More than a century ago, psychology was defined as the science of human mental manifestations and phenomena. However, it was psychometrics (the science of measuring these mental manifestations and phenomena) that made psychology scientific. Thus, psychometrics is a purely psychological area of research.

From a historical point of view, psychology branched out from philosophy as an independent university discipline at the close of the nineteenth century. It all started in Leipzig in 1879. Here the philosopher Wilhelm Wundt (1832–1920) established his psychological laboratory at the university. Formally, however, his laboratory remained under the faculty of philosophy. Wundt succeeded in detaching psychology from philosophy, especially freeing it from the influence of Emanuel Kant, an extremely influential philosopher who stated that it is impossible to measure manifestations of the mind in the same way as physical objects (5). With his criticism of pure reason, Kant (1724–1804) established the very important distinction between ‘the essential nature of things’ (things in themselves) and ‘things as they seem’ (i.e., that which we sense or perceive as a phenomenon when faced with the object we are examining).

Figure 1.1 illustrates Kant’s philosophical approaches with reference to present day psychiatry, according to which depression is understood to be a clinical phenomenological perception (shared phenomenology of depressive symptoms) as measured by the six depression symptoms contained in the Hamilton Depression Scale (HAM-D6, see Figure 3.1). Modern neuropsychiatry attempts to describe the depression behind the phenomenological perception, i.e., depression ‘in itself’, as we believe it to be present in the brain, for example, as a serotonin 1A receptor problem (impairment).

The area of research now known as brain research is just such an attempt to measure the processes presumed to be taking place in the brain, that is ‘das
Ding an sich. As pointed out by Sontag, reality has increasingly grown to resemble what the camera shows us (6). It is reality itself when the neuropsychiatric camera demonstrates receptor binding in the brain, while clinical reality is increasingly becoming what the camera visualises for us by means of assessment scales or patient-related questionnaires.

The ability to describe reality as it is in itself, i.e., looking at the world unclouded by any preconception of it, has been debated by such neo-Kantians as Wittgenstein and Quine (7). The quantification of endophenotypes or deep phenotypes is probably the most scientific image of the world. However, we do not have endophenotypes to tell us whether we indeed can describe reality, e.g., the brain, as it is itself. Wittgenstein tells us that he does not want to say whether we can or cannot describe reality as it is in itself. He wanted, as stated by Putman to bring our phenomenological items back to their home in clinical psychiatry. This is what clinical psychometrics is about (7).

Figure 1.2 shows a correlation between the so-called psychotic symptom items in an American rating scale (see Appendix) and serotonin 2A receptor binding, which it is now possible to measure by means of positron emission tomography (PET) scanning (8). The figure shows a correlation coefficient of $-0.57$; this is statistically significant but not clinically significant, as the variance on the ordinate axis (the 'psychosis' scale) can explain only about 32% of the variance on the axis of abscissas (serotonin 2A receptor binding). If the two patients at the far left are excluded as outliers, then the negative correlation value is halved, so that less than 10% of the variance is explained.
The scale in Figure 1.2 shows the positive symptoms in a schizophrenia scale. In the early 1970s, the American psychiatrist Nancy Andreasen found it important to label those schizophrenic symptoms on which medication had an effect as positive. In clinical psychiatry, these were termed productive symptoms as they were often the reason for hospitalisation in a mental institution. Later on, Nancy Andreasen became interested in neuropsychiatric brain imaging methods [Computer Assisted Tomography (CAT scan), Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET)], which became available in the 1980s and 90s. However, in an interview from 2003, she had to admit that schizophrenia is probably not located in one specific section of the brain (9). Schizophrenia affects many different brain areas that cannot be visualised as ‘das Ding an sich’.

Wilhelm Wundt’s major achievement was to realise that mathematical models of ‘das Ding für uns’ can be used to measure the ‘shared phenomenology’ of the state one wishes to assess quantitatively. During his studies at the Heidelberg faculty of medicine, he obtained a degree in medicine. Wundt then participated in studies in the physiology of perception under Helmholtz (1821–94) and Fechner (1801–87). He observed that it was possible to get subjects to reliably assess sensory impressions when the conditions of the study were standardised, e.g., with increasing light or noise exposure.
Wundt’s philosophical basis was that each manifestation of the mind corresponds to a neurobiological substrate in the brain, but in his opinion the psychometric measurement of this manifestation of the mind should only focus on the psychological phenomena (das Ding für sich) and not include any biological elements in any way. He belonged to the branch of philosophy called non-reductive monism (corresponding to Harald Høffding’s critical monism, which maintains that manifestations of the mind cannot be reduced to purely biological variables) (10). On the other hand, it is of course possible to reduce certain manifestations of the mind to less complicated ones in an attempt to obtain the most reliable or objective measure. He felt that it would be possible in this way to make psychology scientific within the frame of its own descriptive realm, since psychological and biological methods of description are two different ways of viewing reality.

Wundt’s approach was that of descriptive psychology where the various dimensions consisting of individual items (symptoms) can be added to give a total score. He was excluding the immediate, peak-experiences detached from relations, e.g., the spontaneous, stimulus-unrelated, perception-like images in the religious experience of the child, actually referred to as ‘Sensus numinis’ (11,12). The clearest description of Wundt’s scientific approach based on his ‘Grundzüge der psychologischen Psychologie’ is found in Vannerus’ monograph (13).

The psychometric method developed by Wundt is probably the only specific psychological method identified in mental science, i.e., in scientific psychology (14). The two most famous scientists to emerge from Wundt’s psychological laboratory in Leipzig were Emil Kraepelin and Charles Spearman; both of them understood that psychological measurement (psychometrics) and biological measurement are two different ways of viewing nature.

