 14	(3.6) 18	(3.5) 15	(2.1) 1	= 49	2

Dzięki tym dwóm obliczeniom dadzą się stwierdzić rzeczy następujące: przez wprowadzenie współczynników ważności przy drugim obliczeniu przesuwają się różnice między wyczynami. Tak np. jeżeli różnica między B a E wynosi w pierwszym obliczeniu cztery punkty (13 — 9) to w drugim obliczeniu = 61 — 49 = 12 punktów; różnica między wyczynami C i D w pierwszym obliczeniu wynosi 1 punkt (17 — 16) w drugim natomiast 12 punktów (90 — 78). Również różnica między wyczynami C i D jest

wskutek drugiego obliczenia większa, aniżeli przedtem. Obliczenie wyczynów staje się w ten sposób bardziej precyzyjne i sprawdziwszy, a to przy większej liczbie kandydatów może mieć również wpływ na rangę badanych osobników.

Widzimy z tego, że rangi zależne są od liczby badanych osób. Ten sam osobnik może przy mniejszej liczbie osób okazać się gorszym, niż przy większej liczbie (albo też naodwrot). Ażeby usunąć przypadkowy czynnik liczby osób badanych, oblicza się niekiedy *rangi procentowe*. Każdy numer rangi dzieli się przez liczbę osób badanych, a mnoży przez 100. Oto takie obliczenie dla naszego szeregu, wziętego za przykład:

$$\frac{1}{5} \cdot 100 = 20$$

$$\frac{3}{5} \cdot 100 = 60 \text{ itd.}$$

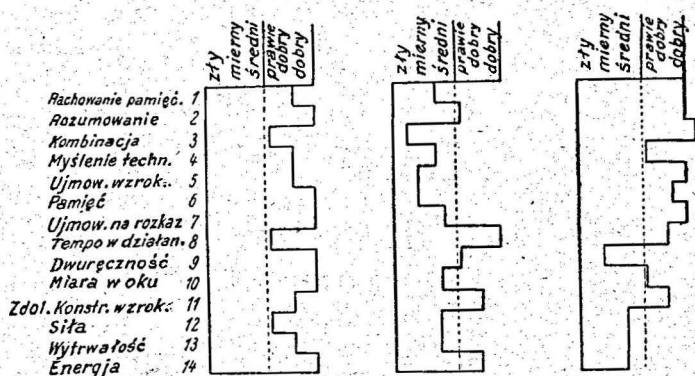
Te rangi procentowe zezwalają na porównywanie rang osobników badanych z jednej grupy z osobnikami z grupy drugiej.

Metoda szeregów rang jest wprowadzić najbardziej rozpowszechniona, lecz nie przez wszystkich psychotechników uznana i stosowana. Poppelreuter¹⁾ i uczeń jego B ü n n a g e l²⁾ podnoszą przeciw nim jako zarzut, że wymagają one skomplikowanych i uciążliwych rachunków, zabierają wiele czasu i sił, a skutkiem tego są nieekonomiczne. (Inne argumenty, podane przez B ü n n a g l a nie są dość silne, dlatego ich tu nie przytaczam). Badacze

1) Richtlinien i t. d. str. 10 i nast.

2) Die Eignungsprüfung in ihrer wirtschaftlichen und sozialen Bedeutung 1923.

ci więc, ażeby zestawić ze sobą wydolność badanych osobników, używają t. zw. *krzywej typowej*, czyli profilu typowego. Dla każdego zawodu rysuje się na podstawie badań większej liczby osób taką krzywą typową; wysokość poszczególnych słupków określa się zapomocą średniej arytmetycznej poszczególnych badań. Dla każdego badanego rysuje się następnie na podstawie badania taką krzywą typową, a krótki rzut oka na jej przebieg wystarczy, ażeby wywnioskować, o ile osobnik badany nadaje się do pewnego zawodu. Krzywa typowa, zdaniem B ü n n a g l a, wystarcza w zupełności, ażeby przeprowadzić selekcję wśród badanych osobników.



Rys. 5.

E. Ustalenie norm.

(Wyczyn średni. — Ocenianie w percentylach).

Taki szereg rang, o którym dopiero co mówiliśmy, jest celowy w tym wypadku, gdy chodzi o *równoczesne porównanie wyczynów* jakiejś mniej lub więcej licznej grupy

osobników. Bardzo często jednak zdarza się, że chcemy przyjąć do pracy tylko jedną osobę i mamy zamiar metodą testów zbadać jej zdolność do zawodu. Wyczyny jej trudno jest osądzić, ponieważ nie mamy ich z czym porównać. Jeżeli taki kandydat, spełniając wyuczył się w przeciągu trzech minut dwunastu szeregów, złożonych z trzech słów — zapamiętał sobie cztery szeregi, to zadamy sobie pytanie: „Czy pamięć jego jest dobra, średnia czy zła”. Musimy więc posiadać taką metodę, która stanowi dobry punkt wyjścia do natychmiastowej oceny takiego wyczynu. Równa się to mniej więcej żądaniu, by uzyskać *stałe normy* (a więc wartości do mierzenia, któreby zatrzymywały na stałe swoją ważność), z którymi można by zawsze porównać — każdy wyczyn.

Aby osiągnąć taką stałą normę, posługujemy się dwoma sposobami:

1) dąży się do stworzenia dla każdego testu t. zw. krzywej wyczynów; dajemy najpierw test do rozwiązania pewnej ilości (np. 50) osób i stwierdzamy nie tylko rozmaitą jakość poszczególnych rozwiązań, lecz także *ilość osób*, które uzyskały pewien stopień rozwiązania. Ze statystyki tej wynika, jaka liczba wartości została najczęściej osiągnięta (stopień częstości) i tę liczbę wartości nazywamy potem *wyczynem średnim*. Jeżeli przyjmiemy, że we wspomnianym teście pamięciowym największa ilość osób zapamiętała pięć szeregów, to właśnie pięć szeregów staje się *wyczynem średnim*, zapamiętanie powyżej 5 szeregów jest wyczynem dobrym, poniżej 5 — wyczynem niedostatecznym. Ta statystyka częstości da się przedstawić graficznie przez wpisanie liczb wartości na oś odciętych — liczb częstości na oś rzędnych. Krzywa przebiega tem prawidłowiej, im większa jest liczba osób badanych, a przy do-

brym teście największa częstość leżeć będzie po środku krzywej, a nie na początku, ani na końcu.

Każdy dowolny wyczyn ocenia się obiektywnie przez porównanie z taką krzywą częstości. W laboratorium psychotechnicznym saskich kolei państwowych w Dreźnie, posługiwano się przy ocenianiu badań maszynistów kolejowych t. zw. *krzywą całkową* (Integralkurve¹⁾); jest to krzywa, powstała ze scałkowania krzywej częstości; posługiwanie się nią okazało się bardzo utrudnione, zbyt ciężkie jest więc zajmować się nią na tem miejscu.

Wyczyn średni możemy otrzymać i w ten sposób, że obliczymy średnią arytmetyczną wszystkich wartości, uzyskanych w badaniu wstępnem. Czasem posługuje się obliczaniem t. zw. *medjany*; wyczyn uклада się w szereg rang i oznacza się wartość, leżącą w środku szeregu jako wartość środkową. (W celu uproszczenia posługujemy się formułą $M = \frac{n+1}{2}$). Medjanę wprowadził do psychologii Kraepelin, który zaliczył zamiast *sumującej* metody wyszukiwania wartości średniej — metodę *wyliczającą* metodę medjany, a to w tym celu, aby usunąć wpływ silnych odchyień²⁾.

2) Obliczenie wyczynu średniego pozwala nam stwierdzić, czy wyczyn w danym wypadku wzięty pod uwagę w porównaniu ze średnim wyczynem zgadza się z nim, czy go przewyższa, czy też do niego nie dorósł. Dla celów

¹⁾ Dr. Ing. A. Schreiber: Mitteilungen aus d. Prüfungs-labor. für Berufseignung bei d. Sächs. Staatseisenbahnen. Nakł. Ver. Deutscher Ingenieure, Berlin 1919.

²⁾ Lipmann rozszerzył metody wyliczające także na inne metody mierzenia psychologii zob. Abzählende Methoden u. ihre Verwendung i. d. psychologischen Statistik Leipzig, 1921.

praktycznych nie zawsze ono wystarcza, gdyż często trzeba poznać *dokładny stosunek* danego wyczynu do normy. okazuje się więc koniecznym jakiś stopniowany podział wyczynów lub właściwości). Claparède¹⁾ zaproponował tu użycie metody oceniania procentowego, percentyli — (fr. le percentilage) której używał już Galton dla innych celów. Obliczanie to okazało się praktycznie użyteczne.

Podobnie, jak przy obliczaniu wyczynu średniego, dajemy i tutaj większej grupie osób (np. 100) jakiś test (np. uderzać ołówkiem możliwie najprędzej o płytę metalową). Jako wyczyn najgorszy 15 uderzeń na minutę. Ale zamiast obliczania wyczynu średniego przez średnią arytmetyczną lub wyszukiwania częstości i t. p. porządkujemy wszystkie wyczyny według jakości wykonania na 100 części = 100 percentyl. Ponieważ badaliśmy 100 osób, każdy wyczyn jednej osoby będzie odpowiadał jednej percentyli. Jeżeli potem uporządkujemy wszystkie wyczyny w szereg rang, to uzyskamy tablicę norm, wyrażoną w percentylach. Percentylom odpowiada ranga²⁾. Da się to przedstawić przejrzyście w tablicy, na której owe 100 osób dzielimy na 10 klas (t. zw. déciles) tak, że każda klasa obej-

¹⁾ Claparède: Profiles psychologiques gradués d'après l'ordination des sujets. Arch. de Psychol. 16 (61) 1916 (także w 8. wyd. Psychol. de l'enfant, str. 344 i nast.).

²⁾ Percentyle nie odpowiadają w zupełności rangom, ponieważ percentyle stawiamy następująco:

0, 10, 20, 30, 40... 100,

podczas, gdy pierwszą rangę musimy oznaczyć nie 0, lecz 1, jeżeli więc rangi znajdują się między 1 a 100, to cała odległość wynosi nie 100, lecz 99, mianowicie: 1. 10.9 20.8 30.7 40.6... 100. Tego małego błędu nie bierzemy jednak pod uwagę.

muje 10 osób¹⁾. Do pierwszej klasy przydziela się tych, którzy wykonali próbę najgorzej, do drugiej tych, którzy wykonali ją lepiej, do ostatniej tych, którzy wykonali ją najlepiej. Z tem stopniowaniem łączymy wyczyny, które również podzielone są na dziesięć klas (zob. tabl. poniż.), przyczem I klasa reprezentuje 10% wyczyn. II klasa 20% wyczyn i t. d. Jeżeli teraz badać będziemy jakąkolwiek osobę, ile uderzeń wykona w minucie i skonstatujemy 80 uderzeń, to na podstawie tablicy możemy natychmiast stwierdzić, że według jakości wyczynu należy ona do V klasy, czyli wyrażając procentami, że jakość jej wyczynu wynosi 50 na 100, rangą jej przeto jest 50.

