First publ. in: Mathematical Models for Social Psychology / ed. by W. F. Kempf and B. H. Repp. London, Wiley, 1977, pp. 14-58

Dynamic Models for the Measurement of "Traits" in Social Behavior

WILHELM F. KEMPF

1.1 The Formal Expression of Psychological Concepts in Psychological Theories

The history of mathematical models in psychology is longer than is often assumed. As early as 1837, Herbart argued for the mathematical formulation of psychological theories, and he attempted to apply to psychology the laws of Newton's mechanics. Although this undertaking was doomed to failure from the beginning, it is nevertheless worth mentioning, since it exemplifies the orientation of early psychological methodology towards the ideals of classical physics. Only a few years later, Gustav Theodor Fechner succeeded in formulating a mathematical relationship between psychological phenomena, Weber's law, that stood firm for a long time in the presence of empirical data. The essential progress represented by Weber's law was that it relied on empirical observations, while Herbart had still denied the possibility of an experimental science of psychology. With the development of experimental psychology and with the rapid advance of mathematical statistics in the first decades of the twentieth century, the application of mathematical models in psychology soon became customary and reached its first peaks in classical test theory and factor analysis.

In the first phase of development, the assumptions of the respective models were characteristically chosen by reason of mathematical simplicity and were only secondarily oriented towards the requirements of the psychological subject matter. At the same time, these models were designed for universal application and, therefore, were frequently misunderstood as "mere methods". Only recently, a clear distinction has been made between the model and the methods of factor analysis (e.g., KALVERAM, 1970a). Nevertheless, this first phase in the application of mathematical models in psychology was of great significance, since it was instrumental in establishing psychology as an empirical science.

Konstanzer Online-Publikations-System (KOPS)
URL: http://www.ub.uni-konstanz.de/kops/volltexte/2008/6024/
URN: http://nbn-resolving.de/urn:nbn:de:bsz:352-opus-60241

The second phase of development began around the end of World War II and grew out of a critique of classical test theory: It is connected with the names of Louis Guttman, Paul Lazarsfeld, and Georg Rasch, among others. Since the publication of Gerhard Fischer's book, "Psychologische Testtheorie" (FISCHER, 1968), it has also gained a foothold in the German-speaking countries. Its most characteristic feature is its increasing emphasis on certain considerations derived from epistemology and the philosophy of science, coinciding with the considerable sophistication that the general theory of measurement has reached in the meantime (cf. SUPPES & ZINNES, 1963; PFANZAGL, 1968). Apart from the undisputed success of psychological test theory, mathematical psychology remained the domain of a few specialists until a few years ago. Although there are few psychological investigations today that can do without the application of mathematical models, the majority of these applications is still restricted to very simple models whose model-like nature, moreover, is frequently overlooked. The mathematical assumptions that underlie many popular statistical methods are only rarely reflected upon and brought in relation to the substantive psychological theories that these methods serve to test. For example, the fact that significant interactions in a multifactor analysis of variance falsify the assumption of additivity of treatment effects rarely receives a psychological interpretation. The likely reason is that, in analysis of variance, the assumption of linearity is made on grounds of mathematical simplicity, and psychological theories usually are only verbally formulated and rarely precise enough to justify assumptions such as the additivity of treatment effects.

The resulting discrepancy between theory and method in psychology is illustrated, for example, by LAUTMANN's (1971) analysis of the concept of social norm. Although sociology and social psychology typically define social norms operationally as group means, Lautmann distinguishes three different concepts of norm:

a) a sociological concept: the norm as guideline;
b) a psychological concept: the norm as frame of reference;
c) a descriptive concept: the norm as average.

In Chapter 5, a probabilistic model for the measurement of social norms is presented which attempts to establish a clear relationship between the methods for measuring social norms and the psychological concept of norm. This model also leads to a clarification of the relation between the concepts, "norm" and "conformity", which so far has been largely undetermined (cf. BRANDT & KÖHLER, 1971). The application of graph theory to Festinger's theory of cognitive dissonance and to Heider's theory of cognitive organization (e.g., BOUDON, 1973) provides a similar example of how the precision of concepts in social psychology can be increased through formal rigor.

We thus arrive at the question: What can the formal expression of psychological concepts contribute towards constructing psychological theories, and what requirements must be satisfied by mathematical models in psychology? It is useful to distinguish here between theoretical and empirical models, where the latter can be further divided into explanatory and predictive models. To add a separate class of descriptive models seems to be unnecessary for two reasons: First, the models just mentioned also have descriptive functions, and secondly, it is doubtful whether purely descriptive models exist at all in psychology. Measurement models, at least, which are often quoted as examples of descriptive models, are generally better classified as explanatory models. As I have discussed elsewhere (KEMPF, 1974a), nearly every measurement transcends experience and therefore cannot be considered as purely descriptive.

An example of a *theoretical* model is game theory which does not strive to explain or predict the actual behavior in games but instead examines the properties of games defined by certain axioms. The significance of game theory for social psychology consists in that it permits the clear definition and analysis of conflict situations and thereby makes the behavior in such situations amenable to psychological analysis. Information theory has fulfilled a similar function in psychology.

Predictive models, on the other hand, serve to predict actual behavior. They are typically applied when certain decisions must be made (e.g., about methods of treatment in clinical psychology) and theoretical explaiations of the relevant behavior are not sufficient for a prognosis. Two examples are multiple regression and the related method of multivariate analysis (COLEMAN, 1964). The concept of validity in classical test theory should also be understood in this sense. Validity does not mean explicability but predictability by means of linear regression. However, investigations in decision theory by CRONBACH & GLESER (1965) have demonstrated that predictive models cannot be considered detached from their decisional context. At least the classical concept of validity, then, is in need of revision, and it seems necessary to generally reconsider the function of predictive models in psychology.

Predictions of behavior can also be derived from *explanatory* models. However, explanatory models are formalized theories and therefore must meet much stricter criteria than predictive models. The most important tasks of explanatory models in psychological research are:

a) to increase the precision of psychological concepts, and
b) to establish a clear relation between psychological theories and the methods employed in testing them.

Ideally, it should be possible to associate certain psychological concepts directly with certain classes of mathematical models, and to gain insight into the relations between the underlying concepts from comparisons between these models. However, mathematical psychology will lead to such strong psychological statements only when its models satisfy a number of prerequisites that go far beyond mere mathematical consistency.

One of the most important postulates for a model is that an unequivocal interpretation of its structure and its parameters must be possible. A model can be considered a formalized theory when *all* its assumptions have some psychological content. Only then it is possible to establish a clear connection between theory and model and to draw conclusions about the validity of the theory from tests of the model. Yet this requirement is often met insufficiently, e.g., when additional parameters are introduced whose only function is to improve the fit of the model to empirical data. Other models are so complex that the interpretation of the results of their application exceeds the psychologist's capabilities. For example, a methodologically satisfactory interpretation of the results of a factor analysis is usually prevented by the fact that it would have to take into account all possible scale transformations, i.e., all rotations of the matrix of factor loadings (cf. KEMPF, 1972b).

The second postulate to be set forth here concerns the mathematical manageability of the model. The following operations must be possible and must lead to unambiguous outcomes:

a) estimation of the parameters in the model;
b) comparisons between the parameters (if the theory to be formalized makes statements about relations between parameters); and
c) tests of the structure of the model.

These criteria can again be demonstrated with factor analysis as an example. KALVERAM (1970a) has shown that the general factor model with p common factors and q specific factors allows the actual computation of a factor analysis only under the assumption – which is not empirically testable – that all intercorrelations between the specific factors and all correlations between specific and common factors are zero. Even then, it is not guaranteed that the true factor structure will be discovered, since the communalities are not unequivocally determined by the covariance matrix. Nevertheless, computation of a factor analysis invariably leads to some outcome, and this is the reason why many authors consider it a tautological model, as the following quotation shows:

> "Indeed, there are many examples of models that cannot be falsified, e.g., factor analysis in psychology. Whatever the available data may be, factor analysis always provides a solution which is accepted or rejected on the basis of criteria that do not depend on an agreement between the model and the observations, since this agreement is by definition always present." (BOUDON, 1973, p. 50; translated from German.)

It can be shown, however, that this view derives from a confusion between factor-analytic models and factor-analytic methods. For example, the often-criticized dependence of factor-analytic results on sample characteristics (FISCHER, 1968) does not lie in the model itself but follows only when the prerequisites of the model are not met. These prerequisites include all assumptions made in coping with the communality problem (cf. FISCHER & ROPPERT, 1965; KALVERAM, 1970b; KEMPF, 1972b). Independence from sample characteristics can thus serve as a criterion for the validity of the assumptions of the model, although it must be kept in mind that the communality problem is still unsolved in the general factor model; and restricted factor models, which surmount this problem, are based on very rigorous assumptions. In practical applications, it is very unlikely that restricted factor models will resist attempts of falsification. Therefore, these models – which I have thoroughly discussed elsewhere (KEMPF, 1972b) – turn out to be rather "unrealistic" (FISCHER, 1968).

A model is called "unrealistic" when it can be refuted by almost any data. It is "realistic" when sets of data can be found for which it needs not be rejected. The progress that the formalization of psychological concepts could bring to psychological theories will depend on whether realistic models can be constructed. The dilemma of mathematical psychology consists in finding models that can be tested and at the same time have a chance of resisting attempts of falsification. This has led to a clear preference for probabilistic models over deterministic models in psychology, which should not distract from the fact that there are also deterministic theories in psychology that have been successful, such as SCANDURA's (1973) theory of structural learning. At present, there seems to be a general trend towards more restrictive models in psychology, which includes not only deterministic models but also recent advances in probabilistic test theory (cf. FISCHER, 1968) which have led to a general probabilistic theory of observations in psychological experiments (MICKO, 1970; SCHEIBLECHNER, 1971a; FISCHER, 1972; KEMPF, 1972a). However, in contrast to the restrictive factor models mentioned above, the restricted nature of these models derives not only from mathematical necessities but primarily from psychological and metatheoretic considerations. The better a restrictive assumption can be justified, the smaller is the danger that it will prove to be unrealistic, and the greater is the gain in knowledge that results from an eventual falsification of the model. For models that rest on arbitrary assumptions, empirical tests are a necessary evil that decides on their applicability. Models whose assumptions represent psychological concepts, on the other hand, become important instruments for testing psychological theories.

1.2 Conceptualizations of Individual Differences in Social Psychology: "Aggressiveness"

The assumption of individual "behavioral dispositions", which underlies the psychology of individual differences and personality, is one of the most ancient and widespread attempts to explain human behavior. In social psychology, examples can be seen in concepts such as "attitude" or "aggressiveness", or in the (ultimately unsuccessful) attempt to solve the problem of leadership in terms of personality traits.

Despite their common usage, concepts of psychological disposition are still controversial. Some critics (ROHRACHER, 1963; KRISTOF, 1968) have called them "a resurrection of the 'psychology of faculties' (*Vermögenspsychologie)*", TRAXEL (1964) speaks of "fictitious conceptual realities", and HOLZKAMP (1964, 1965) has pointed out the danger of "an inadmissible duplication of reality" which could be avoided only by considering psychological traits strictly as theoretical constructs which neither denote anything directly observable, nor point towards some hidden reality. Ultimately, all these criticisms point in the same direction: The arbitrariness of the assumption of behavioral dispositions cannot be removed by simple "operational definitions" which are so common in the psychology of individual differences. However, as JÜTTEMANN (1972) has correctly remarked, to speak instead of "theoretical constructs" merely veils the fact that trait concepts are insufficiently determined in psychological scientific terminology. Therefore, such constructs do not contribute towards clarifying the question: under which conditions can the assumption of behavioral dispositions be justified on a metatheoretical basis?