**Emil Kraepelin: Symptom check list and pharmacopsychology**

Kraepelin (1856–1926) had just obtained his medical degree when he applied for a post at Wundt’s laboratory in 1882. As Wundt was unable to finance his salary, Kraepelin also had to take up a post as a locum at the local mental hospital in Leipzig. Thus, Kraepelin held an unsalaried position at the Wundt laboratory. Kraepelin’s purpose was to introduce scientific psychology into psychiatry so that his career as a psychiatrist would be furthered by his studies at Wundt’s psychological laboratory. In his job application to Wundt, he wrote that he would give a kingdom for a [research] topic; Wundt then gave him the opportunity to examine the influence of psychoactive substances
Classical psychometrics

Kraepelin’s symptom checklist from his Zählkarten (counting cards)

- Nervousness
- Restlessness
- Irritability
- Depression
- Psychomotor retardation
- Aggression
- Grandiosity
- Negativistic behaviour
- Hallucinations
- Paranoid ideas


**Figure 1.3** The assessment scale or checklist used by Kraepelin (10)

such as alcohol and the hypnotic drug chloral hydrate on volunteer research subjects. Kraepelin set out to demonstrate a dose response curve using reaction time measurements as the psychological response and psychoactive substances as the stimuli, so that increasing amounts of alcohol (number of drinks) led to lengthening reaction times. Since Wundt could see that Kraepelin had his heart set on psychiatry, he encouraged Kraepelin to employ this objective scientific method when subsequently assessing the various symptoms presented by patients suffering from mental disorders.

Kraepelin published his first Psychiatric Compendium as early as 1883. In this he attempted to focus on the symptoms presented in the different disorders (Compendium der Psychiatrie. Verlag von Amber Abel, Leipzig, 1883). After leaving the Leipzig laboratory and starting on his career as a psychiatrist in Munich, Kraepelin published several compendiums or textbooks on psychiatry. He revised his textbook almost bi-annually and in the 6th edition in 1899, he was able to describe two disorders with different symptom profiles: manic-depressive disorder and schizophrenia.

Figure 1.3 shows the checklist Kraepelin used when systematically monitoring his patients over several years in order to ascertain which symptoms possessed ‘shared phenomenology’ over this period of time. These are called
checklist symptoms, as Kraepelin only determined whether the symptom was present or absent. This type of scale is called a nominal scale. Using this method, Kraepelin was able to demonstrate that during a period of about six months, some patients presented with the first five or six symptoms in Figure 1.3, while in other episodes of shorter duration (up to three months) they had the next two symptoms (aggression and delusions of grandeur), along with restlessness, sleep disturbance and irritability. Between these episodes of depression or mania, these patients were discharged from hospital and were socially well-functioning. Other patients, who were often lifetime residents in asylums, had the last three symptoms in Figure 1.3. Kraepelin described them as suffering from dementia praecox (now schizophrenia), as the disorder typically started when they were about 20 years of age and was chronic in nature, often with an influence on intellectual functions as well. But these were consequences, not elements, of the schizophrenic symptomatology. Manic-depressive disorder, on the other hand, did not typically emerge at a specific age. Based on the original registrations by Kraepelin on his ‘Zahlkarten’ (counting cards) including the checklist symptoms in Figure 1.3, Jablensky et al made a comparison using the Present State Examination (PSE). From the PSE scores the ICD-9 diagnoses of schizophrenia and manic-depressive disorder can be made. In total Jablensky et al identified 721 patients assessed by Kraepelin and found a concordance for the diagnoses of schizophrenia and manic-depressive disorder of approximately 80% with the ICD-9 diagnoses (15).

In his thesis: ‘Über die Beeinflussung einfacher psychischer Vorgänge durch einige Arzneimittel’ (Jena, Fischer Verlag 1892), Kraepelin established the area of research he designated pharmacopsychology.

In the 8th edition of his textbook, written between, 1909–13, Kraepelin added reflections on the psychotherapeutic effects of certain drugs such as morphine, phenemal and chloral hydrate. However, he found that the effects of these drugs on schizophrenia and manic-depressive disorder were extremely poor. He was thus able to observe the spontaneous course of illness in these two disorders.

In the schizophrenic patient, as stated previously, the condition was unremitting, while manic-depressive disorder was characterised by episodes with specific symptoms and then periods between episodes of a year or more in which the patients were completely without symptoms and thus able to function normally. In these descriptions, Kraepelin determinedly avoided including the various theories on disease circulating at that time, such as hereditary elements, stress burden and so on.

Kraepelin’s textbooks were not widely known outside Germany, as the two world wars made German psychiatry less acceptable. His system only began to make an international impact after World War II, not least in the USA.
During his research at Wundt’s Leipzig laboratory, Kraepelin conceived the idea of establishing pharmacopsychology. He thought it important to describe the symptoms found to be reversible during a course of pharmacological therapy. However, as mentioned previously, no therapeutically adequate drugs were developed during Kraepelin’s lifetime, so this research area was scaled down. It is a fact of great interest that Kraepelin was among the first to propose the use of dose response comparisons as an essential pharmacological criterion when determining the clinical effect of a drug.

The Rorschach test

Until the breakthrough of modern psychopharmacology in the 1960s, Danish psychometric research was heavily influenced by the Rorschach Shape Interpretation test, published by the Swiss psychiatrist Hermann Rorschach in 1921. The Rorschach test consists of 10 symmetrical inkblots, which do not represent recognisable images per se, but are used as indefinite visual stimuli open to many different interpretations, in the same way as with abstract painting. No psychometric theory underlies this ‘inkblot test’, but in the hands of a trained psychologist it may provide an opening for the psychodynamic theories propounded with reference to Freud’s psychoanalysis. Psychoanalysis was an accepted method of treatment in psychiatry during the period between the two world wars. However, an inherent limitation of the Rorschach test is that the scoring is heavily dependent on the testing psychologist, so that the Rorschach test has very poor inter-observer reliability (agreement).