X	kl. 100%	165 — 150 uderzeń
IX	„ 90%	150 — 135 „
VIII	„ 80%	135 — 120 „
VII	„ 70%	120 — 105 „
VI	„ 60%	105 — 90 „
V	„ 50%	90 — 75 „
IV	„ 40%	75 — 60 „
III	„ 30%	60 — 45 „
II	„ 20%	45 — 30 „
I	„ 10%	30 — 15 „

Tego uciążliwego rachunku można jednak uniknąć, obliczając percentylę zapomocą wzoru

$$R. = 1 + \frac{(Pn - 1)}{100}$$

przyczem P. oznacza wartość procentową a n liczbę osób.

¹⁾ Jeżeli zbadaliśmy więcej osób, niż 100, np. osób z tem, że $n = 100 + X$, to każda z 10 klas obejmować będzie $100 + X / 10$ osób.

Ten sposób oceniania, poza tem że pozwala dokładnie wyznaczyć każdemu wyczynowi i miejsce w tablicy norm, pozwala również na porównanie rozmaitych osób, przynależnych do rozmaitych grup, a także na porównanie ze sobą wyczynów dwóch różnych grup. Jeżeli jednego roku zgłasza się grupa terminatorów A. a terminator X. uzyskał za pewien wyczyn percentylę 24,6, jeżeli w następnym roku w grupie B terminator Y uzyskuje tę samą percentylę za wyczyn znacznie lepszy, to można z tego wnioskować, że grupa terminatorów B jakościowo stoi znacznie wyżej. Naturalnie ma to znaczenie dla oceny personelu, który zgłasza się w rozmaitych odstępach czasu do tego samego zakładu przemysłowego.

Wyżej wspomniane stopniowanie procentowe zezwala jeszcze na następujące ważne stwierdzenie: jeżeli osobnik badany ma wykonać kilka prób, a jego wyczyny wykazują: w teście pamięci punktów 40, w teście na uwagę 60, w próbie zręczności 70, to te liczby same przez się mówią bardzo mało. Jeżeli jednak liczby te przedstawimy w percentylach i uzyskamy rezultat — dla pamięci 20%, dla uwagi 15%, dla zręczności 40% wartości, to jasne jest, że główna zaleta osoby badanej leży w zręczności.

Percentyle zezwalają przeto na sprowadzenie wszystkich wartości do tego samego szeregu stopni (do tego samego mianownika), a w ten sposób umożliwiają poznanie ilościowego stosunku (stosunku wielkości) poszczególnych właściwości.

Użycie metody percentyl do ustalania norm jest w Niemczech mało rozpowszechnione, mimo to propagował je bardzo Giese¹⁾. Częściej uwzględnia się wyczyn śred-

¹⁾ Giese. Psychol. Normen in Grundschule und Berufsbera-

ni. W Ameryce, szczególnie od wprowadzenia badań psychotechnicznych w wojsku, używa się jako normy tak zwanych *wartości stałych*, albo *wzorów wartości*. Są to wartości przeciętne, uzyskane na podstawie bardzo szerokich i bardzo starannych zastosowań testu. Zanim używano się wartości stałe dla badań w armii, przeprowadzono dwa badania wstępne 30-u testów, zaproponowanych przez komisję, złożoną z najwybitniejszych psychologów. Z tych 30 testów pozostało tylko 10. Te dziesięć testów dano do rozwiązania około 5.000 żołnierzy i osób z poza wojska (niedorozwinięci, studenci z kolegów i uniwersytetów, nauczyciele szkół wojskowych). Wyniki badania opracowano statystycznie, a równocześnie testy zastosowano do osób, które już innymi metodami psychologicznymi były badane; jeśli badano dzieci szkolne i studentów, posługiwano się dla porównania opiniami nauczycieli. W ten sposób porównano wyniki badań wstępnych owych „Army Intelligence Tests” z innymi miarami inteligencji. Dopiero na podstawie zgodności wyników badania ze sobą i z opiniami nauczycieli, podzielono sposoby rozwiązywania na kilka stopni, na stopnie inteligencji (A, B, C, C⁺, C, C — D, D —), przyczem każdemu stopniowi odpowiadała określona liczba punktów. Cała skala obejmuje od 0 do 212 punktów; bardzo wysoki stopień inteligencji ma 135 do 212 p., wysoki 105 — 134 p., ponad przeciętny 104 — 75 p., przeciętny 74 — 45 punkt., poniżej przeciętny 44 — 25 p., niski 24 — 15 p., bardzo niski 14 — 0 p. Te wartości w punktach stały się *wartościami stałymi*¹⁾). Każ-

tung, Halle 1916; także: Handbuch der psychotechn. Eignungsprüfungen, 1925. Nr. 652.

¹⁾ C. S. Joakum and R. M. Yerkes: Army mental Tests. New York, Holt et Co 1920, str. 17.

de badanie określono według tych wartości. Na tej samej podstawie próbuje się obecnie w Ameryce ustanowić wartości stałe dla badań specjalnych jak np. dla znajomości języków, zdolności matematycznych, technicznych i t. d.

Należy jeszcze wspomnieć o tem, że niektórzy badacze opracowali swoje własne sposoby oceniania i używali ich, choć nie przyjęły się one w zupełności. Do nich należy Lippman¹⁾ i Link²⁾.

F. Graficzne przedstawienie wyników badań.

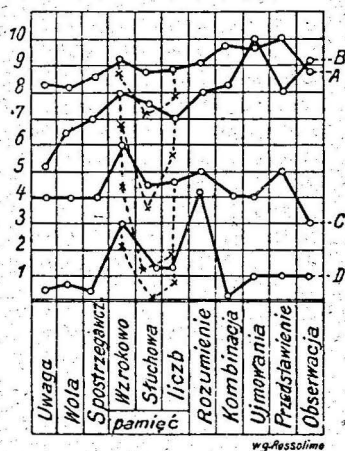
Wyniki badań dadzą się uzmysłowić w graficznym przedstawieniu, przez co unaocznia się ogólny obraz badanych właściwości. Zasłużył się w tym kierunku znany rosyjski lekarz Rossolimo, on bowiem pierwszy spróbował narysować *profil psychologiczny* zdolności duchowych. Stopnie, które dawał każdemu osobnikowi badanemu za rozwiązanie pewnej liczby testów, (były to stopnie 0 — 10) przenosił na papier milimetrowy, jako wyższe lub niższe linie, stosownie do stopni. Gdy połączyło się — punkty szczytowe tych linii ze sobą, otrzymywało się linję, której wyrazista forma zygzakowata, nazwana została *profilem*. Zamiast, jak było z początku, rysować linje odpowiadające ocenom, można wyznaczyć tylko punkty, oznaczające wierzchołki linii i punkty te łączyć ze sobą (p. rys. 6).

Przez połączenie punktów profil wywołuje czasem złudne wrażenie co do „wyczynu dominującego”, zależnie

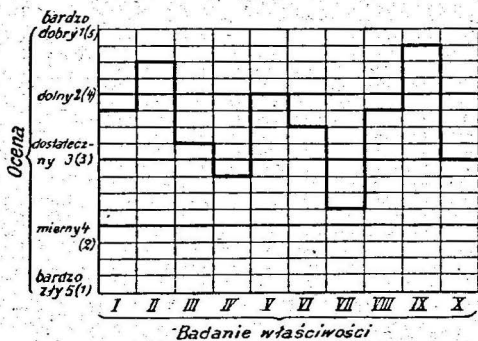
¹⁾ Lipmann i Stolzenberg: Methoden zur Auslese hochwertiger Facharbeiter der Metallindustrie SchrPsBerWi 11.

²⁾ Link: Eignungspsychologie, übers. von Witte, München. Oldenbourg 1922.

od tego, czy badany osobnik przedstawia się dobrze, lub źle pod względem zdolności, umieszczonych w najbliż-



Rys. 6.

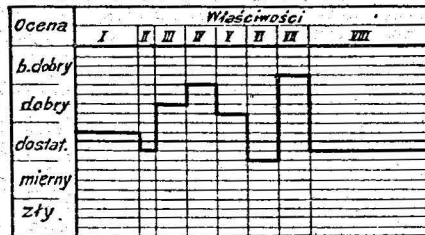


Rys. 7.

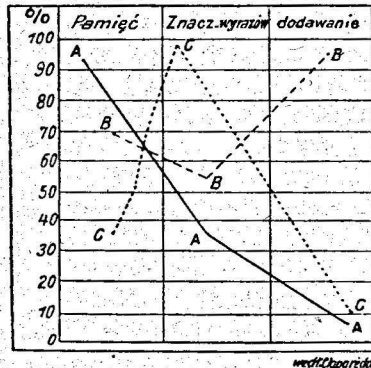
szem sąsiedztwie; dlatego też lepiej jest ocenę rysować nie jako punkt, lecz jako linię prostą t. zw. (belka, rys. 7).

Czasem nadaje się belce szerokość, odpowiadającą ważności pewnej zdolności (rys. 8).

Wobec tego, że oceny nie były sprowadzone do tego samego mianownika, przeto profil, na nich oparty, nie może dać poprawnego obrazu właściwości. C l a p a r è d e



Rys. 8.

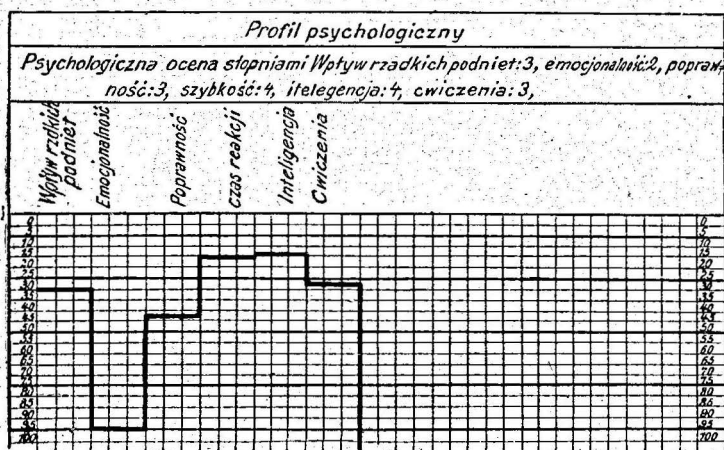


Rys. 9.

poprawił więc ten błąd, przeliczując oceny na percentyle i rysując profil na ich podstawie (p. rys. N. 9).

Laboratorium psychotechniki przemysłowej na Politechnice w Charlottenburgu posługiwało się takim gra-

ficznem przedstawieniem wyników badań i nazywało je *krzywą właściwości* (Eigenschaftskurve¹). Na prostokątnym układzie współrzędnych wyznaczono na osi odciętych badane właściwości, na osi rzędnych liczby wartości, wynikiem tego jest krzywa schodowa, której punkty szczytowe przedstawiają punkty szczytowe uzdolnienia. Czasem wyznacza się oceny na osi odciętych, a właściwości na osi



Rys. 10.

