

2. Historisches	1
2.1. Vor der Veröffentlichung des BINET-SIMON-Tests	1
2.2. Von der Veröffentlichung des BINET-SIMON-Tests bis zum Ersten Weltkrieg.....	2
2.3. Vom Ersten bis zum Zweiten Weltkrieg.....	3
2.4. Vom Zweiten Weltkrieg bis heute.....	5
2.5. Fragen zur historischen Entwicklung von Tests.....	6

2. Historisches

Drei wichtige Perioden in der Entwicklung des Testens sind zu verzeichnen:

- BINET-SIMON-Test
- Erster Weltkrieg
- Zweiter Weltkrieg

2.1. Vor der Veröffentlichung des BINET-SIMON-Tests

Ort der Entwicklung der Idee des Testens war die Psychiatrie in Frankreich und Deutschland. Die Ideen des Arztes Pinel (1794) waren der Anstoss, dass Geisteskranke nicht mehr als Verbrecher, sondern als Kranke betrachtet wurden. Esquirol (1838) unterschied weiter zwischen Geisteskrankheit und Schwachsinn, wobei letzterer hauptsächlich durch einen Mangel an intellektuellen Fähigkeiten charakterisiert wurde.

Seguin befasste sich mit der Erziehung schwachsinniger Kinder und entwickelte ab 1848 einen Test zur Prüfung motorischer und sensorischer Funktionen - den "Performance-Test".

In Deutschland versuchte Ebbinghaus 1897 die Ermüdbarkeit in einem Lückentext zu messen.

In der Psychiatrie wurden also Methoden gesucht, zwischen verschiedenen Geisteskrankheiten zu unterscheiden (Differentialdiagnose) und die verschiedenen Ausprägungsgrade des Schwachsinnns zu bestimmen.

Die frühe deutsche experimentelle Psychologie wirkte förderlich, aber auch hemmend auf die Entwicklung der Testpsychologie. Genaue Beschreibung der Versuchsbedingungen, Kontrolle der Variablen und genaue Verarbeitung der Resultate waren fördernd. Die Unterschiede zwischen den Versuchspersonen wurden allerdings nicht ausgewertet, da man ja an allgemeinen Gesetzmässigkeiten interessiert war. Die interindividuellen Differenzen wurden als Fehler behandelt.

1879 gründete Wundt in Leipzig sein Institut für experimentelle Psychologie mit dem Ziel der allgemeinen Beschreibung sensorischer und motorischer Qualitäten.

Gerade aber die interindividuellen Unterschiede in der Reaktionsdauer standen im Interesse von Cattell, der 1890 in einem Zeitschriftenartikel zum erstenmal das Wort "Test" gebrauchte. Bereits drei Jahre später, 1893, beschäftigte sich die APA (American Psychological Association) damit, Tests zu registrieren und Anwendungsmöglichkeiten zu beschreiben.

Neben der Psychiatrie und der experimentellen Psychologie wurde die Entwicklung des Testens auch durch die Genetik beeinflusst. Angeregt durch Darwin's Buch "On the origin of species" (1859), in dem die Erbllichkeit von körperlichen Eigenschaften behandelt wurde, versuchte der englische Biologe Galton entsprechende Gesetze für psychische Eigenschaften nachzuweisen. Durch Galton's Arbeit ergaben sich drei wichtige Gesichtspunkte für die Testentwicklung. Neben der Forderung, individuelle Verschiedenheiten zu untersuchen, sah er die Notwendigkeit einer Systematisierung der Untersuchungsverfahren, und schliesslich beschrieb er seine Ergebnisse in den mathematischen Formulierungen der Wahrscheinlichkeitsrechnung als Abweichungen, "deviations", vom Mittelwert. Sein jüngerer Kollege Pearson perfektionierte die Auswertungen. Auf ihn gehen Korrelations-, Faktorenanalysen und weitere Verfahren zurück.

Zusammenfassend kann man festhalten, dass in der Psychiatrie Tests für die Art und den Grad geistiger Abnormalität benötigt wurden, dass aus der experimentellen Psychologie die standardisierten Untersuchungsmethoden übernommen wurden, und dass in der Genetik das Interesse für menschliche Verschiedenheiten entstand. Die Ergebnisse waren zwar theoretisch interessant, wurden aber kaum praktisch angewandt, was sich erst durch die Idee Binets änderte.

2.2. Von der Veröffentlichung des BINET-SIMON-Tests bis zum Ersten Weltkrieg

Binet war es schon sehr früh klar, dass die Intelligenz nicht nur mit einfachen sensorischen und motorischen Prüfungen gemessen werden konnte. Er erhielt 1904 vom Unterrichtsministerium den Auftrag, eine Möglichkeit zu finden, um zwischen Schülern unterscheiden zu können, die nicht lernen konnten und die nicht lernen wollten. Die 30 Aufgaben, die er mit SIMON zusammenstellte, waren stark auf die Anforderungen der Schule ausgerichtet, z. B.:

- Benennung von Gegenständen
- Ergänzung unvollständiger Sätze
- Verstehen von Sätzen

Die Aufgaben wurden an 50 normalen und einigen schwachsinnigen Kindern ausprobiert, und eine Rangreihe der Aufgabenschwierigkeit wurde erstellt. Der erste praktisch angewandte Intelligenztest war erfunden.