Other authors, such as MERZ (1960), have tried to do without "theoretical constructs" and have championed the view that, e.g., "aggressiveness" means that one can order persons according to the degree to which they possess this trait. KEMPF (1973a) has further pursued this idea and has demanded that behavioral dispositions be considered strictly as metric concepts whose introduction presupposes that a scale of measurement already exists. ***Behavioral dispositions can be justified as an assumption only when they can be measured.*** Thereby it is understood that measurement is not simply the "art of assigning numbers to phenomena" but rather – according to the general theory of measurement – a mapping of the objects measured and of the empirically identifiable relations between them onto (real) numbers and relations between these numbers, respectively. Hence, as a minimal prerequisite for the assumption of behavioral dispositions, an empirical order relation must be found. In the case of "aggressiveness", for example, it would be defined as the two-valued predicate, "more aggressive than", where it is understood that one of two persons shall be considered more aggressive if and only if he or she tends more towards aggressive acts than the other person. What exactly is meant by such a definition will be rendered more precise below. However, one must also take into account the empirically well-founded notion that situations in which aggressive behavior may take place frequently differ in their incentive to aggression. Therefore, persons can be compared with respect to their aggressiveness only when this comparison is not distorted by situational factors. In line with the methodological tradition of psychology, this can be achieved by relating the comparison of two persons to a particular situation (S) and by thus introducing the predicate, "more aggressive than", only relative to this situation. The influence of situational factors is not eliminated by such a conditional definition; rather, a specific form of aggressiveness is postulated for each situation. Situational factors will be successfully eliminated only when one makes the additional assumption that aggressiveness is invariant within a defined universe of situations (Σ), so that statements about the relation between two persons do not depend on which situation S has been selected from Σ and has formed the context of the comparison. If there exists a situation in Σ in which person P_1 is more aggressive than Person P_2, then this relation must also hold for all other situations in Σ:

(1.1) $$(\exists S \epsilon \Sigma: P_1 \text{ more aggressive than } P_2) \rightarrow \\ \rightarrow (\forall S \epsilon \Sigma: P_1 \text{ more aggressive than } P_2).$$

With this assumption of the "Σ-invariance" of aggressiveness, new aspects come to light: Statements about the aggressiveness of persons no longer are just tautological reformulations of relations observed in a specific situation, but they assume a higher degree of universality and can be generalized to other situations, whereby the universe Σ determines the situations to which one may generalize. However, taken by itself, the assumption of Σ-invariance is not legitimate, since it has no consequences that could be falsified. No person can be in a situation S and at the same time in another situation S'. Therefore, the assumption of Σ-invariance requires the additional postulate of relative temporal invariance, so that the relations between two persons P_1 and P_2 are preserved over definite time intervals τ_1 and τ_2, i.e., they are valid at arbitrary times t_1 and t_2 within these time intervals:

(1.2) $$(\exists t_1 \epsilon \tau_1, t_2 \epsilon \tau_2, S \epsilon \Sigma: P_1 \text{ more aggressive than } P_2) \rightarrow$$
$$\rightarrow (\forall t_1 \epsilon \tau_1, t_2 \epsilon \tau_2, S \epsilon \Sigma: P_1 \text{ more aggressive than } P_2).$$

Thus, in addition to generalizability over all situations in Σ, there is generalizability within definite time intervals. This agrees with our intuitive understanding of "aggressiveness" (cf. SELG, 1968). To summarize: A methodologically satisfactory definition of behavioral dispositions such as "aggressiveness" is only possible when they are considered as relatively permanent traits that are manifested not only in particular contexts but in precisely defined classes of situations. Only when these prerequisites are met can statements about behavioral dispositions be generalized, and only then do they have any explanatory psychological value[1]. However, it still must be clarified what is meant by the statement that one person P_1 (at time t_1) in a given situation S tends more towards aggressive behavior than another person P_2 (at time t_2). As far as I know, only two relevant possibilities have been seriously discussed in psychology. The first consists in defining the predicate, "more aggressive than", by a fixed relation between the manifest forms of behavior that the persons show in situation S. KEMPF (1974a) has analyzed this possibility in detail and has shown that, in this case, the assumptions of Σ-invariance and relative temporal invariance must be presupposed, in order to arrive at a rank order at all. For, if one decides that, of two persons in situation S, the one should be considered "more aggressive" who

[1] At first sight, these prerequisites seem too strict, since statements about the aggressiveness of a person obviously can also be generalized when aggressiveness is subject to lawful change. However, even in this case we can speak of an invariant „initial aggressiveness" wich then undergoes a lawful process of change.

in fact acts aggressively, while the other person does not behave aggressively,

$$(1.3) \quad \left\{\begin{array}{l} P_1 \text{ acts aggressively and} \\ P_2 \text{ acts non-aggressively} \\ \text{in situation S} \end{array}\right\} \rightarrow \left\{\begin{array}{l} P_1 \text{ is more} \\ \text{aggressive than} \\ P_2 \text{ in situation S} \end{array}\right\},$$

then a single situation S yields at best a binary division into classes of "aggressive" persons that respond to the situation aggressively, and "non-aggressive" persons whose response to the situation is non-aggressive. On the other hand, when the assumptions of Σ-invariance and relative temporal invariance are added, the relation (1.3) defines a Guttman scale (GUTTMAN, 1950). In this scale, the persons P_v are brought into a rank order with regard to their aggressiveness, ξ_v, and at the same time the situations S_i are brought into a rank order with regard to their "inhibition of aggression", δ_i:

$$(1.4) \quad P_v \text{ acts in } S_i \begin{cases} \text{aggressively} & \text{if } \xi_v \geqslant \delta_i, \\ \text{non-aggressively} & \text{if } \xi_v < \delta_i. \end{cases}$$

The greater the "inhibition of aggression" in a situation can be before a person is prevented from reacting aggressively, the stronger is this person's tendency towards aggressive behavior. The greater the aggressiveness of a person must be in order to result in an aggressive response to a situation, the stronger is the "inhibition of aggression" that this situation exerts. Two persons P_1 and P_2, with $\xi_1 \geqslant \delta_i > \xi_2$, are brought into a rank order by the situation S_i. Any other situation S_j, with $\xi_1 \geqslant \delta_j > \xi_2$, leads to the same result. When two persons P_1 and P_2 are observed in the same situations $S_1, S_2, \ldots, S_i, \ldots, S_k$ selected from Σ and within the temporal intervals τ_1 and τ_2, that person is considered more aggressive who reacts more frequently in an aggressive way. If two persons show the same number of aggressive reactions, it cannot be decided who is more aggressive.

At first sight, this seems to be identical with a traditional operational definition of aggressiveness. This is not true, however, since operational definitions determine the predicate, "more aggressive than", by a particular test, i.e., by a particular selection of situations. In Guttman's quasi-scale, on the other hand, statements about the relations between two persons are essentially independent of which situations in Σ formed the basis of the comparison. All possible subsets of situations in Σ lead to the same result. At most, the precision of the statements will vary, e.g., when a "test" does not contain any situation with $\xi_1 \geqslant \delta_i > \xi_2$ and therefore does not allow any decision about which of two persons is more aggressive than the other. However, there is no subset of situations that could lead to a reversal of rank positions. If the rank order of the situations with respect to their "inhibition of aggression" is already known, the Guttman scale even allows the comparison of persons observed in different situations in Σ. For, according to the relations (1.2–1.4), an aggressive response to a certain situation S_i implies an aggressive response to all situations with less "inhibition of aggression" than S_i.

At the same time, however, there are unusually strict requirements with respect to the empirical data, and the Guttman scale is therefore generally considered to be an unrealistic model. In my opinion, this evaluation is justified not so much because of the deterministic structure of the Guttman scale but because it introduces deterministic assumptions into areas of psychology that are still far from being able to make predictions of manifest behavior with any degree of "certainty". We need not speculate here whether, for example, psychological research on aggression will ever succeed in making deterministic statements about the conditions of aggressive behavior. Certainly, such deterministic statements would presuppose an extremely high degree of sophistication, while the concept of "aggressiveness" is comparatively crude and weak in content. This is especially true when aggressiveness is merely based on a Guttman scale which

a) determines only the rank order of persons, and
b) whose statements can be generalized only within very narrow and specific classes of situations.

It is obvious, therefore, that the definition of the predicate, "more aggressive than" that has been adhered to so far should be replaced with a probabilistic definition. It is then stipulated that a person P_1 (at time t_1) in a certain situation S tends more towards aggressive acts than another person P_2 (at time t_2) if and only if the former responds aggressively to the situation with a higher probability than the latter:

$$(1.5) \quad \left\{\begin{array}{l} P_1 \text{ acts aggressively in S with} \\ \text{higher probability than } P_2 \end{array}\right\} \Leftrightarrow \left\{\begin{array}{l} P_1 \text{ is more aggressive in S} \\ \text{than } P_2 \end{array}\right\}$$

It follows from this definition – for a similar proposal, see KAUFMANN (1965) – that there exists, for each person P_v and each situation S_i in Σ, a strictly monotonic, positive relation between the aggressiveness of the person and the probability P_{vi} that the person will respond aggressively to the situation. The prerequisites of Σ-invariance and relative temporal invariance now assume the form of a probabilistic law:

$$(1.6) \qquad (\exists t_1 \epsilon \tau_1, t_2 \epsilon \tau_2, S_i \epsilon \Sigma: p_{1i} > p_{2i}) \rightarrow \\ \rightarrow (\forall t_1 \epsilon \tau_1, t_2 \epsilon \tau_2, S_i \epsilon \Sigma: p_{1i} > p_{2i}).$$

Statements of stochastic lawfulness cannot be verified or falsified in a strict sense but can only be decided upon according to a methodological convention. Since only probabilistic statements can be deduced from probabilistic models, the statistical falsification of a stochastic model can occur only with a certain error probability which, however, can be exactly stated. The procedure is well known: First, the distribution of some test quantity is deduced from the model (which contains the so-called null hypothesis, among others). Then, the deviation of the empirically found value of the test quantity from the "ideal" value predicted by the model is determined. Finally, the model is rejected if the probability of obtaining this (or a more extreme) deviation is smaller than a certain value agreed upon beforehand.

According to definition (1.5), the relation between two persons with respect to their aggressiveness can only be estimated with the help of statistical procedures, and this requires a sample of observations. The assumptions of Σ-invariance and relative temporal invariance, which were originally introduced because of other considerations, now prove to be necessary in order to be able to estimate the rank order of persons in terms of their aggressiveness. In order to obtain a more powerful scale of aggressiveness, multi-valued relations between persons must be introduced in addition to the relations (1.5) and (1.6), e.g., by postulating a certain form of functional connection between aggressiveness and the probabilities p_{vi}.

It can now be shown that these assumptions must satisfy certain metatheoretical and mathematical-statistical criteria and therefore cannot be made arbitrarily. For example, according to assumption (1.6) it is theoretically of no consequence on which situations in Σ the comparison of two persons rests, and it must therefore be postulated that the distribution of the estimation functions that determine the relation between the two persons be likewise independent of the sample of situations. If this prerequisite is not met, the sampling procedure represents an "interfering condition" (cf. HOLZKAMP, 1968), and it becomes possible to interpret any arbitrary statement about the relation between two persons against the empirical findings. As the Danish statistician Georg Rasch has shown (RASCH, 1965), this minimal requirement –

which he calls "specific objectivity"[2] – leads necessarily to a very special class of probabilistic measurement models, i.e., to very distinct assumptions about the form of the connection between aggressiveness and the probabilities p_{vi}. When specific objectivity is not met, the connection between a statement about the relation between two persons and observational data remains obscure. Holzkamp's view, that behavioral dispositions are purely constructs of imagination without any relation to reality, is then well justified. Therefore, it must be demanded that specific objectivity be understood not only as a prerequisite for the scientific validity of statements about the relations between persons, but also as a prerequisite for scientific statements about behavioral dispositions in general. Concepts such as aggressiveness are legitimate in the scientific language of psychology only when their measurement satisfies the principle of specific objectivity.