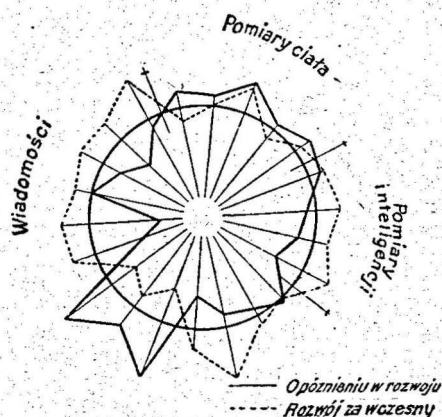
rzędnych, jak to czyni A. E. G.²). Krzywej profilowej, obliczanej w procentach używał Giese; metodę tę przyjęło laboratorium psychotechniczne Ludowego Komisarjatu Pracy w Rosji; załączamy tutaj taki profil procentowy

¹) B. Herwig: Auswertungsverfahren bei der psychologischen Eignungsprüfung. PrakPs. 2 (2), str. 57 — 58.

²) Heilandt A.: Psychotechnische Eignungsprüfungen bei der Einstellung gewerblicher Lehrlinge in der Werkschule der A. E. G. Fabriken. Der Betrieb 3 (1).

motorniczego, opracowany według moskiewskiego laboratorium (rys. 10).

Próbowano również graficzne przedstawienie wyników badania ująć nie tylko w formę linearną, lecz także w formę koła. B a l d w i n¹⁾ porównywał w ten sposób wyczyny dzieci wyjątkowo uzdolnionych i opóźnionych w rozwoju z wyczynami normalnymi; obraz uzyskany, jak wskazuje



Rys. 11

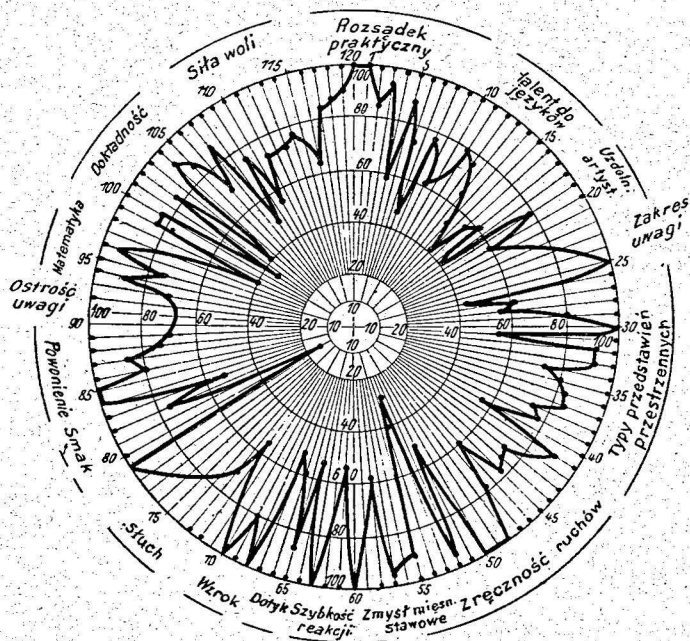
rys. 11 jest bardzo przejrzysty. Tą samą metodą posługiwali się Jakób van Ginneken¹⁾ i Lämmel²⁾. Właściwości osobnika normalnego przedstawia się jako koło. Po stwierdzeniu zdolności t. j. po wymierzeniu jej war-

¹⁾ The Johns Hopkins University Studies in Education N. 3.

²⁾ J. v. Ginneken: Zielerleken. Der Beroepsken anschauweliggheit. Overgedrukt uit „Studiën” Okt. 1922.

³⁾ Lämmel R.: Intelligenzprüfung und psychologische Berufsberatung. München — Berlin, Oldenbourg 2 wyd. w r. 1923.

tości przedłuża się lub skróca promień koła. Krzywą, która łączy te długości promieni i przedstawia odchylenia od obwodu koła nazewnątrz, albo dowewnątrz, nazwał L ä m m e l *ingenogramem*. W załączeniu podajemy takie dwie krzywe kołowe Ginnekena (rys. 12a i 12b)

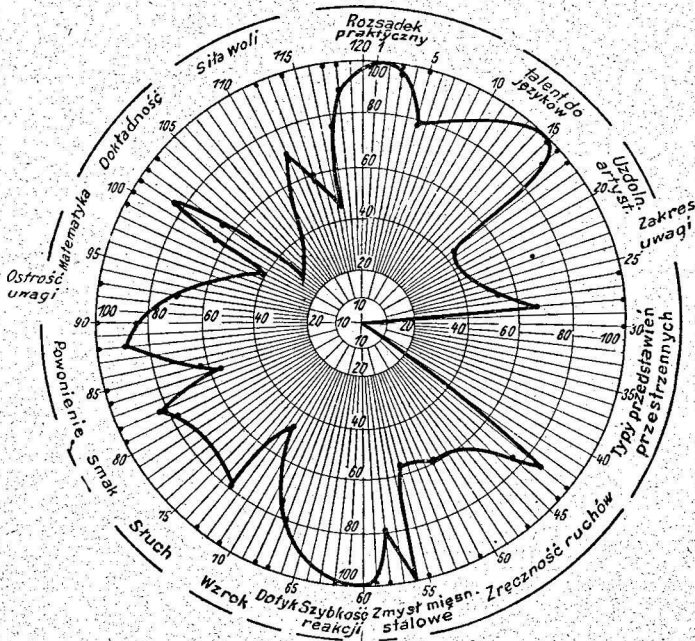


Rys. 12a.
(Harmonijny rozwój zdolności)

i ingenogram L ä m e l a, (rys. 13), który przedstawia właściwości osobnika badanego w porównaniu z typem normalnym.

Profile pozwalają na uzyskanie już na pierwszy rzut oka obrazu uzdolnień osobnika badanego, a na podstawie

dokładniejszego przeglądu na przeprowadzenie paraleli, między wyczynami dwóch lub więcej kandydatów. Według poglądu Giese'go mają one także znaczenie dla oceny psychologicznej, ponieważ umożliwiają zaoszczędzenie cza-

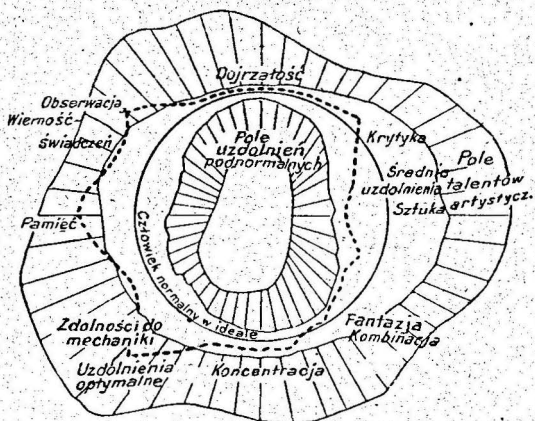


Rys. 12 b.
(Wielkie uzdolnienia obok wielkich braków)

su i pracy. Zamiast wypisywać ocenę, wystarczy na formularzu wyznaczyć każdorazowo procenty i połączyć punkty końcowe. Wyciągnięcie krzywej profilu ma tę zaletę, że jest ona zrozumiała nawet dla laika¹⁾. Wbrew te-

¹⁾ Giese: Psychotechnische Eignungsprüfungen an Erwachsenen, str. 312 — 313.

mu należy zaznaczyć, że Poppelreuter zdecydowanie wystąpił przeciw używaniu profilów i to zarówno ze względów teoretycznych jak i praktycznych. Uważa za bardziej wskazane zastąpić świadectwo psychologiczne orzeczeniem opisowem — na wzór świadectwa lekarskiego.



Rys. 13.

Słusznie zaznacza, że profil daje laikowi złudzenie dokładności, której w gruncie rzeczy profil ten nie posiada¹⁾.

5. ŹRÓDŁA BŁĘDÓW W BADANIU TESTAMI.

Chociaż jakiś test już został gruntownie zbadany pod względem wartości symptomatycznej i dajęnostycznej, to jednak zastosowanie go nie jest jeszcze zupełnie pewne i wolne od zarzutów; jego pewność zależy jeszcze od dwóch czynników. Czynników tych należy szukać:

- 1) w osobie badanego,
- 2) w osobie badającego.

¹⁾ Richtlinien i t. d. str. 29 — 32.

A. *Niestałość osobnika badanego.*

Przy badaniu testami rozmaici badacze stwierdzili, że ten sam test, stosowany do tej samej osoby kilkakrotnie w bezpośrednim następstwie, albo po upływie kilku dni, dawał rozmaite odpowiedzi dobre i złe. Osobnik badany na przykład w pierwszym dniu, bardzo dobrze podolał próbie pamięci, następnego dnia rezultat tego samego zadania jest całkiem niezadowolający, trzeciego dnia znowu zadowolający. Zjawisko to może całkiem obalić zaufanie do badania testami, bo przecie zrodzi się pytanie: któryż wyczyn odpowiada rzeczywistym uzdolnieniom osobnika badanego, maksymalny, czy minimalny? którego mamy się trzymać? Jeżeli zbadamy większą liczbę osób i uporządkujemy je według rang, to jakąż stałość posiada ten szereg?

Jeżeli się niestałość taka odnosi do prób, podejmowanych w przeciągu kilku dni lub tygodni, to skłaniamy się do wyjaśnienia, że w pewnych dniach jest się lepiej, w innych gorzej usposobionym, przygnębnym troskami lub usposobionym beztrosko. Jeżeli jednak takie wahania możliwe są w przeciągu kilku minut, to podejmowanym próbom musi się bezwzględnie odmówić immanentnej stałości. W takim wypadku test nie posiada zupełnie wartości prognostycznej, o co przecież głównie chodzi w badaniu zdolności do zawodu. I rzeczywiście brak stałości doprowadził do tego, że zakwestjonowano wogóle praktyczną wartość używania testów (Robert a g). Krüger — i Spearman podjęli „próbę obliczenia współczynnika pewności” jako kryterjum stałości testu¹⁾. A b e l-

¹⁾ Die Korrelation zwischen verschiedenen geistigen Leistungsfähigkeiten. Zeitschr. für Psych. Bd. 44 1906.

son¹⁾ żądał, aby tego samego testu używać w bezpośrednim następstwie większą ilość razy (niektóre testy 8 razy (a potem oceniać wartość średnią tych badań. Jones²⁾ zaproponował, ażeby rozróżnić dwa wskaźniki: *reliability index*, jeżeli chodzi o stałość wyników badań następujących po sobie i *stability index*, jeżeli chodzi o badania przedzielone znaczniejszymi okresami czasu, mające za cel zbadanie stałości pewnych właściwości u kandydata. Claparède przeprowadził cały szereg eksperymentów, by zgłębić to zagadnienie³⁾, a wynikiem tego jest stwierdzenie, że niektóre testy posiadają większą stałość, drugie mniejszą. (Np. pamięć wykazuje mniejszą stałość niż szybkie pisanie). Pozatem doszedł on do wniosku, że istnieje różnorodność indywidualna, skutkiem której jeden osobnik wykazuje większą stałość w odpowiedziach aniżeli inny. Claparède uważa za najsłuszniejsze badać tę samą właściwość zapomocą różnych testów i uważać wartość średnią tych wyników za reprezentatywną⁴⁾

W ostatnich czasach H. Bogen⁵⁾ zajmował się kwestją stałości wyników testów. Na podstawie swych badań doszedł do przekonania, że djaagnostyczna pewność szeregu rang jest tem mniejsza, im trudniejsze było zada-

1) Mental ability of backward children. Brit. Journ. of Psychol. IV. 1911.