In Denmark experimental psychology with stimulus response trials dominated research. Alfred Lehmann (1858–1921) was the founder of experimental psychology in Denmark. He had worked together with Kraepelin at Wundt’s laboratory. He established a psychological laboratory at Copenhagen University in 1886; Kraepelin paid a visit to it in 1901. The first professor of clinical psychology at Copenhagen University, Lise Østergaard (1924–1996) used the Rorschach test in her doctorate thesis on formal thought disturbances in schizophrenia at the University of Copenhagen, but the clinical experience she had gained under the supervision of the consultant psychiatrist Erling Dein turned out to be more rewarding than her Rorschach results (16). In the introduction to her thesis, Lise Østergaard correctly states that Kraepelin with his symptom checklist was the first person able to delimit schizophrenia by its characteristic symptomatology. Kraepelin had emphasised that the symptom profile was rarely quite alike from one patient to another, but in chronic schizophrenics the course of their disorder was completely homogenous.

Lise Østergaard then adds that Kraepelin’s description of these patients could ‘have a rather sterile and external appearance’. She finds Kraepelin’s
mode of description ‘marked by the stiffness and paucity of nuances that characterised Germanic psychology (Wundt). Kraepelin was not open to the new currents in the psychology of his period (i.e., the psychodynamic theories).

However, Lise Østergaard was forced to conclude that it was Kraepelin’s consistent, clinical descriptions of psychiatric patients that made it at all possible to delimit both the schizophrenic as well as the manic-depressive disorder.

With the introduction of modern psychopharmacology, it became vitally important to follow Kraepelin’s clinical but somewhat sterile measuring of symptoms, and as a consequence psychometrics had to reject the Rorschach test on a scientific basis (lack of reliability and validity) and to go on to promote the use of symptom rating scales based on Kraepelin’s checklist.

Clinical reality, as described by Kraepelin at the start of the 20th century, was ousted by Freud’s psychoanalysis, and only reinstated in the 1950s when modern psychopharmacology appeared on the scene. This made the clinical reality Kraepelin had described perfectly obvious to everyone, as well as the fact that Freud’s clinical theorising had been dismissed. Because clinical psychology was so slow to realise this, its range became very limited. Thus, it is hardly a paradox that clinical psychiatrists were the ones to develop clinical rating scales.

Charles Spearman: Factor analysis and intelligence tests

In 1906, the English psychology student Charles Spearman (1863–1945) finalised his studies at Wundt’s laboratory with a PhD thesis, but in 1904 he had already published his first paper on the correlation method that was to become the starting point of factor analysis (17).

Spearman then moved back to England and took up a London professorship. His psychological field of interest was that of intelligence tests for use with primary school pupils. Spearman is generally regarded as the first actual promoter of psychometrics via his attempt to define certain dimensions of intelligence through factor analysis. His idea was to use mathematical factor analysis to identify the factors that make up the concept of intelligence. Factor analysis is a method by which one may get an indication of which tests belong together and which do not. Thus, it is not a method of measurement but a classification of the different tests (factor structures).

Worldwide, however, factor analysis was soon elevated to the status of an important psychometric proof of validity of a rating scale, i.e., that the scale was scientifically valid.
If it was possible to show by the use of factor analysis, which tests pointed the same way and which pointed in other directions, then a scientific analysis had been performed.

In 1927, using factor analysis, Spearman was able to identify two factors of intelligence: a general factor and a specific factor (18).

The principle of Spearman's factor model is first to compute the correlations between different intelligence tests, identifying those factors that best describe the connection. The weighting of the tests comprising a certain factor is computed (factor loadings). The first factor is usually a general factor. The second factor is a specific factor, which shows in which areas the person in question has their strong points.

An attempt to use the Spearman factor analysis tradition for empirical research with different intelligence tests showed that the model does not describe the real world. One of the problems was that factor analysis is very sensitive to the range of variance in the sample being tested. If the analysis is an attempt to determine factors in subjects who are all very intelligent (i.e., a very narrow range of variance), too many factors will be identified. In a very large population sample with very different levels of intelligence (i.e., a very great range of variance), usually only a single general factor emerges.

The fundamental element in factor analysis is the correlation coefficient. Computation of the first factor will provide a rough estimate as to the size of the correlation coefficients of the individual items in a scale; these are given as factor loadings. When all the items have positive factor loadings (as is the case with the first factor in Hamilton's Anxiety Scale, see Table 1.1) then a general factor is present (general anxiety factor in Table 1.1). Should one wish to ascertain whether some items have a higher mutual correlation coefficient (loadings) than others, then the second factor will provide this information, through its contrast between positive versus negative loadings. In Table 1.1 the psychic anxiety symptoms have positive loadings while the physical (bodily) anxiety symptoms have negative loadings. The sign direction in itself is of minor importance and should not be dwelt upon; as the significant element here is that the symptoms with the same sign have a higher mutual correlation than the items with the opposite sign. The result shown in Table 1.1 has a very high clinical validity when assessing antianxiety effect in a drug.

In short, it is the identification of the first two factors that is of clinical significance. Typically, the first factor will demonstrate that the symptoms selected obviously have varying degrees of positive correlations; therefore this factor is called the general factor. The second factor is the bipolar factor according to the factor analysis literature as it attempts to establish two discriminatory symptom groups, namely the group with positive factor
loads and the group with positive factor loadings. Hence this factor is called the bipolar factor. As this term has nothing to do with bipolar affective disorder, it is now labelled the dual factor. According to Spearman, in intelligence tests this dual factor would typically discriminate between people with language skills and people with maths skills.

### British versus American factor analysis

Spearman founded a special British approach to factor analysis, in which factor analysis is used to interpret the first two factors of a rating scale analysis (the general versus the dual). In contrast, an American approach rapidly emerged in which factor analysis was used to identify as many factors as possible. In the following, emphasis will be on the British method. The American tradition of factor analytical tradition particularly refers to Guilford’s classical monograph, which first appeared in 1936 (5) and in a revised version in 1954 (19).