1.3 Outline of Rasch's Theory of Specifically Objective Measurement

So far, our discussion has led us to base the concept "aggressiveness" on a strictly increasing relation between the probability of aggressive behavior and the aggressiveness of persons. At the same time, we have kept in mind that analogous considerations also apply to other behavioral dispositions, such as the intelligence or the attitudes of persons. In the following, we will leave the example of aggressiveness and use the terminology of psychological test theory. Instead of situations, we will now speak of "items", and instead of aggressive and non-aggressive reactions, of "positive" and "negative" responses to items. We will replace the concept of aggressiveness by the more general concept of "latent trait" (ξ) which is defined by a potentially infinite universe of items ($i = 1, 2, \ldots$) and strictly increasing functions $f_i(\xi)$ such that $p(+ \mid v,i) = f_i(\xi_v)$ and $p(- \mid v, i) = 1 - f_i(\xi_v)$ for a particular person v. If we write $a_{vi} = 1$ for a positive response and $a_{vi} = 0$ for a negative response, then we can summarize these relations as follows:

$$(1.7) \qquad p\{a_{vi}\} = f_i(\xi_v)^{a_{vi}} \cdot (1 - f_i(\xi_v))^{1 - a_{vi}}.$$

[2] Originally, Rasch assigned an even more fundamental interpretation to the principle of specific objectivity. The present definition derives from KEMPF (1974a) and is in congruence with the statistical argumentation in RASCH (1965).

The functions $f_i(\xi)$ are called "item characteristic functions" in psychological test theory. For the time being, we will make no assumptions about their form but simply require that the items are locally stochastically independent of each other, i.e., that the probability that person v responds positively to item i does not depend on this person's responses to other items. In formal terms, this assumption – which is made throughout probabilistic test theory – means that the item intercorrelations observed in a sample of persons are solely due to the connection of the items with the latent dimension ξ, and therefore will disappear for constant ξ. At first sight, practice, fatigue, or position effects seem to contradict the principle of local stochastic independence. However, if one assumes that these are general effects that are equally pronounced in all persons, the contradiction can be eliminated by the assumption that the form of the item characteristic functions is determined not only by the relevant properties of the items, but also by their positions within the test (KEMPF, 1970). Nevertheless, the assumption of local stochastic independence brings about a considerable restriction of theoretical formulations, and there are a number of theoretical concepts in psychology, such as "catharsis" and "learning contingent on success", which are not compatible with the assumption of local stochastic independence. In the last three sections of this chapter, I will introduce some probabilistic measurement models which do not assume local stochastic independence and thereby permit the formalization of dynamic theories.

If no assumptions beyond monotonicity are made about the item characteristic functions, persons can be compared on the latent dimension ξ only by bringing them into a rank order with respect to the frequency with which they respond positively to the items of a given test. Let k be the number of items in the test; then we define

$$(1.8) \qquad a_{vo} = \sum_{i=1}^{k} a_{vi}$$

and obtain for the expected value of the test score a_{vo} the relation

$$(1.9) \qquad E(a_{vo}) = \sum_{i=1}^{k} E(a_{vi}) = \sum_{i=1}^{k} f_i(\xi_v),$$

which is likewise a strictly increasing function of the latent dimension ξ, since each of the functions $f_i(\xi)$ is strictly increasing. Thus we see that the prerequisites for the assumption of behavioral dispositions, introduced in the previous section, permit an unbiased comparison of persons. Nevertheless, a comparison of persons on the basis of the number of positive responses in a test may not be very useful, e.g., when the test score a_{vo} contains only a fraction of the information that is contained in the responses a_{vi}. When a comparison is based on total scores a_{vo}, one should therefore demand that a person's score contain the complete test information about the unknown person parameter ξ_v, i.e., that it be a sufficient statistic for the person parameter.

In mathematical terms, this postulate means that the response vector $(a_{vi}) = (a_{v1}, ..., a_{vk})$ of a person may not contain any information about the person parameter ξ_v which is not fully exhausted by the statistic a_{vo}, so that the conditional probability

$$(1.10) \qquad p\{(a_{vi}) \mid a_{vo}\} = \frac{p\{(a_{vi})\}}{p\{a_{vo}\}}$$

is no longer dependent on ξ_v. FISCHER (1974) has proven that, given local stochastic independence of the items, the relation (1.10) leads necessarily to RASCHS (1960) special logistic test model whose item characteristic functions (1.11) are all of the same type and differ from each other only by an "item easiness parameter" ϵ_i:

$$(1.11) \qquad f_i(\xi_v) = \frac{\xi_v \, \epsilon_i}{1 + \xi_v \, \epsilon_i} \qquad (i = 1, 2, ...).$$

In summary: If there is local stochastic independence, comparisons of persons on the basis of the number of positive responses can be performed without loss of information only when the model (1.11) holds. When the model does not hold, such comparisons still lead to an unbiased estimate of the rank order of the persons on the latent dimension ξ.

In the following, we will show how the principle of sufficient statistics can be exploited to perform specifically objective comparisons of persons or items. We start by assuming that n persons with person parameters $\xi_1, ..., \xi_n$ have responded independently to k items with item easiness parameters $\epsilon_1, ..., \epsilon_k$, so that the total response pattern can be represented as a response matrix $((a_{vi}))$ with n rows and k columns. The likelihood of this matrix is, according to (1.7) and (1.11),

$$(1.12) \qquad p\{((a_{vi}))\} = \prod_{v=1}^{n} \prod_{i=1}^{k} p\{a_{vi}\}$$

$$= \frac{\prod_{v=1}^{n} \prod_{i=1}^{k} (\xi_v \epsilon_i)^{a_{vi}}}{\prod_{v=1}^{n} \prod_{i=1}^{k} (1+\xi_v \epsilon_i)} .$$

We note that, in the numerator, the parameter ξ_v occurs k times, each time with the exponent a_{vi}, so that its total exponent is a_{vo}. The parameter ϵ_i occurs n times, also with the exponent a_{vi}, and therefore the total exponent is $a_{oi} = \sum_{v=1}^{n} a_{vi}$. Hence, we can write (1.12) in the simpler form,

$$(1.13) \qquad p\{((a_{vi}))\} = \frac{\prod_{v=1}^{n} \xi_v^{a_{vo}} \cdot \prod_{i=1}^{k} \epsilon_i^{a_{oi}}}{\prod_{v=1}^{n} \prod_{i=1}^{k} (1+\xi_v \epsilon_i)}$$

and we see immediately that the likelihood of the data matrix $((a_{vi}))$ depends only on the marginal vectors $(a_{vo}) = (a_{1o}, \ldots, a_{no})$ and $(a_{oi}) = (a_{o1}, \ldots, a_{ok})$. All possible data matrices with identical marginal vectors (a_{vo}) and (a_{oi}) are therefore equally probable, so that

$$(1.14) \qquad p\{(a_{vo}), (a_{oi})\} = \begin{bmatrix} (a_{vo}) \\ (a_{oi}) \end{bmatrix} \cdot p\{((a_{vi}))\},$$

where

(1.15) $\begin{bmatrix} (a_{vo}) \\ (a_{oi}) \end{bmatrix} =$ the number of possible data matrices for given marginal vectors (a_{vo}) and (a_{oi}).

The conditional probability of a specific response matrix $((a_{vi}))$, given the marginal vectors (a_{vo}) and (a_{oi}), thus no longer depends on the parameters of the model. Information that tells which person has responded positively to which item does not reveal anything additional about the parameters and therefore is irrelevant to their estimation. The marginal vectors (a_{vo}) and (a_{oi}) are sufficient statistics for the parameter vectors (ξ_v) and (ϵ_i). If only one of the two marginal vectors is held constant, we obtain conditional likelihood functions of the data matrix $((a_{vi}))$ which then depend only on those parameters whose sufficient statistics are not fixed. As RASCH (1966c) has shown,

$$(1.16) \qquad p\{((a_{vi})) \mid (a_{vo})\} = \frac{\prod_{i=1}^{k} \epsilon_i^{a_{oi}}}{\prod_{v=1}^{n} \gamma_{a_{vo}}(\epsilon_i)},$$

where $\gamma_{a_{vo}}(\epsilon_i)$ denotes the elementary symmetric function of order a_{vo} of the parameters $\epsilon_1, ..., \epsilon_k$:

$$(1.17) \qquad \begin{aligned} &\gamma_0(\epsilon_i) = 1 \\ &\gamma_1(\epsilon_i) = \epsilon_1 + \epsilon_2 + ... + \epsilon_k \\ &\gamma_2(\epsilon_i) = \epsilon_1\,\epsilon_2 + \epsilon_1\,\epsilon_3 + ... + \epsilon_{k-1}\,\epsilon_k \\ &\cdots\cdots\cdots\cdots \\ &\gamma_k(\epsilon_i) = \epsilon_1\,\epsilon_2 \ldots \epsilon_k, \end{aligned}$$

In analogous fashion,

$$(1.18) \qquad p\{((a_{vi})) \mid (a_{oi})\} = \frac{\prod_{v=1}^{n} \xi_v^{a_{vo}}}{\prod_{i=1}^{k} \gamma_{a_{oi}}(\xi_v)},$$

where $\gamma_{a_{oi}}(\xi_v)$ stands for the elementary symmetric function of order a_{oi} of the parameters $\xi_1, ..., \xi_n$. The separability of parameters – the formal equivalent of specific objectivity – is thus proven. According to (1.18), a comparison of persons with respect to the latent trait ξ can be based on a conditional probability distribution of the observed data $((a_{vi}))$ which depends only on the parameters to be compared.

As for the estimation of the model parameters, RASCH (1966b) has suggested that the method of maximum likelihood be applied to the conditional likelihood functions (1.16) and (1.18), and ANDERSEN (1973a) has justified this procedure in a comprehensive mathematical theory of "conditional inference". The resulting estimation functions for the parameters of the model (1.11) are special cases of the algorithms – discussed in Chapter 2 of the present book – for the multicategorial test model of RASCH (1961):

$$(1.19) \qquad p\{h \mid v, i\} = \frac{\xi_{vh}\epsilon_{ih}}{1 + \sum_{q=1}^{m-1} \xi_{vq}\epsilon_{iq}},$$

where $p\{h \mid v, i\}$ stands for the probability that person v responds to item i with category h (out of m possible response categories).

Computer programs for the numerical solution of the estimation equations can be found in FISCHER & ALLERUP (1968) and in FISCHER & SCHEIBLECHNER (1970) for the two-categorial test model (1.11), and in SCHEIBLECHNER (1971b), ANDERSEN (1972) and ALLERUP & SORBER (1974) for the multicategorial test model (1.19).

To demonstrate the significance of the method of conditional maximum likelihood (CML) for psychological statistics, FISCHER (1971) and ANDERSEN (1973b) refer to the distinction between structural and incidental parameters (NEYMAN & SCOTT, 1948). This distinction essentially rests on the relation between the number of parameters in a probabilistic model and the number of observations that depend on these parameters:

a) Structural parameters are of finite number, and for each of them a potentially infinite number of observations can be obtained which depend on this parameter.
b) Incidental parameters, on the other hand, are potentially of infinite number, and for each of them there is only a finite number of observations.