2) E. S. Jones: The influence of age on correlations Baltimore 1917, str. 2.

3) Claparède: De la constance des sujets à l'égard des tests d'aptitude Arch. Psych. 17. 1920.

4) ibid str. 344. Zob. także Berliner Anna: Wiederholung einer psychologischen Untersuchung nach einjähriger Pause. ZangPs. 19. 1921.

5) Zur Frage der Rangreihenkonstanz bei Begabungs- und Eignungsprüfungen, ZangPs. 20 (3/4) 1922.

nie¹⁾, ²⁾. Za przyczynę niestałości uważa on wahania w rozwoju ogólnej energii psychofizycznej; odróżnia on energję według *stopni* (nazywa ją *Disponibilität*) i według *rodzajów* (nazywa ją *Disposivität*, czyli dyspozycją strukturalną). Widzimy z tego, że przyczyną niestałości testów przypisuje się *niestałości osobnika badanego*. Powstaje pytanie, jaka cecha testów zdołałaby usunąć tę niestałość badanego? Jakie cechy musiałby posiadać taki test? W ten sposób dochodzimy znowu do omówionego już zagadnienia symptomatycznej, dżagnostycznej i prognostycznej wartości testu, przede wszystkim do problemu kontroli testu (a pośrednio znowu do metody korelacji³⁾). Bardzo mało dokonano dotychczas zapomocą badań w tym zakresie dla celów praktycznych. W praktyce uwzględnia się to całe zagadnienie w ten sposób, że przy wielu próbach, szczególnie wtedy, kiedy chodzi o testy aparatuowe, powtarza się próbę kilkakrotnie i stosuje się wartość średnią bardziej wyczerpująco (mówię o tem niżej). Przy serjach testów stosuje się ponadto próby nie w dniach po sobie następujących, lecz po upływie 8 dni (zastosowano w Berlinie, badając wybitnie uzdolnionych). W badaniach pojedynczych uwzględnia się ponadto pewne przyczyny osobiste, a więc stan zdrowia, zmęczenie, przygnębienie, troski, dążąc do usunięcia ich wpływu względnie do przesunięcia badania na inny dzień.

¹⁾ Str. 167.

²⁾ Zgadza się to ze zdaniem Lippmanna, kiedyś wypowiedzianem: im niższy jest rodzaj pewnej czynności, tj. im bardziej jest ona czystą funkcją zmysłową a im mniej wymaga wyższych duchowych właściwości, jak inteligencja i uwaga, w tym mniejszym stopniu ulega ona wahaniom dyspozycyjnym i wpływom ćwiczenia.

³⁾ Na ostatnim IV Międzynarodowej Konferencji Psychotech-

B. *Nastawienie i zmiana nastawień osoby badanej
względem badania.*

Ogromne znaczenie dla wyników badania ma *nastawienie* osobnika badanego względem badania.

Już postawa cielesna osobnika badanego ma wpływ na wynik rozwiązywania testów. W dziale psychotechnicznym wystawy przemysłowej Tow. Inżynierów Niemieckich (w r. 1923) znajdował się rysunek Frytza Wernera, przedstawiający wadliwą i poprawną postawę osobnika badanego, która zaznaczyła się wyraźnie w wynikach jego prób.

Jeszcze bardziej zaznacza się *nastawienie psychiczne* kandydata podczas badania. Kandydat poddaje się chętnie lub niechętnie stwierdzeniu swej zdolności (z rozmaitych przyczyn osobistych: jedni lubią wysiłek — drugim wysiłek przychodzi ciężko), a stosownie do tego u jednych występuje „gotowość” do poddania się badaniu, co ułatwia rozwiązanie, u drugich pewne „otamowanie”, która rozwiązanie utrudnia. Zagadnienie odrębne stanowi stosunek *robotnika i pracownika* do badania. Młodociany wykonywa automatycznie rozkaz dorosłego (nauczyciela lub przełożonego), lecz robotnik jakiegoś zakładu przemysłowego nie przyjmuje bezkrytycznie nieznanych mu dotychczas rozporządzeń. Dla stanowiska kierowników badań dadzą się tu zauważyć: jedni informują zgodnie z prawdą o celu badań (tak np. laboratorja drezdeńskie dla maszynistów kol.) inni natomiast (głównie w Ameryce) informują błęd-

nicznej w Paryżu (październik 1927) pytania te omawiano bardzo żywo. Okazały się ogromne powikłania w terminologii i określeniach pojęć, które dopiero przez pracę specjalnej komisji mają być wyjaśnione.

nie. P a t e r s o n, który badał niższych urzędników, mówił im: „testy nie mają nic wspólnego z pańskim zawodem — prosimy pana tylko o pomoc w przeprowadzeniu kontroli testu”¹⁾. Jest więc rzeczą zdolności kierownika badań żeby pozyskał zaufanie, jeżeli robotnik podda się chętnie badaniu. Robotnikowi może się test wydawać „interesującym” albo „niedorzecznym” i zbudzić w nim przez to samo życzliwość, nieżyczliwość lub obojętność. Choć zadanie może być łatwe, to w tej chwili, kiedy zacznie się je traktować właśnie jako zadanie, może ono wywrzeć działanie podniecające albo hamujące. W osobniku badanym może się zjawić nagle przytomność umysłu, która fałszywie określi stopień i siłę jego właściwości, albo znów pozostaje on zamknięty w sobie tak, że nic z niego wydobyć nie można²⁾.

Takie przeciwległe nastroje, znane są wszystkim z egzaminów, wystąpień publicznych, wogóle z tych wypadków, gdzie chodzi o jakąś próbę. Wspomnijmy tu o typie „esprit d'escalier”, któremu najlepsze pomysły i odpowiedzi przychodzą dopiero wtedy, kiedy już nie mogą być zastosowane (przypomnijmy sobie podobne skargi J. J.

¹⁾ P a t e r s o n: File Clerks Tests JPeRes 1 (12) 1923. Str. 552.

²⁾ Niektórzy autorowie, jak Hildegard Sachs (w „Psychologie u. Berufsberatung” Langensalza 1925) twierdzą: „Eksperyment laboratoryjny wywołuje oniesmielenie lub podniecenie” (str. 96) i dlatego odmawiają badaniom eksperymentalnym wartości symptomatycznej. Wniosek ten nie jest słuszny. Jeżeli badany jest oniesmielony czy podniecony, to w tem okazuje się jego natura, która również objawi się i w innych sytuacjach życiowych. Ważne jest właśnie, by tę naturę wykryć zapomocą badania i wziąć pod uwagę przy udzielaniu wskazówek co do zawodu. Ludzie emocjonalni nie wchodziliby nigdy do tych zawodów, gdzie potrzebna jest zimna krew.

Rousseau w jego „Confessions”); niezapomnieć też trzeba o „nieśmiałyach” których opisuje Dupras¹⁾. W M. Brown²⁾ mówi o „caution faktor” (nazywa go „Czynnikiem Zeta”), który stanowi określenie dla stopnia *ostrożności* badanego. Z pomiędzy osób badanych przeto jedni dają z siebie maximum, inni zaś minimum; badanie więc w zasadzie winno uwzględnić owo plus lub minus w wy-czynię, ażeby uwzględnić właściwy obraz rzeczywistych możliwości kandydata.

Dodamy jeszcze kilka słów o charakterystycznym wypadku w nastawieniu osobnika badanego, który wynika z samej natury niektórych testów. Testy takie jak t. zw. „Mut-prüfer” (badanie odwagi) lub badanie bojaźliwości i t. p., — wogóle testy, w których wydobywa się na jaw uczucia — w zupełności zawodzą. Badany ma najgłębsze przekonanie, „wszystko to jest zabawka, a nie rzeczywistość, nic złego mi się tu nie stanie”; w ten sposób rozwija się w nim w stosunku do badania pewna obojętność, czasem nawet niechęć do wykonywania testów. Nie jest również wykluczone, że podobne uczucie „nierealności” zbudzi się także w stosunku do innych testów, a więc rozwiązanie znowu nie będzie właściwym obrazem zdolności osobnika badanego. Drugi rodzaj nastawienia polega na przekonaniu kandydata, że zadanie, które otrzymał, choć wydaje mu się bardzo trudne, musi mieć jednak jakieś rozwiązanie; w przeciwnym razie nie dawanoby go. Nastawia się więc na wyszukanie rozwiązania. Można mu dać kilka nieprawidłowych części i powiedzieć, że ma z tego złożyć prostokąt,

¹⁾ Dupras: Les grands timides. Paris, Alean 1922.

²⁾ W. M. Brown: A Study of the „caution” factor and its importance in Intelligence test performance. AmJPs 35 (3) 24.

a on, wychodząc ze swego założenia, będzie próbował go złożyć; jeżeli mu się nie uda, to powie „nie mogę tego wykonać” (a nie — „to jest rzecz niemożliwa”). W rzeczywistym życiu niema się wrażenia, że każda sytuacja musi znaleźć rozwiązanie bez reszty. To „nastawienie na rozwiązanie” przy teście ułatwia wprawdzie zadanie, ale wytwarza także pewne wrażenie nierealności sytuacji.

Nietylko nastawienie względem badania, ale także zdolność szybkiej zmiany nastawień ma wielkie znaczenie dla wyników badania. Allers mówi o zdolności „formującej” („Modellfähigkeit”) pewnych osób badanych, które szybko podczas badania wczuwają się w fikcyjną sytuację, naśladującą na kształt modelu pewne zabiegi należące do pracy zawodowej. Drudzy wczuwają się z trudnością, inni znowu zupełnie tego nie potrafią¹⁾. Tych czynników nieuwzględniono jednak dotychczas nigdy przy badaniach.

C. *Wpływ osoby badającej.*

Wpływ osoby badającej na wynik badania może być dwójakiego rodzaju: 1) przez jego osobiste oddziaływanie na osobnika badanego, 2) przez sposób prowadzenia badań zapomocą testów.

Już przy laboratoryjnych eksperymentach psychologicznych zauważono niejednokrotnie, że osoba badającego wywiera wpływ na wypowiedzenia osobnika badanego (na przykład zauważono, że badani mężczyźni doznają zahamowania wobec badających kobiet, podobne kobiety wobec badających mężczyzn. Najniesłuszniej w świecie nie

1) R. Allers: „Psychologische Berufsberatung” w: „Laubers Handbuch der ärztlichen Berufsberatung”. Wien. 1923, str. 9.

uwzględniano tego zupełnie w naukowych dociekaniach. W badaniach dla celów praktycznych, okazało się, że z osobą badającego należy się bardzo liczyć, jeżeli chce się przeprowadzić badania bez zarzutu. Giese zrobił pod tym względem doświadczenie, że robotnicy komunistyczni w Halle, którzy się nastawiali nieprzychylnie w stosunku do badań, zmienili swoje stanowisko, gdy badania zaczęły przeprowadzać kobiety. Wyciągnięto z tego wnioszek, że do badań zdatności do zawodu nadaje się wogóle o wiele lepiej kobieta, aniżeli mężczyzna. Ma ona cierpliwość, zdolność wczuwania się, łagodniejszy sposób obcowania z ludźmi nawet w stosunku do trudnego materiału, do osób histerycznych, do niesfornych nieuków. Raczej zawodzi kobieta przy badaniu innych kobiet. To, ciekawe samo przez się twierdzenie, nie może mieć jednak ogólniejszego znaczenia¹⁾. Znane są zakłady przemysłowe, gdzie właśnie robotnicy protestowali przeciw badaniu przez kobietę i do tego stopnia ją wysmiewali, że musiano badań zaprzestać. Robotnicy czują się mianowicie „poniżeni” propozycją, żeby badała ich kobieta, którą uważają za niższą w porównaniu z mężczyzną. Z drugiej strony są takie robotnice, które najchętniej wyjawiają przed kobietą swoją nagą duszę. Ażeby więc zapomocą badań wytworzyć sobie poprawny obraz o zdolnościach badającego, trzeba wpierv zorjentować się w zasadniczem stanowisku osobnika badanego w stosunku do osobnika badającego.