In the American tradition, Thurstone (20) recommended noting down the factors in order to find more simple structures, while Guilford recommended an ‘orthogonal’ rotation, i.e., factors may not inter-correlate (must be at right
angles to each other). Cattell, on the other hand, suggested a less rigorous approach with the use of ‘oblique’ rotation, permitting a certain degree of inter-correlation between factors (21). This basic attempt to eliminate negative loadings through rotation is called ‘positive manifold’ (22). In contrast, British tradition advocates an initial simple description of the principal component analysis. According to this the entire core of Spearman’s factor analysis must be examined before performing any rotation. In this ‘Spearman’ algebra, the first factor (the principal component) is a general factor that indicates the degree of positive correlation among the different items in a scale. The second factor is frequently a bipolar or dual factor (i.e., with negative loadings on some items and positive loadings on other items). One might claim that the British tradition is less invasive, less ‘manipulative’ than the American.

When focusing on the landmarks in the development of factor analysis over the first 50 years, Vernon concludes that Hotelling’s principal component analysis is mathematically more accurate than Spearman’s method, but that its greater complexity implies tedious calculations (23). However, with the SSPS or SAS programs, a century after Spearman’s factor analysis, we may now actually start with Hotelling’s method before we perform all the many rotations within factor analysis. The paradox is that we have difficulty in understanding the mathematical superiority of Hotelling’s method over that of Spearman. Therefore we do not realise that the first and second principal components identified by Hotelling’s method are often sufficient. In other words, we are often unable to provide an argument for making all the rotations inherent in the factor analytic method.

**Harold Hotelling: Principal Component Analysis**

It was the American mathematician Harold Hotelling (1895–1973) who became the best advocate for the British (Spearman) algebra of concentrating on the initial simple correlation matrix, focusing on the first two factors; the general factor and the bi-directional factor.

Hotelling received his PhD in Mathematics from Princeton University in 1924. In 1927, he wrote a review in the *Journal of American Statistical Association* on the first edition of Fisher’s *Statistical Methods for Research Workers* and subsequently visited Fisher in London in 1929. In 1933, from his new base at Columbia University, Hotelling introduced his Principal Component Analysis as a pure mathematical approach to factor analysis in an attempt to simplify the structure of a large number of items in a rating scale (24,25) (see Calculus Example 1).

The best description of Hotelling’s Principal Component Analysis (PCA) has been made by Dunteman (26). PCA is an attempt to identify a few
components explaining most of the variance in the scores for individual items in a rating scale in the original sample. Because PCA is conducted on rating scales that contain items with some degree of positive inter-correlation, the first component might explain up to 50% of the variance while the second component explains 10–15% of the variance. PCA has no underlying statistical model, but employs a mathematical focus to explain the total variance in the item scores, thereby capturing most of the information within the items of the rating scale. The first (general) component is a straight line in the correlation matrix with closest fit to the total variance, and the second component is a straight line of closest fit to the residuals from the first principal component. Since both principal components are uncorrelated, each one makes an independent contribution to accounting for the variance of the original items. The correlations of items within the principal components are called loadings, a term borrowed from Spearman’s factor analysis. Whereas the eigenvalue of the first principal component is usually higher than 1.0, the eigenvalue of the second principal component need not be higher than 0.7 (26).

The first principal component must be orthogonal to the second component, which will have alternative loadings, i.e., as many negative as positive loadings (bi-directional, or dual), thereby contrasting the two groups of items that are mutually most correlated.

PCA should be clinically interpreted as a method of classifying items, rather than a method to validate the problems of measurement. The presence of a general factor or component is not an argument for summing all items of a rating scale so that the total score is a sufficient statistic for measuring severity on a dimension.

PCA is a way to group items according to the second, bi-directional component, for example into typical and atypical depression. In this context, Bertrand Russell’s ramified hierarchy of typology is the best way to illustrate the clinical meaningfulness of PCA (27). The example used by Russell is the definition of a typical \textit{versus} an atypical Englishman. It is clear that most Englishmen do not possess all of the properties that most Englishmen possess. Therefore, a typical Englishman, according to this definition, might be atypical. The problem raised by Russell is that the word ‘typical’ has been defined by a reference to all properties. It is in this situation that Russell introduced his ramified hierarchy in order to deal with the apparent circularity (27). Being a typical Englishman should not refer to the totality of properties (all potential items) but to a sub-totality of the predictive items for which over 50% of the properties are captured by the concept of a typical Englishman. The PCA can be considered as a method of ramified hierarchy in which the second component has identified the predicative items by contrasting items with negative and positive loadings.
In conclusion, with reference to Russell’s theory of typology, the general component or factor identified by principal component analysis is the description of being an Englishman, whereas the bi-directional second principal component or factor is the description of being a typical or an atypical Englishman by the contrasting positive versus negative loadings of the second, bi-directional factor, e.g., positive = typical and negative = atypical.

**Hans Eysenck: Factor analysis and personality questionnaires**

In the autumn of 1945, Eysenck (1916–1997) was appointed Chief Psychologist at the Psychiatric Institute in London, which is affiliated with the Maudsley psychiatric hospital (1).

Eysenck set out to evaluate the validity of the psychological tests used in clinical psychiatry in the late 1940s. This has quite neutrally been summarised by Schafer, who concludes that if the results of a psychological test diverge from the diagnosis made by the psychiatrist, this does not necessarily mean that the test is incorrect (28). A clinical diagnosis, e.g., in depression, was not at that time clear-cut, as psychiatrists often found it difficult to distinguish between neurotic and psychotic depression. This mirrored Kraepelin versus Freud in their understanding of ‘neurotic’ and ‘psychotic’. The above-mentioned Rorschach interpretation in schizophrenia is a good example of this (16).