In einer weiteren Version wurde der Begriff des "Intelligenzalters" zur Beschreibung der geistigen Entwicklung eingeführt. Der Test wurde schnell in mehrere Sprachen übersetzt, und William Stern setzte das Intelligenzalter mit dem tatsächlichen Lebensalter in Beziehung.

Terman entwickelte den Test in seiner amerikanischen Fassung weiter ("Stanford-Binet-Test"). Sein Test enthielt nun standardisierte Instruktionen und Normen einer repräsentativen Stichprobe, sowie den Intelligenzquotienten ($IQ = \frac{\text{Intelligenzalter}}{\text{Lebensalter}} * 100$) zur Beschreibung des Intelligenzniveaus des untersuchten Kindes.

Daran anschliessend kam immer wieder die Frage auf, was nun eigentlich die Intelligenz in Wirklichkeit ist, die man so gut erfassen konnte. Spearman wollte das Problem empirisch lösen und verwendete die durch Pearson möglich gewordene Faktorenanalyse zur Untersuchung der Testaufgaben. Er ermittelte zwei Faktoren, einen generellen (g-) Faktor, der in allen Intelligenztests zu beobachten ist, und einen spezifischen (s-) Faktor. Thurston stellte dem später seine "multiple Faktoretheorie" gegenüber.

Andere Forscher wie Thorndike (ab 1910, vgl. 1949) in den USA und Burt (ab 1913, vgl. 1962) in England beschäftigten sich mit der Erfassung von Schulerfolg. Hier ging es nicht um die Anlage, wie bei der Intelligenz, sondern um das Niveau der Kenntnisse und des Schulfortschritts.

In dieser Periode wurde versucht, objektive Beurteilungen von vor allem Schulleistungen zu erzielen. Gruppentests und Validitäts-Untersuchungen, bei denen die Testresultate mit äusseren Kriterien verglichen werden, gab es nicht.

2.3. Vom Ersten bis zum Zweiten Weltkrieg

Wie in vielen Bereichen der Wissenschaft war traurigerweise auch in der Testpsychologie der Krieg Anstoss für neue Entwicklungen. In Deutschland, England und Frankreich wurden grosse Menschengruppen Tests unterzogen, um Autofahrer, Piloten oder ähnliche Gruppen zu selektieren.

In den USA bestand durch die heterogenen Bevölkerungsgruppen eine sehr komplexe Situation für die Entscheidung der Auslese bzw. Verwendung für den Kriegsdienst. Die APA entwickelte einen "paper-and-pencil" Intelligenztest: den "Army-Alpha-Test". Dieser erste Gruppentest enthielt mehrere Arten von Aufgaben: Rechenaufgaben, Denkaufgaben und Worterklärungen. Die Zuverlässigkeit und der Vorhersagewert in Bezug auf Eignung waren befriedigend, und der Test wurde ständig weiterentwickelt.

Nach dem Ersten Weltkrieg wurden in Europa eher individuelle Beobachtungstests zur qualitativen Beschreibung der Person eingesetzt (Phänomenologische Betrachtungsweise).

In Amerika wandte man sich eher den quantitativ auswertbaren Gruppentests zu (bedingt durch das behavioristisch-positivistische Weltbild). Die Integration der vielen Einwanderer war die gesellschaftliche Aufgabe, die es zu lösen galt. Da die Testaufgaben in der englischen Sprache formuliert waren, bestand allerdings keine Chancengleichheit für die Untersuchten. Bereits 1918 wurde deshalb als zweiter schriftlicher Gruppentest der "Army-Beta-Test" konstruiert, dessen Instruktion pantomimisch erteilt werden konnte und der keine Sprachkenntnisse voraussetzte.

Aber auch das individuelle Testen wurde weiterentwickelt. Terman und Merrill veröffentlichten 1937 eine neue Fassung des Stanford-Binet-Tests. Er hatte nun zwei Parallelformen, war neu normiert und hatte zusätzliche Aufgaben für das 2. Lebensalter bis zum Erwachsenen.

Die Klinische Psychologie setzte einen Intelligenzstrukturtest zur Diagnose verschiedener Krankheitsbilder ein, der von Wechsler 1939 aus 11 Untertests entwickelt worden war.

Praktisch parallel wurden in England die "Progressive Matrices" von Penrose und Raven (1938) entwickelt.