Ale nietylko płeć badającego, lecz także jego indywidualność wywiera wpływ na kandydata. Ten sam badający może na jednego badanego oddziaływać podnieca-

¹⁾ Giese: Zur Betriebsführung psychotechnischer Prüfstellen. PrakPs. 3 (1), str. 4.

jaco, na drugiego hamujaco. Bowiem między dwiema osobami, które zetkną się ze sobą, powstaje mniej lub więcej sympatyczny lub niesympatyczny stosunek, a to uczucie przyciągania lub odpychania uzewnętrznia się również w mniej lub więcej życzliwym zachowaniu się osoby, poddanej badaniu. Jest to jedna z przyczyn, dlaczego ten sam badany wobec rozmaitych osób badających często rozmaicie się zachowuje¹⁾.

Ponadto trzeba zwrócić uwagę na wpływ sugestywny, który wywiera osoba prowadząca badania na badanych. Jeżeli badający wchodzi żwawo energicznym krokiem do pracowni, to i badani odczuwają przypływ pewnej energii. Jeżeli przy przyrządach do badania wykonywa szybko, pewne ruchy, to i u osób badanych zjawi się dążenie do wykonywania ruchów w ten sam sposób. I naodwrot, niedbała opuszczona postawa, powolne ruchy wywołują i u badanych odpowiednią postawę, obojętność, niechęć do badań. (Często nerwowy badający wywołuje zdenerwowanie badanego). Te prawa naśladownictwa, które w ba-

¹⁾ Uderzający przykład miałam na następującym przypadku: chłopca około piętnastoletniego, który wykazywał jako praktykant handlowy jaknajgorsze wyniki, przesłano do poradni zawodowej, ażeby poddać badaniu jego zdolności. Badający dał mu już kilka testów, które jednak były rozwiązane zupełnie źle, gdy nagle zmuszony jakąś sprawą, niecierpiącą zwłoki, wyszedł, a mnie przekazał doprowadzenie badania do końca. Chłopiec rozwiązywał wszystkie testy znakomicie, tak, że przedłożyłam mu powtórnie testy, które przed chwilą rozwiązywał źle i te także rozwiązał obecnie zupełnie dobrze. Na moje pytanie jaka jest przyczyna tej zmiany, odpowiedział, że ten pan zupełnie mu się nie podoba tak samo jak i jego szef w sklepie, u którego naumyślnie tak źle pracuje; ażeby go opuścić jaknajprędzej. Gdyby więc przypadkiem mój kolega kończył egzamin, ocena wypadłaby zupełnie fałszywie.

daniach obowiązują w równym stopniu, co we wszystkich innych stosunkach życiowych, nie powinny być tutaj również pomijane. Wywierają one wpływ na każde badanie.

Oprócz tych nieokreślonych bliżej, osobistych wpływów, oddziaływa jeszcze badający na wyczyn:

1) Przez swoje *zachowanie* się. W ostatnich czasach R. W. Schulte dokonał niewielkiej próby, celem ustalenia, jaki wpływ wywiera jego przemowa podniecająca, budząca zaufanie badanego do samego siebie, a przemowa krytyczna, zabijająca zaufanie do samego siebie. 19 osób miało rozsegregować 100 kart na małych przegródkach i miało wykonać tę pracę (między innemi), pozostając pod wpływem sugestywnym raz hamującym, drugi raz podniecającym. Okazało się prawie wyłącznie polepszenie wyczynu po przychyłnej sugestji, pogorszenie wyczynu po nieprzychyłnej sugestji. „W odniesieniu do zakresu zmienności między-osobniczej wszystkich osób badanych, wpływ ten należy uznać jako 20% — 30%; może on się jednak podnieść nawet do 100%, jakto wykazały poszczególne wypadki¹⁾).

2) Przez swą znajomość lub nieznanomość technicznego obchodzenia się z testami. Test musi być przede wszystkim dobrze przygotowany, powtórnie dobrze przeprowadzony, po trzecie dobrze opracowany. Użycie testów, szczególnie aparatuowych wymaga wielu wiadomości i wprawy, które zdobyć można tylko przez doświadczenie. Nie rzadko spotkać można psychotechników, którzy nie umieją się dobrze obchodzić z aparatami do badania, a jeszcze

¹⁾ Untersuchungen über den Einfluss des Versuchsleiters auf das Prüfungsergebnis. IndPste. 1 (9), str. 289 — 291.

mniej rozumieją psychologiczne znaczenie testu, którego używają¹⁾. Również nie mogą oni zdać sobie sprawy z tego, jak silnie oddziałują każda drobnostka na rezultat badania. Jaskrawym przykładem tego jest *dawanie instrukcji*. Jeżeli badający nie potrafi ująć jej w możliwie najjaśniejsze, ściśle wyrażenia, to będzie źle zrozumiany. Jeżeli daje instrukcję dwu badanym, których wyczyny ma ze sobą porównać, nie w ten sam sposób, nie w tym samym tonie (np. jednemu szeroko wyjaśnia, drugiemu podaje instrukcję krótko i zwięźle), to wytwarza warunki różne dla rozwiązania zadania i t. p. A. Mayer wykonał jedno badanie dwukrotnie; raz instrukcja wzywała: „pracuj prędko i dobrze”, drugi raz: „pracuj prędko”; wykazał on w ten sposób, że zależnie od instrukcji powstały dwa różne nastawienia u uczniów, które odbiły się wyraźnie na ich wyczynach²⁾. K. Lewin dał dwóm grupom terminatorów dwie instrukcje: „szukaj stosownego klucza” i „szukaj dwóch stosownych kluczy”; rezultat wypadł całkiem różnie: pierwsza próba nie zgadzała się, druga natomiast zgadzała się w zupełności z wrażeniem subiektywnem co do inteligencji osobnika badanego. H. R. Crosland wykazał oddziaływanie rozmaitych instrukcji: „możliwie

¹⁾ Słuszne jest, że nad tem zagadnieniem dyskutowano w Ameryce i stawiano propozycję, jak zapobiec złu; zob. prace: Geissler L. R. A plan for the technical Training of consulting psychologists. JApplPs. 2 (1) 1918, „Communications regarding A Plan for the technical training of consulting psychologists” ibid 2. (2), str. 174 — 178 i „Further Communication regarding A plan for the technical training of consulting psychologists” ibid. 2 (3), str. 280 — 285.

²⁾ Mayer A. Über Einzel-und Gesamtleistung der Schulkinder. ArchgesPs. 1903. Tom I.

„jaknajszybciej“, „możliwie jaknajdokładniej“ „uważaj na treść“ — na wykonanie korekty¹⁾).

Nie mniej wpływają na badanych niezamierzone sugestje. Binet wypowiedział bardzo ważne słowa: „Un mot, un simple mot, dit par l'expérimentateur, peut changer complètement les dispositions mentales du sujet“²⁾). Co więcej Binet twierdzi, że nieuwzględnianie sugestji, płynącej z rozmów, prowadzonych podczas badania, jest podobnem zaniedbaniem jak u bakterjologa prowadzenie kultury bakterjologicznej w brudnym naczyniu³⁾). Wynika z tego, że badania, dokonywane w obecności większej ilości osób, narażone są na silne zarzuty. G. S. Gates przeprowadziła znamieny eksperyment w tej sprawie. Trzem grupom osób badanych dała dwa testy. Jedna grupa pracowała tylko w obecności badającego, druga przez połowę czasu pracowała również w jego obecności, a przez drugą połowę w obecności czterech do sześciu osób, trzecia przez drugą połowę czasu pracowała w obecności 27 — 30 osób. Okazało się, że najlepsi badani najczęściej w swych wyczynach zostali zahamowani przez obecność obcych⁴⁾).

Do dobrego operowania techniką testów należy dar dobrej obserwacji, ażeby stwierdzić sposób, w jaki badany rozwiązuje zadanie. Badani szczególnie, jeśli chodzi o inteligentniejszych, próbują rozwiązywać trudniejsze

¹⁾ Harold R. Crosland: An investigation of proofreaders illusions. University of Oregon Publication 2 (6) 1924.

²⁾ Binet: De l'intelligence, Paris, Alcan 1922. (Neue Auflage), str. 300.

³⁾ ibid.

⁴⁾ Gates, Georgina Stickland: The effect of an audience upon performance: JbnPs. 28 (4) 1924.

zadania zapomocą „tricków”. Zauważono np., że badani szybko zorjentowali się przy tremometrze Moedego, że głębsze wciśnięcie sztyłtu powoduje najczęściej wyczynny błądne, prowadzili go więc po samej powierzchni płyty, a następstwem tego był naturalnie błąd w ocenianiu ich wyczynu. Giese wspomina, że przy ćwiczeniach restytucyjnych dobre rezultaty nie miały nic wspólnego z wprawą, natomiast badany wyuczał się odpowiedniejszego i zręczniejszego układania ręki na guziku¹⁾, służącym do badania reakcji.

Wyniki testów zmieniają się zależnie od sztuki oceniania, którą posiada badający. Przedewszystkiem istnieje cały szereg badań, które w zupełności poddane są subiektywnemu sądowi badającego (prawie wszystkie t. zw. próby zręczności, jak np. wyginanie drutu, wycinanie i rysowanie figur według wzorów, testy inteligencji jak np. definicje, tworzenie zdań z podanych słów i t. d. Alicja Descœudres w pięknym doświadczeniu zademonstrowała niepewność sądu subiektywnego. Kazała ona 24 osobom psychologom i studentom ocenić 154 odpowiedzi dzieci na trzy testy. Najostrzejszy sędzia ocenił 33% odpowiedzi jako dobre, a jego łagodniejszy kolega 68%. Między temi skrajnościami znajdowały się najrozmaitsze oceny. Nie znalazło się ani dwóch osób, któreby oceniły zupełnie jednakowo²⁾, ³⁾). Także i testy aparatuowe podlegają sub-

¹⁾ Giese: Handbuch psychotechn. Eignungsprüfungen 1925, str. 755.

²⁾ Archives de Psychologie. Tome XVI.