In psychological test theory, for example, the item parameters must be considered structural parameters. Each test contains only a limited number of items, and by increasing the sample of persons it is possible to obtain arbitrarily many observations that depend on the item parameters. Any addition to the sample of persons, on the other hand, implies the introduction of an additional person parameter, and each individual person always responds only to a relatively small number of items. Therefore, the person parameters must be considered incidental parameters. FISCHER (1971) has pointed out that the reverse situation is conceivable, so that individual parameters assume the property of structural parameters, e.g., when one makes the assumption, as in psychophysics, that one can perform arbitrarily many threshold measurements on a single subject. In Chapter 5 of this book we will discuss a model for the measurement of social norms in which parameters of equal content – viz., the normative achievements of subgroups of a society relative to the norms of the larger social system – are sometimes considered as structural parameters and sometimes as incidental parameters, depending on whether the norms of large groups or of small groups are under investigation.

In summary, almost every psychological theory contains incidental parameters, at least when it admits individual differences. As is evident from the work of NEYMAN & SCOTT (1948) and ANDERSEN (1973a, b), this results in fundamental statistical problems of estimation: Traditional methods of parameter estimation fail when the number of parameters to be estimated does not converge upon a fixed value as the number of observations increases towards infinity. In such a case, the maximum likelihood method does not even provide consistent estimation functions. However, if one applies the method of maximum likelihood to conditional likelihood functions, given (minimally) sufficient statistics for the incidental parameters, the structural parameters can be estimated without requiring simultaneous estimation of the incidental parameters. As we have seen in the model (1.11), these conditional likelihood functions are independent of the incidental parameters, and the CML method has generally the same properties as the direct maximum likelihood method without incidental parameters (ANDERSEN, 1970): It provides consistent estimation functions which are asymptotically normally distributed, unbiased, and have a known error variance. On the basis of CML estimates for the structural parameters, empirical tests of the structure of the model and of psychological hypotheses about the relations between the structural parameters of the model can be carried out in the form of conditional likelihood ratio tests (ANDERSEN, 1971, 1973c, and this volume).

1.4 A Dynamic Measurement Model with Separable Parameters

An essential generalization of probabilistic test theory consists in dropping the assumption of local stochastic independence. This extension of the model is essential for the measurement of aggressiveness, since the "catharsis hypothesis" says that the completion of an aggressive act reduces the probability of further aggressions. In more recent formulations of the catharsis hypothesis (cf. FESHBACH, 1956; BRAMEL, TAUB, & BLUM, 1968), it is presupposed that a person has first "been provoked to act aggressively; i.e., the behavior of another person has awakened aggressive tendencies in him that are primarily directed against the provocator" (DANN, 1971, p. 60; translated from German). If we adopt these more recent formulations of the catharsis hypothesis, it becomes evident that our considerations so far are applicable only to unprovoked aggressive tendencies. When provoked aggressiveness is to be assessed, the concept of catharsis contradicts the assumption of the local stochastic independence of situations, as well as the postulate of the relative temporal invariance of the relation "more aggressive than", as defined in (1.5). However, this contradiction is resolved when (1.5) is appropriately modified: If two persons are compared in the i-th situation after a provocation, it will be assumed that these two persons have acted identically or equivalently (in a well-defined sense) in all preceding situations.

The formal consequence of this dynamic extension of probabilistic test theory is that the probability of a certain response vector (a_{vi}) can no longer be expressed as

$$(1.20) \qquad p\{(a_{vi})\} = \prod_{i=1}^{k} p\{a_{vi}\}$$

but takes the form

$$(1.21) \qquad p\{(a_{vi})\} = \prod_{i=1}^{k} p\{a_{vi} \mid s_{vi}\},$$

where s_{vi} denotes the partial response vector $(a_{v1}, \ldots, a_{v(i-1)})$. The item characteristic functions $f_i(\xi)$ are replaced by conditional item characteristic functions

$$(1.22) \qquad f_{i \cdot s_{vi}}(\xi) = p\{a_{vi} = 1 \mid (a_{v1}, \ldots, a_{vi(i-1)}) = s_{vi}\},$$

and we will assume, in accordance with the modification of definition (1.5) stated above, that each of the functions $f_{i \cdot s_{vi}}(\xi)$ shall grow strictly monotonically with the latent dimension ξ. If the conditional item characteristic functions $f_{i \cdot s_{vi}}(\xi)$ are all equal for fixed i, local stochastic independence results as a special case of the more general model (1.21–1.22).

As a more specific formulation of the model (KEMPF, 1974b, c), we assume that the conditional item characteristic functions $f_{i \cdot s_{vi}}(\xi)$ depend on the responses to the preceding items only *via* the number of positive responses,

$$(1.23) \qquad r_{vi} = \begin{cases} 0 & \text{for } i = 1 \\ \sum_{j=1}^{i-1} a_{vj} & \text{for } i = 2, 3, \ldots, k \end{cases}$$

so that

$$(1.24) \qquad f_{i \cdot s_{vi}}(\xi) = f_{i \cdot r_{vi}}(\xi)$$

holds for all partial response vectors s_{vi} with identical total scores r_{vi}. Under the assumptions (1.23) and (1.24), all possible partial response vectors with the same total score r_{vi} are thus to be considered equivalent with respect to their influence on the probability of a positive response to the i-th item.

For the conditional item characteristic functions $f_{i \cdot r_{vi}}(\xi)$ we postulate the model structure

$$(1.25) \qquad f_{i \cdot r_{vi}}(\xi) = \frac{\xi_v + \psi_{r_{vi}}}{\xi_v + \sigma_i}$$

where $\psi_r < \sigma_i$ is presupposed, and σ_i represents the difficulty of the i-th item: the larger σ_i, the smaller the probability of a positive response to the i-th item (cf. Figure 1.1).

The influence that the positive responses to r_{vi} preceding items exert on the probability of a positive response to the i-th item is expressed in the ψ_r parameters which, henceforth, will be called *transfer parameters.* The larger the numeric value of the transfer parameter ψ_r, the higher is the (conditional) probability of a positive response (cf. Figure 1.2):

a) If the numerical values of the transfer parameters ψ_r are in a monotonically increasing relation to r, then transfer can be interpreted as "learning gain".

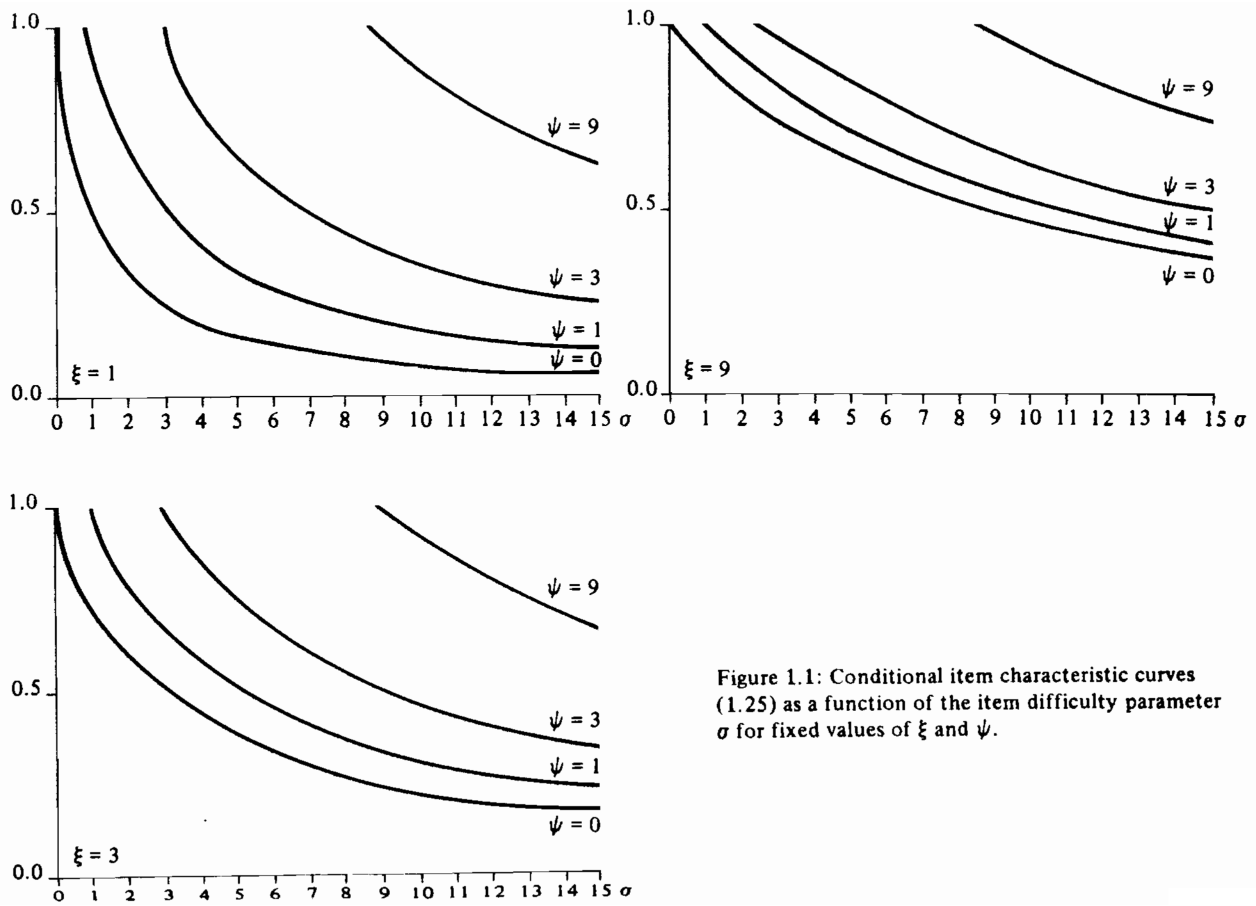

Figure 1.1: Conditional item characteristic curves (1.25) as a function of the item difficulty parameter σ for fixed values of ξ and ψ.