In this connection, it is imperative to understand that Eysenck did not himself treat patients and that his contact with clinically experienced psychiatrists led him to perceive Freud’s psychoanalysis as both a theory of personality and a treatment model (1). Eysenck soon realised that as a treatment method, psychoanalysis lacked clinical effect. In his personality questionnaire studies, however, his reference frame was to be found in Freud’s and Jung’s psychoanalytic models of personality rather than in true clinical reality. In his trials with factor analysis, he adhered to Spearman’s British tradition by examining the first two factors (the general versus the dual), while using Hotelling’s principal component analysis.

As mentioned previously, it had become a tradition among psychologists to use the test constructed by the psychiatrist Herman Rorschach (1884–1922) (the Rorschach test). In the area of personality, Rorschach had discovered that vision can be influenced by the personality behind the ‘glasses’. He thus thought that coloured inkblots are especially stimulating for the extrovert personality (extroversion dimension), while non-coloured inkblots, with less movement of the figures, are connected to the introvert personality (introversion dimension).
Eysenck demonstrated that these Rorschach theories could not be empirically reproduced using the Rorschach test, as interpretations of the test varied a good deal from one psychologist to another. In the field of psychometrics, Eysenck adopted the position that it is important to work with consistent personality dimensions. Using an empirical approach, he demonstrated that it is possible to ask people what they are experiencing.

By using questionnaires, Eysenck was able to eliminate investigator influence on testing behaviour, and he felt that the use of factor analysis would ensure that the interpretation of the questionnaire response profiles would not be influenced by the interpretation of the individual investigating psychologist. Eysenck made use of lay subjects (initially often young men up before the medical board prior to military service), but rarely included patients with a valid diagnosis. His questionnaires had qualitative response options on a nominal scale, in which only a ‘Yes’ or a ‘No’ were required. One of the reasons for this was the limited capacity of the computers available in the 1950s and 1960s; nowadays, we have access to the necessary power when using quantitative response categories.

Eysenck drew on both Jung’s personality theory of extroversion/introversion (as used by Rorschach), as well as on Freud’s personality theory of neuroticism, as the basis of his psychologist approach.

As a psychologist working on a theoretical basis, Eysenck was not sufficiently aware of the fact that both Jung and Freud were primarily clinical experts. Thus, Freud perceived neuroticism as a particularly pronounced degree of normal behaviour, not as the qualitative remove from normal behaviour seen in schizophrenia or the psychotic forms of depression or mania. As shown by Kline (29), Eysenck attempted to validate his questionnaire dimensions, e.g., neuroticism and extroversion/introversion, within the field of learning psychology, not in the clinical reality that formed the basis of Freud’s and Jung’s theories.

Among these personality dimensions (30), Eysenck’s neuroticism factor proved the most definite (31). Figure 1.4 gives an abbreviated version of Eysenck’s Neuroticism Scale with the nine items that best show the structure of the anxious neurotic personality. Of the remaining questions in Eysenck’s Neuroticism Scale (23 in all), many are closely associated with depression. A psychometric analysis of Eysenck’s Personality Questionnaire (EPQ), based on a study with persons experiencing relatively rapid remission after posttraumatic stress (32) and a corresponding control group (N = 1353 persons), gave a Loevinger coefficient of homogeneity of 0.42, proving that it is acceptable to use the total score of the nine questions as a measure of neuroticism. Another study, with patients suffering from differing types of affective disorders, showed that only Eysenck’s neuroticism scale was in accord with an experienced psychiatrist’s assessment of the degree of neurosis (33).

Eysenck found that those persons specifically suffering from anxiety had a response pattern that was very sensitive to negatively formulated
Eysenck’s Neuroticism Scale

Item numbers in accordance with the EPQ (30)
The questions below address how you would describe yourself in general

<table>
<thead>
<tr>
<th>No.</th>
<th>Symptom</th>
<th>Yes (= 1)</th>
<th>No (= 0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>Are you an irritable person?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Are your feelings easily hurt?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>Would you call yourself a nervous person?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>Are you a worrier?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>Do you worry about awful things that might happen?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>Would you call yourself tense or “highly-strung”?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>Do you worry about your health?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>Do you suffer from sleeplessness?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>Do you worry too long after an embarrassing experience?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total score

Figure 1.4 Scoring sheet for Eysenck’s neuroticism questionnaire

questions – such as those dealing with symptoms: the higher the number of affirmative responses, the more neurotic the subject is.

When commencing his research with these questionnaires, Eysenck labelled the Rorschach test the idiographic method of measurement and his own questionnaires, the nomothetic method.

The idiographic method is concerned with what is of unique significance to one individual with no relevance for others and Eysenck therefore correctly stated that the idiographic method cannot be used in measuring, since to measure is precisely to observe individuals with reference to a common scale. In contrast, the nomothetic method centres on what can be measured. Eysenck’s use of factor analysis to prove the fact of a nomothetic measure is a paradox, because factor analysis is not a method of measurement. Thus, in modern research factor analysis is used in idiographic analyses, e.g., when describing an individual’s quality of life (34).

It is of great importance to understand that Eysenck’s intensive personality questionnaire research using factor analysis actually confirms Spearman’s results within the field of intelligence tests, in that especial focus should placed on the first two factors identified by the analysis. Thus, Eysenck found that the first factor was a general neuroticism factor (Figure 1.4), while factor 2 was a dual factor discriminating between extroversion versus introversion (30). It was Eysenck’s attempts to explain the remaining factors and to relate these to the psychoanalytic perception of personality rather than to clinical reality that blurred his results.
Around 1970, the American psychologist Charles Spielberger developed a questionnaire to measure anxiety (35). In this he attempted to discriminate between dispositional neurotic personality and present state anxiety. The first of these he termed ‘trait’ anxiety and the second ‘state’ anxiety.