³⁾ Eksperyment Descœudres obala więc celowość propozycji, uczynionej niegdyś przez E. Borela, by tworzyć sąd ogólny według sądu kilku osób (E. Borel: Le calcul des probabilités et la methode des majorités. AnPs. 1907).

jektywnej ocenie. Często się słyzy zdanie, że ponieważ aparat automatycznie wskazuje wynik badania, albo ponieważ dla każdego rozwiązania testu już wyliczono wartość — przeto zadanie badającego polega poprostu na notowaniu wartości. Tymczasem odczytanie wyczynu na aparacie nie odbywa się poprostu i całkiem bez wątpliwości. Weźmy jakiś prosty przykład: każemy badanemu zareagować na pewną podniecie i mamy odczytać na stoperze czas trwania reakcji; wskazówka stopera może stanąć pomiędzy 0,6 a 0,7 sekundy. Od badającego zależy, czy wybierze 0,6 czy 0,7. Taka ocena nie jest przypadkowa; już dawno przekonano się, że istnieją typy, oceniające w kierunku plus albo w kierunku minus; przyczem jeden podwyższa, a drugi obniża wyczyn (tendencje do przeceniania i niedoceniania). Nie pominiemy tutaj także, że wśród badających, np. nauczycieli, sędziów i t. d. niekiedy występuje specjalna skłonność (czasem pochodzenia sadystycznego), ażeby jak najwięcej osób „ściąć”. Z drugiej strony znowu, zdarzają się „litościwi”, którzy wzdrygają się przed udzielaniem złych ocen, ażeby kandydatowi nie szkodzić.

Kingsbury wymienia jeszcze tendencję do zbyt szybkiej oceny, niechęć do rewizji oceny, wreszcie tendencje do tego, ażeby osoby, dla których jedno wyliczenie wypadło nisko lub wysoko, oceniać wogóle nisko lub wysoko¹⁾.

Osądzanie innych jest w wysokim stopniu zależne od osobistych uczuć sympatji lub antypatji, które badany wzbudza w badającym; wtedy prawie niechęący i nieświadomie wypada ocena przychylniej lub nieprzychylniej.

1) Forrest N. Kingsbury: Analyzing ratings and training raters. WeRes 1 (1) 1922 cyt. według ZangsPs. 23, str. 376.

Także przypadkowe czynniki jak np. zły humor, niezdro-
wie, zmęczenie, wpływy atmosferyczne, troski osobiste,
zniecierpliwienie przy końcu eksperymentu i t. d. odbijają
się w ostrzejszej lub łagodniejszej ocenie¹⁾. Piéron,
który zajmował się tą przypadkowością oceny, ujął swój
sąd w takie sarkastyczne zdanie: „C'est que la machine
humaine à juger n'est pas une machine parfaite”²⁾. A
wreszcie badanie zależy również od tego, czy badający
posiada znajomość ludzi, czy jest zdolny wczuć się w ba-
danego, w jego indywidualność, w jego istotę, w jego chwi-
lowy nastrój. Dobry znawca człowieka będzie zawsze do-
brym badającym.

6. CELOWA POSTAĆ BADANIA.

Z wszystkich tych rozważań, które przytoczyliśmy
w toku powyższych wywodów, wyłania się wniosek, że
test powinien przybierać taką postać i być tak wycehowa-
nym, żeby wykluczał wszelkie źródła błędów, a jaknajle-
piej spełniał swoje właściwe zadanie; jest niem dżagnoza
i prognoza co do uzdolnień człowieka. Dlatego to często
stawiano pytanie, jakie cechy powinien posiadać dobry
test. Zbieramy tutaj w jedno odnośne propozycje rozma-
itych autorów.

1) Konstrukcja testu musi być ekonomiczna, jego wy-

¹⁾ Marbe mówił na ten temat w dyskusji na „Posiedzeniu
Psychotechnicznym” w r. 1923, że jest tendencją, stwierdzoną sta-
tystycznie, że sędzia w wypadkach przestępstw za które prawo
przewiduje karę więzienia od 2 do 20 lat, z zamiłowaniem stosuje
liczby równe (2, 4, 6) albo wyznacza zazwyczaj kary 5 i 7 lat.

²⁾ H. Piéron: La Psychologie dans l'Orientation Profession-
nelle. W tomie zbiorowym: Dix conférences sur l'orientation prof.
Paris, Alcan, 1923, str. 51.

konanie nie może wymagać specjalnych urządzeń, które byłyby uciążliwe, albo skomplikowane, nie może również wymagać długich wyjaśnień. Im test jest bardziej zrozumiały, im mniej skutkiem tego wymaga dokładnych wyjaśnień badającego, im prostsze są warunki wykonania, także im łatwiej jest skontrolować zachowanie wszystkich warunków przez badanych nawet podczas badań masowych, — tem bardziej celowa jest postać testu.

Kwestją ogromnej wagi jest przytem także ocena testu: czy test nie wymaga skomplikowanych metod obliczania, w których łatwo jest o błąd, a które zabierają wiele czasu i wysiłków.

Oto pytanie, które trzeba uwzględnić wtedy, kiedy niema się pomocników przy badaniach, gdyż niekiedy z tego powodu musimy zrezygnować z testu, choćby nawet posiadał on wielką wartość djagnostyczną. Z tych powodów praktycznych należy uważać za lepsze wyliczanie rezultatów od mierzenia rezultatów, mierzenie zaś za lepsze od oceniania jakościowego, które często wymaga długiego namysłu przed wydaniem sądu; stosownie do tego należy testom nadawać odpowiednią postać.

2. Należy wziąć pod uwagę stronę *gospodarczą* testów. Nie chodzi tu o zagadnienie kosztów badania, ponieważ zawsze zależy to od zakładu przemysłowego, czy chce onłożyć wydatki na badanie indywidualne, czy przez oszczędność poleca przeprowadzać badania masowe. Ale zarówno w badaniu pojedyńczem, jak zbiorowem, można używać testów, które wymagają niewielu materiałów, a temsamem obniżają koszty badania. Szczególniej Rupp kładł nacisk na gospodarczą stronę testów i podkreślił w tej sprawie kilka stanowisk. Nie zawsze można się z nim zgodzić, kiedy np. każe, o ile możliwości, unikać ba-

dań indywidualnych, ponieważ za bardzo absorbują badającego; ale w wielu punktach ma Rupp słuszność; wtedy mianowicie, kiedy żąda uproszczenia konstrukcji aparatów, a temsamem obniżenia ich ceny; kiedy żąda, aby aparatów nie budować z większą dokładnością, aniżeli jest to potrzebne. Tak np. optometr z dokładnością do 1/1000 mm. używany w bardzo wielu zakładach, jest zupełnie niepotrzebny. Drogie przedmioty winno się zastępować rysunkami (formularzami) łatwymi do powielania, tak np. w zadaniu: szukaj zatyczek odpowiednich do otworów". Także w samem badaniu powinno się stosować odpowiednie oszczędnościowe metody. Dla miary w oku np. zamiast *metody odtwarzania*, czyli zamiast odtwarzać na aparacie odcinki określonej długości — można zastosować *metodę cceniania*, przy której należy uszeregować i porównać ze sobą stałe podniety rozmaitych stopni. Używanie kołeczków, ciężarków, sprężyn i t. p. zamiast niektórych aparatów, w których skład wchodzi powyższe części jest o wiele tańsze¹⁾. Giese np. celem zaoszczędzenia materiału zamiast dawać do wykonywania zadania rachunkowe na przygotowanych drukach używał zadań już wykonanych z błędami, przyczem kandydat miał wskazać błędy.

Także ważnem zagadnieniem jest czas *trwania* testów. Psychologowie niejednokrotnie zastanawiali się nad tem, o ile krótki test może dać wyobrażenie o zdolnościach lub wyczynachadanego. Hansen i Rean wykonali w tym celu następującą próbę: dwie grupy studentów, jedna złożona z 48, druga złożona z 80 osób otrzymały jeden test (test ten był to powiązane w swych czę-

¹⁾ H. Rupp: Wirtschaftliche Durchführung von Eignungsprüfungen. Der Maschinenbau — Betrieb 5 (22) 1922/23).

ściach zadanie, którego pełne wykonanie wymagało 25 minut). Osoby badane obowiązane były po każdych 5 minutach zrobić znak ,ile w przeciągu tego czasu wykonały. Określono potem w punktach każdy okres 5-cio minutowy i przekonano się, że wyczyny ,wykonane w ciągu 5 minut, były proporcjonalne do wyczynów 25 minutowych. Eksperymentatorzy wyciągnęli stąd następujący wniosek „The intelligence test can be made short or long according to the time of the disposal of the examiner”. Różnice wyczynów wychodzą najaw już w pierwszych 5 minutach. Dłuższego testu należy używać na to, ażeby uzyskać całkowitą pewność wyników¹⁾.

Z ową zasadą „gospodarczości” należy jednak być ostrożnym. Mogłaby się w pewnych warunkach sprawdzić na niej owa anegdota o koniu, który właśnie zdechł wtedy, kiedy w myśl życzenia swego pana odzwyczaił się jedzenia. I tak praktycy, wyposażeni w dar obserwacji, podnoszą, że „metoda odtwarzania” wywołuje u kandydatów wrażenie samodzielnego czynu działania, co dla praktyki ma ogromne znaczenie. i chociaż w całym szeregu badań pojedynczych „metoda odtwarzania” da się zastąpić „metodą oceniania”, to jednak nie należy tego traktować jako reguły²⁾. Tak samo krótkie testy wprawdzie zaoszczędzają wiele czasu, ale tracimy przez nie sposobność do dokładniejszego i pewniejszego scharakteryzowania badanego.

Wyżej wymienione postulaty, dotyczące cechowania testu są zapewne najważniejsze. Ale i pod tym względem

1) Hansen C. F. et Rean M. J.: The productive value of short intelligence tests. JApplPs.-5 (2), str. 184 — 186, 1921.

2) Blumenfeld: Eignungsprüfungen an Lehrlingen der Metallindustrie im Psychotechn. Institut a. d. T. H. in Dresden. Maschinenbau-Betrieb. 5 (22) 1923, str. 249.

niema zgody. Już niejednokrotnie próbowano ustanowić kryterja „dobrego” testu. W. Stern jeden z pierwszych sformułował takie zasady w sprawie wymagań stawianych możliwie doskonałemu testowi albo serii testów. Według jego poglądu powinien test: 1) mieć możliwie najwyższą wartość symptomatyczną (musi stanowić możliwie jednoznaczne znamię badanej właściwości); 2. musi posiadać możliwie najszerszą wartość symptomatyczną (powinien charakteryzować badaną właściwość w całości); 3. winien zezwalać na umieszczenie badanego w szeregu rang) w naszych czasach określa się to wyrażeniem „test musi różnicować” albo „rozszewać”, to znaczy zezwalać na stwierdzenie możliwie jaknajliczniejszych różnic; 4. winien być łatwy do zastosowania; 5. musi być „obszerny” to zn. nadawać się do zastosowania na wielu najrozmaitszych osobach¹⁾. U Claparèda znajdujemy aż 18 wymagań, którym powinien odpowiadać idealny test: a mianowicie:

1. Test winien być interesujący.
2. Obiektywny w stosowaniu (t. j. niezależny od osoby badającej).
3. Umożliwiać jednoznaczną instrukcję (ażeby wszyscy ją jednakowo rozumieli).
4. Ocena jego powinna być obiektywna (wyników jego nie można oceniać subiektywnie).
5. Powinien wykluczyć przypadkowość (rozwiązanie powinno się opierać *wyłącznie* na badanych właściwościach).
6. Dający się stopniować.
7. Być wycechowany (na podstawie zbadania) wielkiej liczby osób).