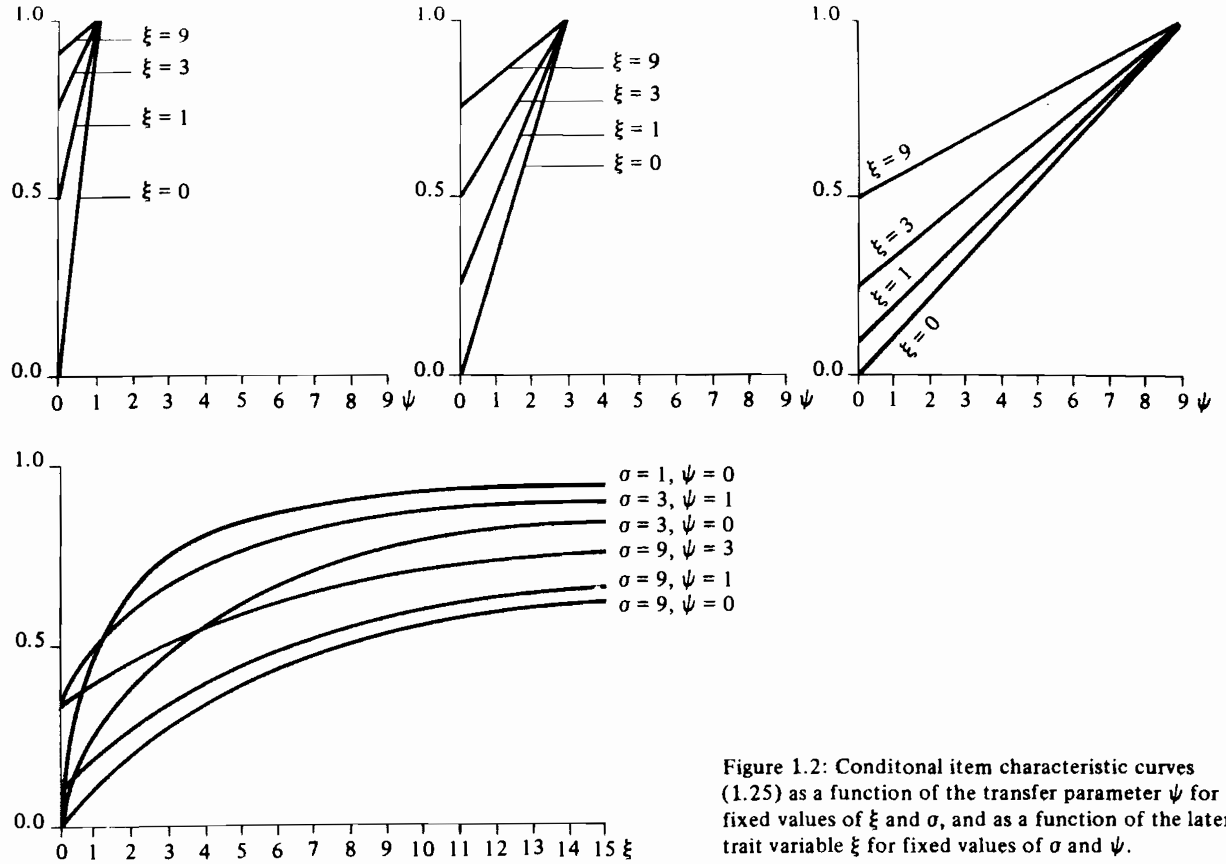

Figure 1.2: Conditonal item characteristic curves (1.25) as a function of the transfer parameter ψ for fixed values of ξ and σ, and as a function of the latent trait variable ξ for fixed values of σ and ψ.

b) If, on the other hand, the relation of the transfer parameters to r is monotonically decreasing, transfer is to be interpreted as "response inhibition". (In this sense, catharsis can be understood as inhibition of aggressive responses).

c) If the relation between ψ_r and r is not monotonic, we may speak of "fluctuation", which can be explained by simultaneous learning and inhibition processes that develop at different rates.

The parameters of the model are not fully determined by the model equation (1.25), i.e., the validity of (1.25) and the meaning of the parameters remain unchanged when we multiply the parameters with a positive constant and add to the person parameters another arbitrary constant which we simultaneously subtract from the transfer parameters ψ_r and from the item parameters σ_i. Thus, the parameters of the dynamic test model (1.25) are measured on interval scales, and it becomes necessary to standardize the scales. We may assume, without restriction of generality, that

$$\text{(1.26)} \qquad \text{Min}\,(\psi_r) = 0$$

and

$$\text{(1.27)} \qquad \prod_{i=1}^{k} \sigma_i = 1,$$

whereby we introduce conditions that correspond to those customary in the special logistic test model (1.11) of RASCH (1960). Rasch's model, with $\psi_o = \psi_1 = ... = \psi_{k-1} = 0$ and $\epsilon_i = \sigma_i^{-1}$, results as a special case of (1.25). We will show that the dynamic test model (1.25) essentially has the same properties as the test models of RASCH (1960, 1961). In (1.25), too,

a) the number of positive responses, a_{vo}, is a sufficient statistic for the parameter of person v;
b) comparisons of items and controls of the model can be performed with specific objectivity; and
c) CML estimation functions exist for the structural parameters.

Sufficient Statistics. To prove that the number of positive responses, a_{vo}, in the model (1.25) is a sufficient statistic for the person parameter ξ_v, we must show that the conditional probability (1.10) of the response vector (a_{vi}), given the total score a_{vo}, no longer depends on ξ_v. First, we write (1.25) in the equivalent form

$$(1.28) \qquad p\left\{a_{vi} \mid r_{vi}\right\} = \frac{(\xi_v + \psi_{r_{vi}})^{a_{vi}} (\sigma_i - \psi_{r_{vi}})^{1 - a_{vi}}}{\xi_v + \sigma_i}$$

and obtain by substitution into (1.21):

$$(1.29) \qquad p\left\{(a_{vi})\right\} = \frac{\prod_{i=1}^{k} (\xi_v + \psi_{r_{vi}})^{a_{vi}} (\sigma_i - \psi_{r_{vi}})^{1 - a_{vi}}}{\prod_{i=1}^{k} (\xi_v + \sigma_i)}.$$

$$= \frac{\prod_{r=0}^{a_{vo}-1} (\xi_v + \psi_r) \prod_{i=1}^{k} (\sigma_i - \psi_{r_{vi}})^{1 - a_{vi}}}{\prod_{i=1}^{k} (\xi_v + \sigma_i)}$$

The probability of a certain total score a_{vo} then follows from (1.29) by summing the probabilities $p\{(a^*_{vi})\}$ over all possible response vectors (a^*_{vi}) that have the same total score $\sum_{i=1}^{k} a^*_{vi} = a_{vo}$:

$$(1.30) \qquad p\left\{a_{vo}\right\} = \sum_{(a^*_{vi}) \mid a_{vo}} p\left\{(a^*_{vi})\right\}$$

$$= \frac{\prod_{r=0}^{a_{vo}-1} (\xi_v + \psi_r) \sum_{(a^*_{vi}) \mid a_{vo}} \prod_{i=1}^{k} (\sigma_i - \psi_{r^*_{vi}})^{1 - a^*_{vi}}}{\prod_{i=1}^{k} (\xi_v + \sigma_i)}$$

where $r^*_{vi} = 0$ for i = 1, and $r^*_{vi} = \sum_{j=1}^{i-1} a^*_{vi}$ for i = 1, 2, ..., k.

For the conditional probability $p\{(a_{vi}) \mid a_{vo}\}$, the following expression results from (1.29) and (1.30):

$$(1.31) \qquad p\left\{(a_{vi}) \mid a_{vo}\right\} = \frac{p\left\{(a_{vi})\right\}}{p\left\{a_{vo}\right\}}$$

$$= \frac{\prod_{i=1}^{k} (\sigma_i - \psi_{r_{vi}})^{1 - a_{vi}}}{\sum_{(a^*_{vi}) \mid a_{vo}} \prod_{i=1}^{k} (\sigma_i - \psi_{r^*_{vi}})^{1 - a^*_{vi}}}$$

which no longer contains the person parameter ξ_v. This completes the proof.

Specific Objectivity and CML Estimation. In the following, we generalize to n persons who respond to k items independently of each other . We obtain from (1.29)

$$(1.32)\qquad p\{((a_{vi}))\} = \prod_{v=1}^{n} p\{(a_{vi})\}$$

$$= \frac{\left\{\prod_{v=1}^{n}\prod_{r=0}^{a_{vo}-1}(\xi_v+\psi_r)\right\}\cdot\left\{\prod_{i=1}^{k}\prod_{r=0}^{i-1}(\sigma_i-\psi_r)^{n_{ri}}\right\}}{\prod_{v=1}^{n}\prod_{i=1}^{k}(\xi_v+\sigma_i)}$$

where n_{ri} stands for the number of persons with $r_{vi} = r$ and $a_{vi} = 0$. As can be seen, the likelihood (1.32) of the data matrix $((a_{vi}))$ depends only on the frequencies a_{vo} ($v = 1, \ldots, n$) and n_{ri} ($i = 1, \ldots, k$; $r = 0, \ldots, i - 1$). Therefore, all possible data matrices with identical marginal person vectors (a_{vo}) and identical marginal item matrices $((n_{ri}))$ are equally probable, so that

$$(1.33)\qquad p\{(a_{vo}), ((n_{ri}))\} = \begin{bmatrix}(a_{vo})\\((n_{ri}))\end{bmatrix}\cdot p\{((a_{vi}))\},$$

where

$$(1.34)\qquad \begin{bmatrix}(a_{vo})\\((n_{ri}))\end{bmatrix} = \text{the number of possible data matrices with a given marginal person vector } (a_{vo}) \text{ and a given marginal item matrix } ((n_{ri})).$$

The conditional probability of a given response matrix for given (a_{vo}) and $((n_{ri}))$,

$$(1.35)\qquad p\{((a_{vi}))\mid(a_{vo}), ((n_{ri}))\} = \frac{p\{((a_{vi}))\}}{p\{(a_{vo}), ((n_{ri}))\}}$$

$$= \frac{1}{\begin{bmatrix}(a_{vo})\\((n_{ri}))\end{bmatrix}}$$

is therefore independent of the model parameters. Information concerning which person has responded positively to which item does not reveal anything new about the parameters and therefore is irrelevant for the parameter estimation. The marginal person vector (a_{vo}) and the marginal item matrix $((n_{ri}))$ fully exhaust the information about the parameters in the model.

Finally, we derive the basis for CML estimation of the structural item and transfer parameters, by substituting (1.31) in

$$(1.36) \qquad p\left\{((a_{vi})) \mid (a_{vo})\right\} = \prod_{v=1}^{n} p\left\{(a_{vi}) \mid a_{vo}\right\}$$

from which we obtain the conditional likelihood

$$(1.37) \qquad p\left\{((a_{vi})) \mid (a_{vo})\right\} = \frac{\prod_{i=1}^{k} \prod_{r=0}^{i-1} (\sigma_i - \psi_r)^{n_{ri}}}{\prod_{v=1}^{n} \sum_{(a^*_{vi}) \mid a_{vo}} \prod_{i=1}^{k} (\sigma_i - \psi_{r_*})^{1 - a^*_{vi}}}.$$

Taking logarithms of (1.37) and setting the first order partial derivatives with respect to the item and transfer parameters equal to 0 finally yields the necessary estimation equations.

Questions concerning the numerical solution of the estimation equations are treated in KEMPF & HAMPAPA (1975, and this volume). A computer program for the iterative computation of the estimation functions is available in KEMPF & MACH (1975).

1.5 Dynamic Processes as Changes in Behavioral Dispositions: The "Catharsis Hypothesis"

In reviewing the results of the last section, it becomes evident that the dynamic measurement model (1.25) has indeed the same desirable properties as the test models of Rasch, but nevertheless with certain restrictions:

a) Unlike the latter, where the parameters can assume arbitrary positive real values, the relations between the parameters are restricted in the dynamic model, since the item and transfer parameters must satisfy the condition

$$(1.38) \quad \psi_r < \sigma_i \text{ for all i and all r.}$$

b) Also, in the test models of Rasch and their multifactorial generalizations (MICKO, 1970; SCHEIBLECHNER, 1971a; KEMPF, 1972a), the estimation of each class of parameters is independent of the estimation of other classes of parameters, but this is not possible in the dynamic model. Item parameters can be determined independently of person parameters but not of the transfer parameters.

From a psychological viewpoint, such dependencies between the parameters and between their estimation functions can be justified only if they have an unequivocal psychological interpretation, and if they can be shown not to contradict the basic principle of specific objectivity.