Figure 1.6 shows Spielberger’s ‘trait’ scale with 9 items selected from the original 20. This selection is based on the criterion of clinical validity, so that it corresponds with Eysenck’s neuroticism scale (Figure 1.4).

Around 1990, an international consensus that a five-factor personality model could adequately cover the whole field was achieved among psychologists (36). This model is called ‘The Big Five’ (37). On the basis of this model, a questionnaire, the NEO-PI-R, was developed. The two first factors in ‘The Big Five’ are based on Eysenck’s EPQ and reflect Eysenck’s Neuroticism Scale and Eysenck’s Extraversion Scale. Neuroticism and Extroversion are usually referred to as ‘The Big Two’; however, the items in the NEO-PI-R do not adequately cover Eysenck’s original dimension. The abbreviated versions of Eysenck’s Neuroticism and Extroversion Scales (shown in Figures 1.4 and 1.5) are sufficient when measuring ‘The Big Two’.

Figure 1.7 shows the nine NEO-PI-R items that correspond most closely to Eysenck’s neuroticism from a clinical point of view as shown in Figure 1.4. Only five out of the nine items in Figure 1.7 are negatively phrased, so the four
### Spielberger’s trait anxiety scale

The statements below address how you would describe yourself in general.

<table>
<thead>
<tr>
<th>No.</th>
<th>Symptom</th>
<th>Yes (1,2,3)*</th>
<th>No (0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>I tire easily</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>I wish I could be as happy as others seem to be</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>I feel that difficulties are piling-up so that I cannot overcome them</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>I worry too much over something that really doesn’t matter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>I am inclined to take things hard</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>I lack self-confidence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>I try to avoid facing a crisis or difficulty</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>I take disappointments so keenly that I can’t put them out of my mind</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>I become tense and upset when I think about present concerns</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Degrees 1, 2 and 3 all give positive replies

**Figure 1.6** Scoring sheet for Spielberger’s trait anxiety questionnaire

### NEO items corresponding with Eysenck’s neuroticism dimension

The questions below address how you would describe yourself in general.

<table>
<thead>
<tr>
<th>No.</th>
<th>Symptom</th>
<th>Yes (1)</th>
<th>No (0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I am not the worrying type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>I scare easily</td>
<td></td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>I seldom feel anxious or uneasy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>I hesitate to show anger, even when apposite</td>
<td></td>
<td></td>
</tr>
<tr>
<td>91</td>
<td>I often feel tense and nervous</td>
<td></td>
<td></td>
</tr>
<tr>
<td>121</td>
<td>I seldom worry about the future</td>
<td></td>
<td></td>
</tr>
<tr>
<td>147</td>
<td>I do not see myself as especially unworried</td>
<td></td>
<td></td>
</tr>
<tr>
<td>151</td>
<td>I often worry about things that might go wrong</td>
<td></td>
<td></td>
</tr>
<tr>
<td>216</td>
<td>Even minor factors can frustrate me</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 1.7** Scoring sheet for modified neuroticism questionnaire (NEO)
remaining items must be 'flipped' when measuring the degree of neuroticism. When this is done Loevinger's coefficient of homogeneity is 0.42.

Max Hamilton: Factor analysis and rating scales

Hamilton (1912–1988) commenced his career as a psychiatrist just after World War II. He had the same starting point as Kraepelin, that of wishing to utilise psychometrics as a means of making clinical psychiatry more scientific in its approach. In 1945 he started working at the Maudsley Hospital in London – at the same time and at the same place as Eysenck. He actually attended Eysenck's PhD courses in factor analysis (1).

His approach was that psychometrics in clinical psychiatry should be considered a scientific discipline parallel to pharmacology and biochemistry. During his career, Max Hamilton was Associate Professor of Psychiatry at Leeds University from 1953–1957. These years saw the founding of modern psychopharmacology, beginning with the establishment of the antimanic effect of lithium compared to placebo, followed by the antimanic and antipsychotic effect of chlorpromazine. Such placebo-controlled, randomised, double-blind clinical trials became more and more common in Britain in the 1950s and Hamilton could see the need for reasonably brief rating scales to be used when measuring the effects of these new psychotropic drugs.

Hamilton held a position as research assistant at Leeds University Hospital from 1957 to 1960 while developing his two rating scales, the Hamilton Anxiety Scale (HAM-A) from 1959 (38) and the Hamilton Depression Scale (HAM-D) (39) from 1960. While Eysenck was interested in the more permanent features of neuroticism, Hamilton was only interested in the symptoms of anxiety or depression that appeared as signs of clinical disorders and were reversible through psychopharmacological treatment. Like Kraepelin, his opinion was that these symptoms provide the best impression of the anxious or the depressive patient.

With both of his scales, the HAM-A (see Figure 1.8) and the HAM-D (see Figure 1.9), Hamilton's purpose was to measure those mental and physical symptoms found by the patient and his or her relatives to be the greatest burden. Hamilton's goal was not to make a diagnosis, only to measure the severity of the anxious or depressive condition. So each week the question was how severe the symptoms listed in Figure 1.8 and Figure 1.9 had been during the past week. Based on these weekly assessments during a course of treatment with antianxiety or antidepressive medication, it would be possible to describe their clinical effects.

Just as Eysenck did, Hamilton made use of factor analysis to demonstrate the scientific value of his scales in his psychometric publications.
For the depression scale, Hamilton found a varying number of factors during his studies (Hamilton, 1960, 1967). The first study population was very homogeneous, namely, depressive patients who were so severely afflicted that they were hospitalised. In the next study, the patient population was more heterogeneous, consisting of depressive patients who were either hospitalised or attending an out-patient clinic. Hamilton could see that, in an increasingly homogeneous patient group, an increasing number of factors could be identified; an unfortunate consequence of the correlation method as a mathematical element of factor analysis.