¹⁾ W. Stern: *Differenzielle Psychologie*, str. 106.

8. Umożliwiać silne zróżnicowanie.
9. Dać się prędko rozwiązywać (wykonanie nie może zabierać dużo czasu).

10) Być jednoznaczny (tylko jedną właściwość powinien badać).

11. Nadawać się do powtarzania (nie powinien się stać wskutek powtarzania łatwiejszym).

12. Być nieznanym t. j. nie powinien być opublikowany, żeby go nie znał osobnik badany.

13. Nie może wymagać wiadomości szkolnych.

14. Winien być stały.

15. Dać się zastosować dla każdego wieku życia.

16. Umożliwiać stosowanie testów tego samego rodzaju (t. zw. testów równoległych).

17. Nie wymagać żadnych aparatów.

18. Badać tylko jeden składnik, a więc albo ilość, albo jakość, a nie obydwie składniki odrazu¹⁾.

O ile wiemy, żaden test nie został zbadany pod względem tych wszystkich wymagań.

Ogromna liczba testów, istniejących we wszystkich krajach, niektórym psychologom nasunęła myśl o unifikacji testów i używania tych samych testów we wszystkich krajach²⁾. Zajął się tem zagadnieniem Ch. S. Myers na II Kongresie Psychotechnicznym w Medjolanie (1922) (In-

¹⁾ Claparède: Comment diagnostiquer les aptitudes chez les écoliers, str. 105 — 106.

²⁾ Przyczyną tej różnorodności testów jest brak ciągłości pracy psychologów. Wydaje się prawie, jak gdyby każdy psycholog miał ambicję wprowadzania swoich własnych testów bez zbadania dotychczasowych. Łatwość skonstruowania testu dopomaga jeszcze w wysokiej mierze temu brakowi ciągłości.

ternational Unification of Vocational Tests¹⁾), wysuwając przeciw takiej unifikacji następujące argumenty: przez tłumaczenie byłyby te testy, choćby pod względem języka, tak przekształcone, że stałoby się niemożliwe ich rezultaty porównać ze sobą jako jednoznaczne. Różnice w wychowaniu i umysłowości każdego kraju sprawiają, że nie można używać wszędzie tych samych testów, ponieważ cechy, właściwe inteligencji, które zawsze są związane z procesami uczuciowymi, nie dadzą się ująć zapomocą tego samego testu, jednakowo u Skandynawa, Włocha lub u wyspiarza z Fidżi. Dla tego też Myers, zamiast za unifikacją, opowiedział się raczej za wyborem testów, którego winna dokonać komisja międzynarodowa, specjalnie utworzona do tego celu. W dyskusji, która nastąpiła potem, Fr. Baumgarten zwróciła uwagę na to, że wprowadzenie unifikacji testów nie jest możliwe, natomiast powinno się ujednolicić *metodę* używania testów, gdyż odnośne publikacje w rozmaitych krajach wykazują daleko idącą różnorodność warunków, w których test został wycechowany i zastosowany. Dlatego też należy żądać od każdego autora ogłoszonego, (a więc zaleconego do użytku) testu, żeby:

1. podawał przyczyny, które skłoniły go do wybrania nowego testu, zamiast używania poprzednio już znanego; w ten sposób podkreśli się symptomatyczną wartość testu;

2. ażeby przytoczył psychologiczną analizę testu (to zn. podał co rzeczywiście bada się zapomocą tego testu);

3. ażeby wycechował test na pewnej oznaczonej liczbie osób (liczbę tę należałoby uzgodnić przez międzynar.

¹⁾ Atti della 3 Conferenza Internazionale di psicotecnica applicata del' orientamento professionale. Milano 1923, str. 82 i nast.

dowy Komitet); przytem powinno się uwzględnić osoby a) rozmaitego wieku, b) rozmaitej płci, c) rozmaitych warstw społecznych;

4. ażeby podawał wyniki kontroli testów.

Taka standaryzacja metody pozwoliłaby oceniać wartość rozmaitych seryj testów¹⁾.

Międzynarodową Komisję testów ustanowiono w następnym roku (1923) na Międzynarodowym Kongresie Psychologicznym w Oxfordzie, o jej działalności jednak dotychczas nic nie wiadomo.

7. TECHNIKA BADANIA.

Na podstawie tego, cośmy dotychczas powiedzieli o czynnikach, wpływających na rozwiązanie testów, zrozumiałe jest, że poza nadaniem testom celowej postaci i samo pojęcie badania połączone jest z pewnemi specjalnemi warunkami. Pomijamy tutaj wpływy zmiany pogody i temperatury, jakoteż i wrażliwości na nie, ponieważ nie potrafimy działania tych wpływów zmierzyć, ani zarejestrować. Ograniczymy się tylko do warunków, których regulowanie i zmienianie leży w naszej mocy.

Oto te warunki:

1. Możliwie zaciszne miejsce, gdzie kandydatowi nie przeszkadzałoby, gdzie nie odwracałyby jego uwagi rozmaite szmery. Ponadto pożądane jest, żeby badany i badający byli sami, żądanie to odrzucają rozmaici Amerykanie. Między innymi L i n k²⁾ twierdzi, że robotnicy mu-

¹⁾ Atti... str. 83 — 84.

²⁾ Wogóle nie badano jeszcze wpływu otoczenia na badanie. Ciekawa jest przeto obserwacja, którą zrobił Charles Richet przy badaniu trzyletniego cudownego dziecka Pepita Arriola. Pepito mógł grać tylko na swoim instrumencie w swoim zwyczaj-

szą wykonywać swoje wyczyny wśród przeszkadzających szmerów i w obecności innych osób, przeto i badania winny odbywać się w możliwie „najnaturalniejszym” otoczeniu robotnika. Link urządził nawet przenośne laboratorium, które ustawiał w sali robotniczej i w niem wykonywał badania. Trzeba jednak zaznaczyć, że właśnie wtedy badał robotników już czynnych w zakładach przemysłowych, którzy przyzwyczaili się do takiego otoczenia tak, że na szmery nie zwracali uwagi. Jego propozycja zatem może mieć więc znaczenie tylko dla takich wypadków. Ale dla robotnika, niezatrudnionego dotychczas w fabryce, albo dla świeżo wstępującego terminatora takie laboratorium nie nadaje się, ponieważ w czasie krótkiego badania, nie potrafi on dostosować się do nowego otoczenia²⁾.

2. Badanie nie należy przeprowadzać w obecności innych kandydatów do posady, gdyż wówczas objawia się współzawodnictwo i działa ambicja. Działanie tych dwu czynników badano dotychczas niewiele, jasne jest jednak, że mamy tu do czynienia z czynnikami, działającymi bardzo rozmaicie. Moede¹⁾ w sprawie współzawodnictwa i Fr. Baumgarten²⁾ w sprawie ambicji mogli zaobserwować, że współzawodnictwo i ambicja u młodocianych, którzy bez tej podniety lepiej pracowali, pomniejszały wy-

nem otoczeniu; w obcym otoczeniu zawodził. Jeżeli zważymy, że pisarze, uczeni i t. d. tylko w pewnym otoczeniu i tylko w pewnych warunkach tworzą swoje najlepsze dzieła (porówn. wiadomość u Cima: *Le travail Intellectuel* Paris, Alcan 1924) to jasnym będzie, że dotychczas zbyt zaniedbaliśmy ten czynnik.

¹⁾ Moede: *Experimentelle Massenpsychologie* Leipzig, Hirzel 1920.

²⁾ Baumgarten: *Eine Ehrgeizprobe*. *PrakPs.* 3 (11) 1922.

czyny. Natomiast ci, którzy wykazywali przedtem złe wyczyny, podwajali, a nawet potrajali pod wpływem ambicji. Psychologja zahamowań zbyt mało jest zbadana, aby określić, czy uczucie ambicji nie wywołuje wewnętrznego niedocenywania się, co wpływa obniżająco na wyczyn; albo, czy nie mamy może do czynienia z różnemi rodzajami ambicji: z jedną wewnętrzną, immanentną, która nas trwale pobudza do wyczynów i drugą, wywołowaną widokiem czynności innych ludzi, która jest często powierzchowną. Jeżeli jednak przeprowadzamy badania nad pracą, którą badany w praktyce będzie wykonywać z całą grupą innych pracowników, to celowe będzie zbadać, w jaki sposób pracuje on w obecności innych ludzi.

3. Badający winien, o ile to możliwe, wprowadzić badanego w stan duchowego spokoju i swobody. Oprócz zachęty w przemówieniu wstępnem, (czy to przy badaniu indywidualnem, czy zbiorowem) dobrze jest wybrać na początek pytania najłatwiejsze, a jednak interesujące, ażeby zająć uwagę badanego. Takie próby wstępne nazywa się „schock—absorber”=„tłumikami strachu”¹⁾. Dopiero stopniowo powinno się wprowadzać zadania trudniejsze.

4. Niedobrze jest gdy kandydata badają rozmaici badający, jak to stosuje się w niektórych laboratorjach: badany idzie od aparatu, do aparatu, gdzie co chwila spotyka coraz to nową osobę. Mimowoli musi się nastawiać wtedy nietylko na nowe zadanie, lecz także na nową osobę i jest bardzo prawdopodobne, że jedna działa nań podniecająco, inna zaś hamująco. Tej okoliczności trzeba przypisać, że stopnie z różnych poszczególnych badań są często krańcowo różne. Jest to często nie dysharmonja zdol-

¹⁾ Link: tłum. J. Witte, str. 79.

ności badanego, lecz różnorodność badających, która wyraża się w ten sposób. Wynik badania nie będzie skutkiem tego dawał właściwego jednolitego wyobrażenia o badanym. Suma takich poszczególnych sądów ma w pewnych warunkach o wiele mniejsze znaczenie, aniżeli obraz, który sobie jeden badający wyrobi o kandydacie w toku większej ilości badań.

Niebezpieczeństwo jednak grozi także, jeżeli ocenia i jeden badający. Jeżeli badany rozwiązał 2 testy fałszywie, to badający nastawia się na to, że i dalsze wyczyny będą fałszywe i wydaje sądy mniej przychylne. Przeciwnego nastawienia należy oczekiwać wtedy, gdy pierwsze testy zostały rozwiązane dobrze.

5. Badającemu nie wolno wogóle ani gestami, ani mianiką, ani żadnym innem uzewnętrznieniem afektów dawać do zrozumienia, czy odpowiedź badanego jest trafna, czy fałszywa. W tym drugim wypadku można kandydata onieśmielić. Dlatego należy unikać takich aparatów, na których kandydat odrazu poznaje zły wynik. Do nich należy więc np. tremometr Moedego, w którym każdy błąd zostaje zdradzony przez dzwonięcie. Szczególniej dzieci, jak to łatwo zauważyć denerwują się tem i wykonywują zadanie o wiele gorzej, aniżeli wtedy, gdy są spokojne. Co innego jest naturalnie w badaniu specjalnem, w którym czysto rzeczowo zwraca się uwagę kandydata na błąd w tym celu, aby stwierdzić, czy potrafi się poprawić, to znaczy, czy potrafi uświadomić sobie swe zadanie, wystąpić czynnie.