For this purpose we define

$$(1.39) \qquad \xi_{vr_{vi}} = \xi_v + \psi_{r_{vi}}$$

as the degree of pronouncedness of the latent dimension for person v after r_{vi} positive reactions, and

$$(1.40) \qquad \sigma_{ir_{vi}} = \sigma_i - \psi_{r_{vi}}$$

as the difficulty of item i after r_{vi} positive reactions, and we write the model (1.25) in the equivalent form,

$$(1.41) \qquad f_{i \cdot r_{vi}}(\xi) = \frac{\xi_{vr_{vi}}}{\xi_{vr_{vi}} + \sigma_{ir_{vi}}}.$$

On the basis of these relations, it is possible to identify the dynamic transfer component of the model (1.25) as the change in the latent dimension ξ, with a simultaneous converse change in item difficulty. Our previous observation that item parameters cannot be determined independently of the transfer parameters thus does not contradict the principle of specific objectivity. For, if a variable changes continuously while it is being measured, it cannot be postulated that its measurement should be independent of the simultaneous measurement of its change. At the same time, an unequivocal interpretation of the relation (1.38) follows from definition (1.40): The maximal possible positive transfer from preceding items to the i-th item must remain smaller than the "initial difficulty" σ_i; otherwise, the relation

$$(1.42) \qquad 1 > f_{v \cdot r_{vi}}(\xi)$$

no longer holds, and the model either assumes deterministic character (if $\psi_{r_{vi}} = \sigma_i$) or no longer obeys the axioms of probability theory (if $\psi_{r_{ri}} > \sigma_i$). In psychological terms, the restriction (1.38) thus means that a probabilistic analysis of the responses to the i-th item is possible only when the item retains a certain "residual difficulty" $\sigma_{ir_{vi}}$, regardless of how many other items have previously been responded to positively.

This relation is interesting mainly because it indicates the possibility of a transition from probabilistic to deterministic statements.

Since the dynamic test model (1.25) describes the transfer from previous responses to later responses as a simultaneous change in the latent trait and in the item difficulties, its application presupposes the definition of an initial psychological state of the dynamic process, as it is provided, for example, in the more recent formulations of the catharsis hypothesis mentioned above. There it is assumed that the completion of an aggressive act leads to a reduction in the probability of further aggressions only when it has been preceded by a provocation. If the initial psychological state of the dynamic process is not defined, as in the "original" catharsis hypotheses (cf. DANN, 1971), then the model (1.25) can be applied only if either

a) the complete "history" of the persons is known, i.e., if it is known for each person how many times he or she has acted aggressively prior to the experiment, or
b) if it is assumed that the transfer effect converges with increasing r against a fixed value, and that each person has acted aggressively sufficiently often prior to the experiment, so that the transfer from pre-experimental behavior to the behavior in the experiment already approximates this fixed value.

While the first of these two possibilites cannot be realized in actual psychological research, the second suggests that no transfer may be found during the period of observation, since according to the assumptions, the ψ_r parameters approximate their asymptotic value. The dynamic model (1.25) is therefore inadequate in principle for testing the "original" catharsis hypothesis. If such a test is desired, one would have to construct a model in which transfer exclusively affects the pronouncedness of the latent dimension, i.e., where catharsis can be explained solely by a change in aggressiveness, so that the aggressiveness of a person can be expressed as a function of his or her previous aggressive acts. It can be shown, however, that such a model is fundamentally in contradiction to the principles of specific objectivity and CML estimation. To prove this, we refer to the previously mentioned theorems of RASCH (1965) and ANDERSEN (1973b), according to which the demand for specific objectivity and the CML method, together with local stochastic independence, necessarily leads to Rasch's logistic test model (1.11). Every dynamic test model that is expected to have comparable properties must therefore be some generalization of Rasch's model. Moreover, if transfer shall be described only as a change in the latent traits of the persons, the following model necessarily results:

To summarize, it can be said that the original catharsis hypothesis fundamentally contradicts the concept of aggressiveness. Moreover, the original catharsis hypothesis necessarily leads to the assumption of individual differences (cf. HILKE et al., 1975) and is therefore not amenable to a methodologically satisfactory test. Nevertheless, legitimate formulations of the catharsis hypothesis can be found in which catharsis corresponds to a reduction in aggressiveness. This is shown in the following considerations, where we will assume that the suppression of provoked aggressive tendencies leads to an accumulation of aggression which is released only when the provocation is finally responded to aggressively. Catharsis, then, is the "equilibration" of the accumulated aggression, i.e., the falling back of aggressiveness to its initial value.

If one attempts to formalize this more precise version of the catharsis hypothesis, each provocation i and each person v can be associated with a sequence of reactions which continues until the provocation has been responded to aggressively. The probability that the sequence of reactions is not yet terminated at time j after the provocation can be written in the abbreviated form,

(1.47) $$q_{vij} = p\{z_{vi} > j\} = \frac{\lambda_{vij}}{1+\lambda_{vij}} \quad \text{for } j = 1, 2, \ldots$$

where z_{vi} stands for that point in time after the i-th provocation at which person v responds aggressively to the provocation, and λ_{vij} is defined by

(1.48) $$\lambda_{vij} = \frac{q_{vij}}{1 - q_{vij}} \quad \text{for } j = 1, 2, \ldots .$$

From the definition,

(1.49) $$\omega_{vij} = \frac{\lambda_{vij+1}}{\lambda_{vij}} \quad \text{for } j = 1, 2, \ldots ,$$

it follows that, for arbitrary successive points in time, j and j + 1, this relation[3] holds:

[3] Note the formal analogy to Luce's Beta Model (LUCE, 1959).

$$(1.50)\qquad p\{z_{vi}>j+1 \mid z_{vi}>j\} = \frac{q_{vi(j+1)}}{q_{vij}} = \frac{\dfrac{\omega_{vij}\lambda_{vij}}{1+\omega_{vij}\lambda_{vij}}}{q_{vij}} = \frac{\dfrac{\omega_{vij}q_{vij}}{1-q_{vij}}}{q_{vij}\left(1+\dfrac{\omega_{vij}q_{vij}}{1-q_{vij}}\right)} = \frac{\omega_{vij}}{1+(\omega_{vij}-1)\cdot q_{vij}},$$

which is a strictly monotonically increasing function of the parameter ω_{vij} (cf. Figure 1.3).

The conditional probability that person v reacts aggressively at time j + 1, after previously suppressing his aggressiveness,

$$(1.51)\qquad p\{z_{vi}=j+1 \mid z_{vi}>j\} = 1 - p\{z_{vi}>j+1 \mid z_{vi}>j\},$$

is therefore strictly monotonically decreasing with ω_{vij} and strictly monotonically increasing with $\delta_{vij} = \omega_{vij}^{-1}$ (cf. Figure 1.4). Therefore, the parameter δ_{vij} shall be called the "aggression increment" which accrues to person v from the suppression of his or her aggressiveness at time j after the i-th provocation.

To show that this interpetation is justified, we must prove that δ_{vij} can indeed be represented as a change of aggressiveness. For this purpose, we refer to the assumption made earlier, that the aggressive response to a provocation should release accumulated aggression, so that the aggressiveness of the person assumes its initial value after the aggressive response. Hence, completed reaction sequences of one person can be regarded as independent of each other. On the basis of the exposition in sections 1.2 and 1.3, the assumption of an "initial value" of aggressiveness can be justified only if the probability that the person responds to a provocation immediately (i.e., at time j = 1) can be expressed in the form

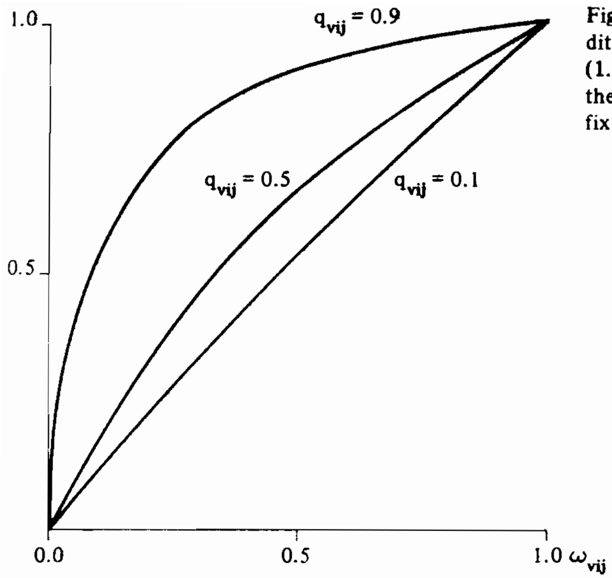

Figure 1.3: The conditional probability (1.50) as a function of the parameter ψ_{vij} for fixed values of q_{vij}.

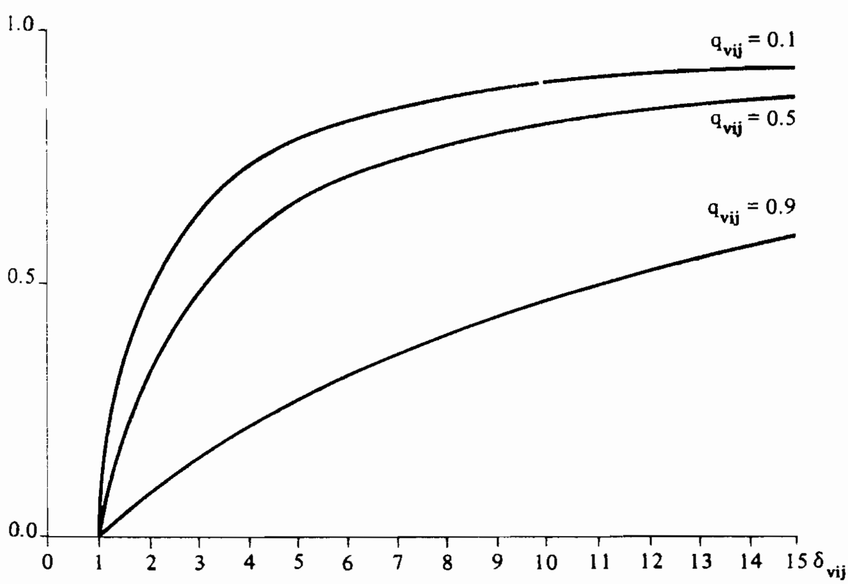

Figure 1.4: The conditional probability (1.51) as a function of the parameter $\delta_{vij} = \omega_{vij}^{-1}$ for fixed values of q_{vij}.

$$(1.52) \qquad p\{+ \mid v, i\} = \frac{\xi_v \epsilon_i}{1 + \xi_v \epsilon_i} \quad \begin{array}{l} \text{for } i = 1, \ldots, k \\ \text{and } v = 1, \ldots, n \end{array}$$

Here, ξ_v signifies the "initial aggressiveness" of the person, and ϵ_i the "incentive for aggression" of the situation in which the person finds himself or herself immediately after the provocation. In general, ϵ_i will also depend on the kind and degree of the provocation.

By substituting (1.52) into (1.47–1.50) we obtain the relations

$$(1.53) \qquad q_{vij} = \begin{cases} 1 - p\{+ \mid v, i\} = \dfrac{\sigma_i}{\xi_v + \sigma_i} & \text{if } j = 1 \\ (1 - p\{+ \mid v, i\}) \displaystyle\prod_{t=1}^{j-1} \frac{\omega_{vit}}{1 + (\omega_{vit} - 1) \cdot q_{vit}} & \text{if } j > 1 \end{cases}$$

where $\sigma_i = \epsilon_i^{-1}$

and

$$(1.54) \qquad q_{vij} = \frac{(\xi_v \cdot \prod_{t<j} \delta_{vit})^{-1} \cdot \sigma_i}{1 + (\xi_v \cdot \prod_{t<j} \delta_{vit})^{-1} \cdot \sigma_i} \quad \text{for } j = 1, 2, \ldots .$$

According to these expressions, δ_{vit} can indeed be represented as a change in aggressiveness. It should be noted that only the "initial aggressiveness" still satisfies the definition (1.5), while the relation "more aggressive than" is defined relative to the j-th point in time after a provocation:

$$(1.55) \qquad \left\{ \begin{array}{l} P_1 \text{ is more aggressive} \\ \text{than } P_2 \text{ at time j} \\ \text{after provocation i} \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} P_1 \text{ has responded aggressively to} \\ \text{provocation i at time j with higher} \\ \text{probability than } P_2 \end{array} \right\}$$

Thus, the definition of aggressiveness proposed here also differs fundamentally from the modification of definition (1.5) expressed in (1.22). In the model (1.54), the conditional probability that a person acts aggressively at time j after a provocation, after previously suppressing all aggressiveness, is not only a strictly monotonically increasing function of his or her initial aggressiveness, but also a strictly monotonically increasing function of the aggressive energy accumulated since the provocation, $\prod_{t<j} \delta_{vit}$. We see from this that the question of an adequate definition of the concept "aggressiveness" depends essentially on the kinds of regularities that underlie catharsis.