With his anxiety scale studies in out-patients suffering from anxiety neurosis, Hamilton found a two-factor model in both the first trial using a 13-item anxiety scale (38) and in the next trial with the final 14-item version (Hamilton 1969) (40). Hamilton's factor analysis showed that the first factor was a general factor while the second factor was dual, as it had negative

### Hamilton Anxiety Scale HAM-A₁₄

1. Anxiety
2. Tension
3. Fears
4. Insomnia
5. Difficulties in concentration
6. Depressed mood
7. General somatic symptoms (muscular)
8. General somatic symptoms (sensory)
9. Cardiovascular symptoms
10. Respiratory symptoms
11. Gastrointestinal symptoms
12. Genito-urinary symptoms
13. Other autonomic symptoms
14. Behaviour at interview

See Appendix 5a for Manual

**Figure 1.8** Scoring sheet for HAM-A₁₄
loadings on the physical anxiety symptoms and positive factor loadings on
the psychic anxiety symptoms (Table 1.1). This was subsequently confirmed
by a French study using the HAM-A (41).

A major international trial with DSM-III panic attack patients confirmed
this HAM-A₁₄ two-factor model (42). On the basis of these results, Hamilton
thought that the first factor is general (i.e., all the symptoms in the scale
concur in measuring one dimension), providing enough evidence to use the
total score as a sufficient statistic. But Hamilton became less confident about
this conclusion when his anxiety scale was not able to distinguish between
placebo and an antianxiety drug (43).

The fact that the second factor in Hamilton’s anxiety scale is bipolar,
or dual, i.e., that some items have negative factor loadings and others have
positive factor loadings, is perhaps the most interesting element in the
factor analysis method. Factor loadings demonstrate the correspondence

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1*</td>
<td>Depressed mood</td>
</tr>
<tr>
<td>2*</td>
<td>Guilt feelings and self-depreciation</td>
</tr>
<tr>
<td>3</td>
<td>Suicidal ideation</td>
</tr>
<tr>
<td>4</td>
<td>Initial insomnia</td>
</tr>
<tr>
<td>5</td>
<td>Middle insomnia</td>
</tr>
<tr>
<td>6</td>
<td>Delayed insomnia</td>
</tr>
<tr>
<td>7*</td>
<td>Work and interests</td>
</tr>
<tr>
<td>8*</td>
<td>Psychomotor retardation</td>
</tr>
<tr>
<td>9</td>
<td>Psychomotor agitation</td>
</tr>
<tr>
<td>10*</td>
<td>Anxiety (psychic)</td>
</tr>
<tr>
<td>11</td>
<td>Anxiety (somatic)</td>
</tr>
<tr>
<td>12</td>
<td>Gastro-intestinal symptoms</td>
</tr>
<tr>
<td>13*</td>
<td>General somatic symptoms</td>
</tr>
<tr>
<td>14</td>
<td>Sexual interest</td>
</tr>
<tr>
<td>15</td>
<td>Hypochondriasis</td>
</tr>
<tr>
<td>16</td>
<td>Loss of insight</td>
</tr>
<tr>
<td>17</td>
<td>Weight loss</td>
</tr>
</tbody>
</table>

See Appendix 3a for Manual

Figure 1.9 Scoring sheet for HAM-D₁₇,
between symptoms and the factor in question, thus implying a psychological insight. This demonstration that the anxiety condition consists of physical and psychic anxiety symptoms with an equal distribution, seven physical and seven psychic anxiety symptoms in HAM-A, proved to be highly significant later on. Hamilton did not look into this because interest was centred on his depression scale in the period from 1969 to 1989. Factor analysis was not able to provide a psychological insight in depressive symptomatology through the factor structure of the HAM-D.

Factor analysis is a psychometric method that reveals a structure in an assessment scale, but not whether it is a dimension in which the total score is a meaningful expression of the severity of a condition. In his monograph on clinimetric methods, Feinstein uses Hamilton’s scales as examples of scales ‘produced by factor analysis’, however, without discussing the nature of this validation procedure (2).

Here it is important to understand that Max Hamilton built on Spearman’s and Eysenck’s factor analysis within the frame of the two explainable factors. Hamilton went on to demonstrate that (particularly in the HAM-A) the first factor is a general factor while the second factor is bi-directional, differentiating between somatic and psychic anxiety symptoms. This dualism between body and mind seems to underline the accepted custom of calling factor 2 a dual factor.

Factor-analytic studies with Hamilton’s Depression Scale have shown that the great difference between different clinical trial results is in the number of factors produced and their item loadings. In other words, the American factor-analytic tradition leads to inferior results. The British tradition (only interpreting the two first factors – the general versus the dual) would seem to result in a fair amount of agreement between different clinical trials. A recent landmark study in this respect is from the STAR-D analysis (44).

**Pierre Pichot: Symptom rating scales and clinical validity**

Pichot obtained his medical degree in Paris in 1947. When he, like Hamilton, chose psychiatry, his purpose was to use psychometrics as a scientific discipline on the same plane as pharmacology and biochemistry. Pichot therefore studied psychometrics at the Sorbonne in Paris immediately after getting his medical degree (3).

He took up a position as registrar at the psychiatric hospital Saint-Anne in Paris under Professor Delay, who was among the first to demonstrate the antipsychotic effect of chlorpromazine.
In 1972, Pichot made it clear that, from a psychometric point of view, using the HAM-D total score in studies on the antidepressive effect of a drug was a dead end. His reason was that factor analysis had not accepted the use of the HAM-D total score. Thus, Pichot did not acknowledge a demonstration of a general factor as sufficient evidence that the total score was a sufficient measure of the degree of depression.

Pichot had worked with the US rating scale, the BPRS (Brief Psychiatric Rating Scale), developed by Overall and Gorham (45). Drawing on a symptom pool of more than sixty symptoms, it had been demonstrated that the eighteen symptoms in Figure 1.10 were especially sensitive to change during a course of chlorpromazine therapy in psychotic patients and imipramine therapy in depressive patients. The BPRS is perhaps the most widely used psychiatric rating scale worldwide. This is perhaps because it is seemingly easy to use; see Kraepelin’s symptom list in Figure 1.3.