Christiaens wyraża przekonanie, że w testach ma się osądzać nie tylko obecny wyczyn, lecz także możliwość udoskonalenia go (wyraża to: „c'est la perfectibilité qu'on cherche"). Przez poprawianie wyczynu kandydata,

które następuje po zwróceniu uwagi na poprzedni błąd, uzyskuje Christiaens obraz zdolności do ćwiczenia danej właściwości u badanego osobnika; dodamy jeszcze od siebie: uzyskuje obraz postawy badanego wobec wyników swojej pracy, co jest charakterystyczne dla *typu* pracownika.

W związku z tem należy wspomnieć o ciekawej pracy Marty Moers, która badała skutki zawiadamiania o błędzie przy sposobności badania miary w oku. Doszła do wyniku takiego, że wogóle zdawanie sobie sprawy z błędu pomniejsza znacznie wielkość błędu; ale z jej dzieśięciu osób badanych dwie wykazywały także powiększenie błędu potem, gdy się dowiedziały, że popełniają błędy¹⁾. Czyli i tutaj zachodzą różnice indywidualne.

6. Instrukcja, t. j. żądanie, aby dany test wykonać w pewien określony sposób, musi być ułożona jaknajjaśniej i jaknajprecyzyjniej; badający musi sobie dokładnie uświadamiać, jak działa instrukcja; (zob. rozdział poprzedni). Badający musi również przekonać się, czy badany zrozumiał dokładnie zadanie. (Przez to zresztą wyjaśni sobie, czy kandydat posiada wogóle zdolność rozumienia zleceń).

7. Należy wymagać *standaryzacji warunków testu*, t. zn. ten sam test powinien być dawany w dokładnie tych samych warunkach wszystkim badanym lub wszystkim grupom badanych (pora dnia, oświetlenie, temperatura). Jeżeli badanie przeprowadzamy w ten sam sposób, ale w rozmaitych warunkach, to uzyskamy rezultaty całkiem

¹⁾ Moers M.: Ein Beitrag zur Untersuchung der Augenmasprüfung ZAngPs. 23 (5/6), str. 270, 273.

różne. Słusznie mówi Whipple: „We cannot always make the conditions ideal, but we can at least try to keep them constant¹⁾).

8. Ważne jest, by badający panował nad swoim złem usposobieniem i nie ulegał jego wpływowi przy ocenianiu badania. Ponadto winien on być jaknajbardziej skoncentrowany na samej rzeczy; rozproszenie badającego udziela się łatwo osobnikowi badanemu.

9. Ale najważniejsze jest to, żeby badający nie zacieśniał uwagi na szczegółowy wyczyn, lecz w badaniu miał na oku *całego* człowieka. Albowiem dla oceny wyczynu ważny jest nietylko fakt rozwiązania lub nierozwiązania testu, lecz także, „jak” został wyczyn dokonany. Poppelreuter mówi na podstawie swych doświadczeń w sposób bardzo znamieny: „moim współpracownikom... brak poprostu zupełnie wiadomości o człowieku, dlatego często stawiają najdziwniejsze błędne dżagnozy. Wogóle poza czystą spostrzegawczością”, „wrażliwością dotykową”, „kombinacją” i t. d. nie widzą człowieka²⁾). Sąd ten odnosi się również do większości badających. Nie ma dość silnych słów, by wpoić w osoby prowadzące badania — przekonanie, że test jest tylko środkiem, prowadzącym do tego celu, aby poznać człowieka. Ma on być wyrównaniem naszej słabej zdolności obserwowania. A każda osoba prowadząca badania powinna uwierzyć głęboko w ten fakt podstawowy, że nawet zły

¹⁾ Whipple G. M. Manual of mental and physical tests, New York 1910, str. 4.

²⁾ Poppelreuter: Allgemeine Richtlinien der praktisch-psychologischen Begutachtung. Leipzig 1923. str. 78.

test w rękach dobrego badającego jest lepszy, niż dobry test w rękach złego badającego.

8. PRÓBA PRACY.

Wielkie braki testów, które skłoniły Whipple'a do powiedzenia, że niema ani jednego takiego testu, do któregoby miał absolutne zaufanie — doprowadziły do stosowania pewnej doskonalszej metody badania, a mianowicie do *próby pracy*.

W próbie pracy nie chodzi o badanie poszczególnych wyczynów np. pamięci, lecz o dokonanie pracy w tej samej skomplikowanej formie, jaką posiada ona w codziennej praktyce, jak np. pakowanie, sortowanie, ważenie, czyszczenie, fałdowanie, znaczenie i t. d. Jest to więc zlecenie praktycznej pracy, istotnego zajęcia, czyli jest to sytuacja rzeczywista. Wykonanie takiej próby pracy wymaga współdziału woli, uczucia, inteligencji, zezwala więc na stwierdzenie całego zachowania się badanego i wnioskowania o jego cechach charakterologicznych; kiedy w badaniu testami idzie o to, czy i w jakim stopniu wykonane rozwiązanie, to w próbie pracy idzie głównie o to, by zaobserwować, jak kandydat przystępuje do pracy, jak próbuje ją wykonać i t. d. Próba pracy więc daje sposobność do jakościowego zaobserwowania kandydata. Można przytem stwierdzić typy woli, nastawienia uczuciowego, inteligencji, właściwości, jak przezorność, pedanterję, organizację, spokój, nerwowość, flegmatyczność, niecierpliwłość, wytrwałość, gnuśność, stanowczość i t. d. Na sposób zachowywania się badanego można wpływać przez pewne techniczne warunki pracy, jnk np. skrępowanie pewnem tempem pracy (przebieg przymusowy) powtarza-

niem (monotonja), pracę jednostkową lub kolektywną. W ten sposób uzyskujemy ogólną, albo pełną ddiagnozę badanego.

Prób pracy używał już Binet w badaniach nad dziećmi, każąc im wykonywać określone zlecenia¹⁾. Klasyczną próbą pracy jest również znana próba rachunkowa Kraepelina²⁾, w której ma się do siebie dodawać liczby w jednym ciągu. Ale na pełne znaczenie próby pracy dla badań zdolności do zawodu zwrócił pierwszy uwagę Poppelreuter³⁾ i użył jej do oceniania zdolności do pracy u rannych w głowę. Wybitny użytek z tej metody zrobił Giese w swych badaniach zdolności u ludzi dorosłych i poświęcił temu zagadnieniu kilka prac⁴⁾. Obecnie prawie każdy psychotechnik używa prób pracy w swoich badaniach.

Nie każdy jednak zabieg roboczy z życia może służyć jako próba pracy. Th. Valentiner, który pierwszy przeprowadził systematyczne badanie młodocianych zapomocą prób pracy, kładzie nacisk na to, że i „próby pracy”, podobnie jak testy, mają rozmaitą wartość symptomatyczną, którą oznaczyć można tylko drogą empiryczną. Dlatego też i do nich stosuje się jako najważniejszy i konieczne wymaganie:

¹⁾ Binet A.: Le développement de l'intelligence chez les enfants. AnPs, 14, 1908.

²⁾ Kraepelin: Die Arbeitskurve. Leipzig 1902.

³⁾ Poppelreuter W.: Die psychischen Schädigungen durch Kopfschuss. Leipzig, Voss, 1917/18. Tom II.

⁴⁾ Giese: Zur Betriebsführung psychotechnischer Prüfstellen PrakPs. 3. (1922). Der Arbeitsgedanke in psychologischer Diagnose. Arbeitsschule 37, 1923. Die Arbeitsprobe. ZangPs. 23 3 (4) 1924.

Przed użyciem nawet najlepiej obmyślanych prób pracy dla celów praktycznych, należy ustalić ich wartość przez praktyczne sprawdzenie¹⁾.

Typologia pracy dzięki „próbom pracy“, niestety, nie wzbogaciła się dotychczas w sposób godny wzmianki.

¹⁾ Valentiner Th.: Zur experimentellen Feststellung der berufswichtigen Willenseigenschaften bei Jugendlichen. *PrakPs.* 4 (1) 1922.





SPIS RZECZY.

	Str.
Przedmowa do wydania polskiego	III
Przedmowa do wydania niemieckiego	V
Wstęp	1
Część pierwsza ogólna	25
I. Terminologia	25
II. Pojęcia zasadnicze	32
1. Psychiczne zróżnicowanie ludzi	32
2. Właściwości i zdolności psychofizyczne	41
A. Definicje	41
B. Podział zdolności	49
C. Stadia rozwoju zdolności	51
3. Właściwości charakteru	59
4. Skłonności (zamiłowanie) do zawodu	67
5. Budowa właściwości i zdolności	67
A. Teoria struktur	74
B. Teoria typów	82
C. Korelacja	85
D. Dynamika zdolności i wyczynów	94
III. Psychologiczne badanie pracy zawodowej	98
1. Definicja pracy zawodowej	100
2. Klasyfikacja zawodów	105
3. Metody badania zawodów	116
A. Metoda swobodnej charakterystyki	116
B. Metoda ankiety i wypytywań	117
C. Analiza zawodów	127
a) Opis analityczno-suma ryczny	127
b) Analizy empiryczne	129
c) Psychotechniczne studjum pracy	139
d) Metoda wykonywania pracy przez samego ba- dacza	141
e) Metoda kolektywna	145
D. Metody eksperymentalne	146
a) Metoda eliminacyjna	147
b) Metoda korelacji	147

	Str.
c) Metoda błędnych wyczynów	148
d) Metoda najwyższych wyczynów	150
e) Metoda patologiczna	152
f) Eksperymentalne badanie nienormalnych warunków pracy	157
g) Naukowa metoda Kitsona	158
E. Metoda charakterologiczna	159
F. Metoda typologiczna	160
G. Metoda struktury	162
H. „Najlepsza” metoda	163
IV. Dobór przy przyjmowaniu personelu	166
1. Zwyczajne sposoby przyjmowania	166
2. Postępowanie naukowe	185
Metoda testów	185
A. Definicja	185
B. Rodzaje testów	185
3. Wartość testów	199
A. Symptomatyczna i diagnostyczna wartość testów	199
B. Zdolność różnicująca testów	212
4. Ocenianie wyczynów, uzyskanych zapomocą testów	215
A. Ocenianie punktami	215
B. Ocenianie testów aparatuowych	229
C. Jakościowa ocena wyczynów	233
D. Tworzenie oceny ogólnej na podstawie metody testów	235
E. Ustalanie norm	241
F. Graficzne przedstawienie wyników badań	248
5. Źródła błędów w badaniu testami	255
A. Niestalość osobnika badanego	256
B. Nastawienie i zmiana nastawień osoby badanej względem badania	259
C. Wpływ osoby badającej	262
6. Celowa postać badania	270
7. Technika badania	277
8. Próba pracy	283