In order to allow a satisfactory estimation of the parameters in (1.54), certain restrictions must be introduced, i.e., it must be assumed that the "aggression increment", δ_{vij}, can be split into two multiplicative factors, one of which (α_{vj}) is independent of i, while the other (β_{ij}) is independent of v:

$$\text{(1.56)} \qquad \delta_{vij} = \alpha_{vj} \cdot \beta_{ij} \qquad \begin{aligned} &\text{for } i = 1, \ldots, k \\ &v = 1, \ldots, n \\ &j = 1, 2, \ldots . \end{aligned}$$

With the definitions

$$\text{(1.57)} \qquad \theta_{vj} = (\xi_v \cdot \prod_{t<j} \alpha_{vj})^{-1}$$

and

$$\text{(1.58)} \qquad \eta_{ij} = \sigma_i \cdot \prod_{t<j} \beta_{ij}$$

the following relation is obtained from (1.55):

$$\text{(1.59)} \qquad q_{vij} = \frac{\theta_{vj}\eta_{ij}}{1+\theta_{vj}\eta_{ij}} \qquad \begin{aligned} &\text{for } i = 1, \ldots, k \\ &v = 1, \ldots, n \\ &j = 1, 2, \ldots . \end{aligned}$$

By using the notation,

$$\text{(1.60)} \qquad x_{vi}^{(j)} = \begin{cases} 1 \text{ if } z_{vi} > j \\ 0 \text{ if } z_{vi} \leqslant j, \end{cases}$$

the reaction sequences of n persons to k provocations can be represented as a sequence of n x k matrices,

$$((x_{vi}^{(1)})), ((x_{vi}^{(2)})), \ldots, ((x_{vi}^{(j)})), \ldots$$

where all observations $x_{vi}^{(j)}$ within each data matrix $((x_{vi}^{(j)}))$ are locally independent of each other, while each row $(x_{vi}^{(j)})$ of a given data matrix is locally dependent on the corresponding rows of the remaining data matrices. Thus the model (1.54–1.56) allows the separate analysis of the data matrices $((x_{vi}^{(j)}))$ by means of conditional inference, and the subsequent factorization of the parameter estimates $\hat{\theta}_{vj}$ and $\hat{\eta}_{ij}$ into their components by the method of least squares. For, according to (1.59), each of the data matrices $((x_{vi}^{(j)}))$ can be described by a Rasch model of the form (1.11). It should be noted, however, that this kind of conditional inference is somewhat different from the one which was applied to the dynamic test model (1.25).

In the present discussion of dynamic measurement models for the investigation of the catharsis hypothesis we have described only two out of a greater number of possible models . We have shown that there is a direct correspondence between the psychological assumptions about the transfer process and the type of conditional inference required.

1.6 Necessary and Sufficient Conditions for a General Dynamic Test Model

As has been pointed out earlier (KEMPF, 1974d, 1976; KEMPF and HAMPAPA, 1975, and this volume), the model (1.25) is only a special case of a more general dynamic test model. We will now derive necessary and sufficient conditions for the existence of sufficient statistics in the general model.

Let I be a finite or countable infinite set of test items. Let the reaction of subject v to the i-th item be $a_{vi} = 1$ in the case of a positive answer and $a_{vi} = 0$ in the case of a negative answer. For any selection of $k \geqslant 2$ items from I and for any sequence of them $(i = 1, 2, \ldots, k)$,

$$(1.61) \qquad p\{(a_{vi})\} = \prod_{i=1}^{k} p\{a_{vi} \mid a_{v1}, \ldots, a_{v(i-1)}\}$$

i. e., local serial dependence is assumed. The probability function of the reaction variable $A_{vi} (i = 1, k)$ may be represented, without loss of generality, as

$$(1.62) \qquad p\{a_{vi} \mid (a_{v1}, \ldots, a_{v(i-1)}) = s_{vi}\} = \begin{cases} \dfrac{\xi_v(s_{vi})}{\xi_v(s_{vi}) + \sigma_i(s_{vi})} & \text{if } a_{vi} = 1 \\[2ex] \dfrac{\sigma_i(s_{vi})}{\xi_v(s_{vi}) + \sigma_i(s_{vi})} & \text{if } a_{vi} = 0 \end{cases}$$

where it is presupposed that $\xi_v(s_{vi}) > 0$ and $\sigma_i(s_{vi}) > 0$.

The set of all possible partial response vectors, $s_{vi} = (-)^4, (0), (1), (0,0), (0,1), (1,0), (1,1), \ldots, (1,1, \ldots, 1)$ may be divided into m disjoint classes, $S_1, \ldots, S_r, \ldots, S_m$. S_r is the class of all partial response vectors that show the property r (e.g., whose last element equals 1 – the Markov assumption –, or that contain a certain number of unity elements, etc.). Let h_{vr} be the number of items – associated with a partial vector $s_{vi} \epsilon S_r$ – that have been answered positively by subject v:

$$h_{vr} = \sum_{i:s_{vi}\epsilon S_r} a_{vi}.$$

The probability (1.62) shall depend on s_{vi} only *via* the class that s_{vi} belongs to, so that

$$(1.63) \qquad p\{a_{vi} \mid s_{vi}\} = p\{a_{vi} \mid r_{vi}\},$$

where r_{vi} is that r for which $s_{vi} \epsilon S_r$.

Theorem: Given the above assumptions, the model assumptions

$$(1.64a) \qquad \xi_v(s_{vi}) = \xi_v + \psi_{r_{vi}}$$

and

$$(1.64b) \qquad \sigma_i(s_{vi}) = \sigma_i - \psi_{r_{vi}}$$

are necessary and sufficient for the score vector $(h_{v1}, \ldots, h_{vm})$ *to be a sufficient statistic for the person parameters* $\xi_v(s_{vi})$.

Proof: First we will show that $(h_{v1}, \ldots, h_{vm})$ is a sufficient statistic for the person parameters if the restrictions (1.64) are met. Substitution of (1.64) into (1.62) and of (1.62) into (1.61) results in

$$(1.65) \qquad p\{(a_{vi})\} = \prod_{i=1}^{k} ((\xi_v + \psi_{r_{vi}})^{a_{vi}} + (\sigma_i - \psi_{r_{vi}})^{1-a_{vi}})/(\xi_v + \sigma_i)$$

$$= \prod_{r=1}^{m} (\xi_v + \psi_r)^{h_{vr}} \prod_{i=1}^{k} (\sigma_i - \psi_{r_{vi}})^{1-a_{vi}}/(\xi_v + \sigma_i).$$

[4] (–) denotes the empty response vector, where no prior items have been responded to.

Consequently,

$$(1.66) \qquad p\{(h_{v\tau})\} = \sum_{(a^*_{vi}) \mid (h_{v\tau})} p\{(a^*_{vi})\}$$

$$= \prod_{r=1}^{m} (\xi_v + \psi_r)^{h_{v\tau}} \sum_{(a^*_{vi}) \mid (h_{v\tau})} \prod_{i=1}^{k} (\sigma_i - \psi_{r^*_{vi}})^{1-a^*_{vi}} / (\xi_v + \sigma_i).$$

The summation on the right side of equation (1.66) extends over all possible response vectors (a^*_{vi}) that are compatible with the score vector $(h_{v\tau})$. It is assumed that the classes $S_1, ..., S_m$ have been defined so that there is, in general, more than one response vector that is compatible with $(h_{v\tau})$. Otherwise, (a_{vi}) can be directly derived from $(h_{v\tau})$, and the conditional likelihood of the response vector, given $(h_{v\tau})$, degenerates.

Finally, we obtain the conditional likelihood:

$$(1.67) \qquad p\{(a_{vi}) \mid (h_{v\tau})\} = p\{(a_{vi})\}/p\{(h_{v\tau})\}$$

$$= \prod_{i=1}^{k} (\sigma_i - \psi_{r_{vi}})^{1-a_{vi}} / \sum_{(a^*_{vi}) \mid (h_{v\tau})} \prod_{i=1}^{k} (\sigma_i - \psi_{r^*_{vi}})^{1-a^*_{vi}},$$

which no longer contains person parameters. This completes the first part of the proof: *The model assumptions (1.64) are sufficient for the existence of sufficient statistics of the form* $(h_{v\tau})$.

Next we will demonstrate that the model assumptions (1.64) are *necessary* for the existence of sufficient statistics $(h_{v\tau})$ when $k = 2$. In this case, there are four possible response vectors: (0,0), (0,1), (1,0), and (1,1). The corresponding partial response vectors s_{vi} for the items $i = 1$ and $i = 2$ are listed in Table 1.1.

There are four possibilities of grouping the response vectors into classes. They are illustrated in Table 1.2 together with the resulting score vectors.

It is immediately evident from Table 1.2 that Cases III and IV are degenerate and do not warrant further consideration. Case I is the special case of local stochastic independence of the two items. Here it follows from a theorem of RASCH (1971; cf. also FISCHER, 1974) that necessarily $\xi_v(s_{vi}) = \xi_v$ and $\sigma_i(s_{vi}) = \sigma_i$ for all s_{vi}, so that the restrictions postulated in (1.64) are trivially met.

Table 1.1. Partial response vectors for k = 2

(a_{vi})	s_{v1}	s_{v2}
(0, 0)	(–)	(0)
(0, 1)	(–)	(0)
(1, 0)	(–)	(1)
(1, 1)	(–)	(1)

Table 1.2. Possible classifications of s_{vi}, and the resulting score vectors

Classification	(a_{vi})	(h_{vr})	
$S_1 = \{(-), (0), (1)\}$	(0, 0)	(0)	I
	(0, 1)	(1)	
	(1, 0)	(1)	
	(1, 1)	(2)	
$S_1 = \{(-), (0)\}$	(0, 0)	(0, 0)	II
$S_2 = \{(1)\}$	(0, 1)	(1, 0)	
	(1, 0)	(1, 0)	
	(1, 1)	(1, 1)	
$S_1 = \{(-), (1)\}$	(0, 0)	(0, 0)	III
$S_2 = \{(0)\}$	(0, 1)	(0, 1)	
	(1, 0)	(1, 0)	
	(1, 1)	(2, 0)	
$S_1 = \{(-)\}$	(0, 0)	(0, 0, 0)	IV
$S_2 = \{(0)\}$	(0, 1)	(0, 1, 0)	
$S_3 = \{(1)\}$	(1, 0)	(1, 0, 0)	
	(1, 1)	(1, 0, 1)	

In Case II, the response vectors (0,1) and (1,0) have the same score vector $(h_{vr}) = (1,0)$. If (h_{vr}) shall be a sufficient statistic for the person parameters, the expression $p\{(0,1)\}/(p\{(0,1)\} + p\{(1,0)\})$ must no longer depend on the person parameters. Hence, the expression

$$(1.68)\qquad \frac{p\{(1,0)\}}{p\{(0,1)\}} = \frac{\xi_v(-)\sigma_2(1)/((\xi_v(-)+\sigma_1(-))(\xi_v(1)+\sigma_2(1)))}{\sigma_1(-)\xi_v(0)/((\xi_v(-)+\sigma_1(-))(\xi_v(0)+\sigma_2(0)))}$$

$$= \frac{\xi_v(-)}{\xi_v(0)}\,\frac{\sigma_2(1)}{\sigma_1(-)}\,\frac{\xi_v(0)+\sigma_2(0)}{\xi_v(1)+\sigma_2(1)}$$

must be independent of the person parameters, so that

$$(1.69)\qquad \frac{\xi_v(-)}{\xi_v(0)}\,\frac{\xi_v(0)+\sigma_2(0)}{\xi_v(1)+\sigma_2(1)} = c$$

must be a constant. Without loss of generality, we may set $c = 1$, so that

$$(1.70)\qquad \xi_v(-)(\xi_v(0)+\sigma_2(0)) = \xi_v(0)(\xi_v(1)+\sigma_2(1)).$$

(If $c \neq 1$, the parameters $\xi_v(1)$ and $\sigma_2(1)$ may be transformed by multiplication by the constant c^{-1}, so that equation (1.70) holds. The item solution probabilites remain invariant under this scale transformation.)