Pichot then recommended the use of the six BPRS depressive symptoms to measure the antidepressive effect of a drug. A major review of the BPRS some years later confirmed that Pichot’s depression factor was an independent factor in the BPRS (46).

Pichot had been brought up in the French school of psychometrics, founded by Alfred Binet (1857–1911) through his intelligence tests for primary school pupils. Binet’s starting point was that school teachers possessed the greatest knowledge about the intellectual abilities of their pupils in the different levels of primary school. So Binet enlisted the aid of the most experienced school teachers when choosing intelligence tests, instead of using Spearman’s factor analysis. Binet ‘outperformed’ Spearman, as the updated versions of Binet’s tests are now generally used.

In 1905 Binet declared that:

> Our aim is, when a child is put before us, to take the measurement of his intellectual ability, in order to establish whether he is normal or if he is retarded. For this purpose we have to study his present condition, and this condition alone… as a result we shall neglect entirely his aetiology… We shall confine ourselves to gathering together the truth on his present condition (47).

Pichot thus held the opinion that rating scales measuring antipsychotic effect, antidepressive effect, or antianxiety effect should be based on the clinical reality of the assessments of experienced psychiatrists and not on factor analysis (3). The version of the BPRS scale shown in Figure 1.10 is identical to Overall’s reference (The semi-structured Brief Psychiatric Rating Scale interview and rating guide) as to symptom description (48). The descriptions applying to absence of a symptom are taken from Turner’s
### Brief Psychiatric Rating Scale

<table>
<thead>
<tr>
<th>Brief Psychiatric Rating Scale</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Somatic concern</td>
</tr>
<tr>
<td>2</td>
<td>Anxiety, psychic</td>
</tr>
<tr>
<td>3</td>
<td>Emotional withdrawal</td>
</tr>
<tr>
<td>4</td>
<td>Conceptual disorganization</td>
</tr>
<tr>
<td>5</td>
<td>Guilt feelings</td>
</tr>
<tr>
<td>6</td>
<td>Tension</td>
</tr>
<tr>
<td>7</td>
<td>Mannerisms and posturing</td>
</tr>
<tr>
<td>8</td>
<td>Grandiosity</td>
</tr>
<tr>
<td>9</td>
<td>Depressive mood</td>
</tr>
<tr>
<td>10</td>
<td>Hostility</td>
</tr>
<tr>
<td>11</td>
<td>Suspiciousness</td>
</tr>
<tr>
<td>12</td>
<td>Hallucinatory behaviour</td>
</tr>
<tr>
<td>13</td>
<td>Motor retardation</td>
</tr>
<tr>
<td>14</td>
<td>Uncooperativeness</td>
</tr>
<tr>
<td>15</td>
<td>Unusual thought content</td>
</tr>
<tr>
<td>16</td>
<td>Blunted or inappropriate affect</td>
</tr>
<tr>
<td>17</td>
<td>Elation/euphoria</td>
</tr>
<tr>
<td>18</td>
<td>Confusion and disorientation</td>
</tr>
</tbody>
</table>

**Figure 1.10** Brief Psychiatric Rating Scale

**Mania scale**
- Grandiosity [8]
- Uncooperativeness [14]
- Hostility [10]
- Increased psychomotor activity [17]
- Intrusive behaviour
- Elevated mood

**Depression scale**
- Somatic concern [1]
- Anxiety, psychic [2]
- Guilt feelings [5]
- Tension [6]
- Depressive mood [9]
- Motor retardation [13]

**Schizophrenia scale**
- Emotional withdrawal [3]
- Conceptual disorganisation [4]
- Suspiciousness [11]
- Hallucinations [12]
- Unusual thought content [15]
- Blunted affect [16]

**Figure 1.11** The three BPRS subscales. In the brackets the item numbers as indicated in Figure 1.10
1963 version (49). The first 18 items make up the BPRS-18. Two extra items are included to allow measurement of mania (Figure 1.11).

A clinical validity analysis of the BPRS would result in a depression factor, a mania factor and a schizophrenia factor, as seen in Figure 1.11. The mania and schizophrenia scales are often combined in a general psychosis factor when assessing the clinical effect of antipsychotics.

In his final work, *Psychology Down the Ages* (1937) (50), Spearman writes that the correlation coefficient developed by Pearson and himself was exclusively comprehended and used in English-speaking countries. In France and especially in Germany it was refuted. Classical psychometrics, which is based on the concept of correlation in factor analysis, and Cronbach’s alpha are thus typically described in the major American standard works on psychometrics: Guilford 1936 (5), Guilford second edition 1954 (19); Nunnally 1967 (51); and Nunnally and Bernstein third edition 1994 (52) as well as Comrey 1992 (22).

It is precisely because these major monographs on factor analysis lie within the American reference frame that the interpretation of Hamilton Depression Scale results is so problematic; this American tradition lays stress on the number of factors, while the British tradition uses Ockham’s razor, i.e., the principle of simplicity, and focuses on the two first factors (the general versus the dual). Hamilton relied chiefly on Hotelling’s principal component analysis.

The English philosopher William Ockham (1285–1349) described the principle afterwards known as Ockham’s razor: the scientific community should only assume what is strictly necessary when working with a scientific hypothesis (the law of parsimony).

This was precisely Pichot’s point; that psychometric analysis of rating scales should avoid the use of factor-analytic methods, as in the American tradition. Such analysis should follow Binet’s model in using experienced psychiatrists as a test of clinical validity, and use item response theory models to determine if it is valid to sum the individual items as a total score. In Pichot’s opinion, Binet had used the same reasoning when developing his intelligence tests as that which lies behind the item analyses published by Rasch in 1960.