Bisher ging allerdings wie so oft die Anwendung der Tests der Theorie voraus. 1938 veröffentlichte Thurstone die Schrift "The reliability and validity of tests", in der er postulierte, dass die Resultate eines Tests zuverlässig sein sollen, d.h. bei wiederholter Vorgabe dieselben Ergebnisse erbracht werden sollen, und wenn man ein Kriterium wie z.B. Eignung vorhersagen will, müssen der Test und das Kriterium zusammenhängen (korrelieren).

Von Russland kamen keine Anstöße zur Entwicklung von Tests. Nach der damaligen ideologischen Vorstellung gab es keine angeborenen Verschiedenheiten in der Anlage der Menschen. Unterschiede wurden auf Motivation und Lerneinflüsse zurückgeführt.

Das Interesse galt eher der Entwicklung von Mittel und Wegen, die Lernleistung anzuheben.

Neben der Entwicklung von Intelligenztests gewannen auch die Persönlichkeitstests immer mehr an Bedeutung. Für das Testen der Persönlichkeit gibt es drei Möglichkeiten:

- Beobachtung
- Persönlichkeitsfragebogen
- projektive Verfahren

Die damalige Beobachtung hatte qualitative Aspekte im Blickpunkt, konnte aber den Gütekriterien von Objektivität und Zuverlässigkeit nicht genügen.

Einer der ersten Persönlichkeitsfragebogen war der "Woodworth Personal Data Sheet" als Vorläufer des 1940 erstellten MMPI (Hathaway & McKinley, 1940, 1943). Er war nichts anderes als ein standardisiertes, schriftlich erhobenes, psychiatrisches Interview. Die so erhaltenen Informationen konnten allerdings leicht verfälscht werden, zudem mangelte es auch hier an der Theorie zu den Tests.

Unstrukturierte Reizsituationen (Bilder, Kleckse usw.), auf die der Untersuchte frei reagieren soll, sind charakteristisch für die projektiven Verfahren. 1921 entwickelte der Schweizer Psychiater Rorschach den gleichnamigen, wohl berühmtesten projektiven Test. 1935 stellten die Amerikaner Morgan und Murray den "Thematic Apperception Test" (TAT) vor.

In dieser Periode wurde auch die Graphologie weiterentwickelt. Sie blieb allerdings bis heute eine empirisch nicht fundierte Methode.

2.4. Vom Zweiten Weltkrieg bis heute

Ironischerweise war es wieder ein Krieg, der Zweite Weltkrieg nämlich, der einen weiteren Aufschwung des Testens brachte. Nicht allerdings in Deutschland, denn hier war die "Reinheit der Rasse" ein wichtigerer Prädiktor als Leistung oder Veranlagung.

In England wurden durch die notwendige Auswahl und Verteilung der Soldaten Ämter für die Personalauswahl errichtet, die nun nicht nur Tests, sondern auch die Biographie, Interviews und Empfehlungen in die Entscheidung miteinbezogen. In den USA wurden die meisten führenden Psychologen für die Selektion, Testkonstruktion und Diagnostik eingesetzt. Im "Army General Classification Test" (AGCT) wurden nun nicht nur Intelligenzniveau untersucht, sondern viele verschiedene Fähigkeiten wurden in Testbatterien festgestellt.

Nach dem Krieg ging es in demselben rasanten Tempo weiter, und es entstand eine ganze Flut von Tests. Buros "Mental Measurement Yearbook" (1938, 1941, 1949, 1953, 1959, 1965) ist nur eine von vielen Bücherreihen, die versuchen, der Informationsflut Herr zu werden.

Aber erst durch den Einsatz von Computern kann die Auswertung von Testbefunden rationell durchgeführt werden. Die USA hatten auch hier die Vorreiterrolle. Die meisten deutschen Verfahren sind aus dem amerikanischen übertragen, wie z.B. der HAWIE und der HAWIK. Der IST von Amthauer (1953) hingegen ist ein originär deutsches Verfahren. Blättert man das von Hiltmann (1960) erstellte Kompendium psychologischer Tests durch, fällt auf, dass die objektiven, standardisierten Tests in der Minderheit sind gegenüber Beobachtungstests und projektiven Verfahren. Hierin zeigt sich die ablehnende Einstellung gegen die Psychometrie, die lange Zeit im deutschen Sprachraum verbreitet war.

Zusammenfassend und etwas vereinfachend kann man sagen, dass für die Testtheorie und die Testkonstruktion das praktische Bedürfnis nach Selektion und Platzierung sowie die positivistische Einstellung der amerikanischen Testpsychologen und natürlich der Einsatz von Computern die Grundlage für die rasende Entwicklung waren.

2.5. Fragen zur historischen Entwicklung von Tests

1. Wer hat den Begriff "Intelligenzquotient" eingeführt, und wie kam es dazu?
2. Durch welche gemeinsame Methode entwickelten Spearman und Thurstone ihre unterschiedlichen Vorstellungen der Struktur der Intelligenz?