Since we are considering a fixed person v, the generality of the argument is not restricted if we define:

(1.71a) $\xi_v(0) = \xi_v + \psi_0,$

(1.71b) $\sigma_2(0) = \sigma_2 - \psi_0,$

(1.71c) $\xi_v(1) = \xi_v + \psi_1,$

(1.71d) $\sigma_2(1) = \sigma_2 - \psi_1,$

where $\psi_0 < \sigma_2$ and $\psi_1 < \sigma_2$ is presupposed. Consequently, $\xi_v(-) = \xi_v + \psi_0$, and since the existence of sufficient statistics of the form $(h_{v\tau})$ shall be independent of the particular selection and sequence of the items, it follows from $p\{a_{vi} \mid (-)\} = p\{a_{vi} \mid (0)\}$ that $\sigma_2(-) = \sigma_2(0) = \sigma_2 - \psi_0$ and $\sigma_1(-) = \sigma_1(0)$. From the definition $\sigma_1 = \sigma_1(0) - \psi_0$ it finally follows that $\sigma_1(1) = \sigma_1 - \psi_1$, so that all restrictions postulated in (1.64) are fulfilled. We may then summarize:

If two arbitrary items are selected from I and presented to a subject in arbitrary order, the model assumptions (1.64) are necessary and sufficient for the existence of sufficient statistics of the form $(h_{v\tau})$ *for the person parameters.*

We show now by complete induction that this result is also valid for larger numbers of items: Let us assume that the model assumptions (1.64) are met for k items $(i = 1, ..., k)$. The response vector of person v to these k items is (a_{vi}). If another item is added to the test, then $s_{v(k+1)} = (a_{vi})$. Without loss of generality, $(a_{vi}) \in S_\rho$. Assume that person v has responded positively to item k + 1. The score vector corresponding to the augmented response vector $(a_{v1}, ..., a_{vk}, 1)$ is $(h_{v\tau})$. The score vector corresponding to (a_{vi}) is then:

$$(h_{v\tau}^{(\rho)}) = (h_{v1}, ..., h_{v(\rho-1)}, h_{v\rho} - 1, h_{v(\rho+1)}, ..., h_{vm}).$$

If $(h_{v\tau})$ shall be a sufficient statistic for the person parameters, the conditional probability

$$(1.72)\qquad p\{(a_{v1}, ..., a_{vk}, 1) \mid (h_{v\tau})\} =$$

$$= \frac{p\{(a_{vi})\}p\{a_{v(k+1)} = 1 \mid (a_{vi})\}}{\sum_{t=1}^{m} \sum_{(a_{vi}^*) \mid t} p\{(a_{vi}^*)\}p\{a_{v(k+1)} = 1 \mid (a_{vi}^*)\} + \sum_{(a_{vi}^{**}) \mid (h_{v\tau})} p\{(a_{vi}^{**})\}p\{a_{v(k+1)} = 0 \mid (a_{vi}^{**})\}}$$

must no longer depend on the person parameters. In (1.72), the summation $\sum_{(a_{vi}^*) \mid t}$ extends over all k-valued response vectors whose corresponding score vector is $(h_{vr}^{(t)})$, and the summation $\sum_{(a_{vi}^{**})\,|\,(h_{vr})}$ extends over all k-valued response vectors whose corresponding score vector is (h_{vr}). Substitution of

(1.73) $$p\{a_{v(k+1)} = 1 \mid (a_{vi})\} = \xi_v(a_{vi}) / (\xi_v(a_{vi}) + \sigma_{k+1}(a_{vi}))$$

(1.74) $$p\{a_{v(k+1)} = 1 \mid (a_{vi}^{*})\} = \xi_v(a_{vi}^{*}) / (\xi_v(a_{vi}^{*}) + \sigma_{k+1}(a_{vi}^{*}))$$

(1.75) $$p\{a_{v(k+1)} = 0 \mid (a_{vi}^{**})\} = \sigma_{k+1}(a_{vi}^{**}) / (\xi_v(a_{vi}^{**}) + \sigma_{k+1}(a_{vi}^{**}))$$

(1.76) $$p\{(a_{vi})\} = \frac{\prod_{r=1}^{m} (\xi_v + \psi_r)^{h_{vr}}}{\xi_v + \psi_\rho} \prod_{i=1}^{k} \frac{(\sigma_i - \psi_{r_{vi}})^{1 - a_{vi}}}{\xi_v + \sigma_i}$$

(1.77) $$p\{(a_{vi}^{*})\} = \frac{\prod_{r=1}^{m} (\xi_v + \psi_r)^{h_{vr}}}{\xi_v + \psi_t} \prod_{i=1}^{k} \frac{(\sigma_i - \psi_{r_{vi}^{*}})^{1 - a_{vi}^{*}}}{\xi_v + \sigma_i}$$

and

(1.78) $$p\{(a_{vi}^{**})\} = \prod_{r=1}^{m} (\xi_v + \psi_r)^{h_{vr}} \prod_{r=1}^{k} \frac{(\sigma_i - \psi_{r_{vi}^{**}})^{1 - a_{vi}^{**}}}{\xi_v + \sigma_i}$$

leads to the result that each of the expressions

(1.79) $$\frac{\xi_v + \psi_\rho}{\xi_v(a_{vi})} \quad \frac{\xi_v(a_{vi}^{*})}{\xi_v + \psi_t} \quad \frac{\xi_v(a_{vi}) + \sigma_{k+1}(a_{vi})}{\xi_v(a_{vi}^{*}) + \sigma_{k+1}(a_{vi}^{*})}$$

for t = 1, ..., m, and

(1.80) $$\frac{\xi_v + \psi_\rho}{\xi_v(a_{vi})} \quad \frac{\xi_v(a_{vi}) + \sigma_{k+1}(a_{vi})}{\xi_v(a_{vi}^{**}) + \sigma_{k+1}(a_{vi}^{**})}$$

must be independent of the person parameters. If we consider an arbitrarily selected response vector (a_{vi}^{**}), we can assume without restriction of generality that the parameters $\xi_v(a_{vi}^{**})$ and $\sigma_{k+1}(a_{vi}^{**})$ have been transformed by admissible multiplications by a common constant, so that the ratio on the right side in (1.80) equals 1. Likewise, we define without loss of generality

(1.81a) $\xi_v(a_{vi}^{**}) = \xi_v + \psi(a_{vi}^{**})$

and

(1.81b) $\sigma_{k+1} = \sigma_{k+1}(a_{vi}^{**}) + \psi(a_{vi}^{**})$,

so that

$$(1.82) \qquad \frac{\xi_v + \psi_\rho}{\xi_v(a_{vi})} = \frac{\xi_v + \sigma_{k+1}}{\xi_v(a_{vi}) + \sigma_{k+1}(a_{vi})}.$$

We also know that the parameters $\xi_v(a_{vi})$ and $\sigma_{k+1}(a_{vi})$ are determined only up to multiplication by a common constant. We may choose this constant so that the ratio on the left in (1.82) equals 1. Then, $\xi_v(a_{vi}) = \xi_v + \psi_\rho$ and $\xi_v(a_{vi}) + \sigma_{k+1}(a_{vi}) = \xi_v + \psi_\rho + \sigma_{k+1}(a_{vi})$. Hence, $\sigma_{k+1}(a_{vi}) = \sigma_{k+1} - \psi_\rho$.

Substitution of this result into (1.80) shows that the relations defined in (1.81) for an arbitrary response vector (a_{vi}^{**}) are also valid for all other response vectors (a_{vi}^{**}) whose corresponding score vector is (h_{vr}).

Substitution into (1.79) finally results in the relations

(1.82a) $\xi_v(a_{vi}^{*}) = \xi_v + \psi_t$

and

(1.82b) $\sigma_{k+1}(a_{vi}^{*}) = \sigma_{k+1} - \psi_t$

for all $(a_{vi}^{*}) \in S_t$, $t = 1, \ldots, m$, whose corresponding score vector is $(h_{vr}^{(t)})$. Since, according to our assumptions, the reaction probabilities $p\{a_{v(k+1)} \mid (a_{vi})\}$, $p\{a_{v(k+1)} \mid (a_{vi}^{*})\}$ and $p\{a_{v(k+1)} \mid (a_{vi}^{**})\}$ depend on the (partial) response vectors (a_{vi}), (a_{vi}^{*}) and (a_{vi}^{**}), respectively, only *via* their respective classes, the restrictions (1.64) follow for all remaining person and item parameters. This completes the proof.

Under the assumption $p\{a_{vi} \mid s_{vi} \epsilon S_r\} = p_{vir}$, *for all* $s_{vi} \epsilon S$ *and for all* $r = 1, \ldots, m$, *the dynamic test model*

$$(1.83) \qquad p_{vir} = \frac{\xi_v + \psi_r}{\xi_v + \sigma_i}, \; \psi_r < \sigma_i \text{ for all } i \in I \text{ and } r = 1, \ldots, m$$

is necessary and sufficient for the score vector (h_{vr}) *to be a sufficient statistic for the person parameters, independently of the selection and the sequence of items from I.*

With this proof, fairly general conditions for the existence of sufficient statistics for the person parameters in dynamic test models have been derived. However, it should be pointed out that these conditions are necessary and sufficient only for specifically objective comparisons of the structural item and transfer parameters (by conditional inference). If a specifically objective comparison of the person parameters is desired, it must be postulated in addition that the marginal distribution of the score vectors (h_{vr}) is not uniquely determined by the marginal item matrix ((n_{ri})). (Here, n_{ri} stands for the number of persons for whom $s_{vi} \epsilon S_r$ and who respond negatively to item i.) As I have shown elsewhere (KEMPF, 1976), this is possible in the dynamic test model (1.25) only when the transfer converges towards a fixed value as the number of solved items increases.

References

ALLERUP, P., SORBER, G.: The Rasch model for questionnaires. Report No. 16, The Danish Institute for Educational Research, 1975.

ANDERSEN, E. B.: Asymptotic properties of conditional maximum likelihood estimators. Journal of the Royal Statistical Society B *32*, 283–301, 1970.

– The asymptotic distribution of conditional likelihood ratio tests. Journal of the American Statistical Association *66*, 630–633, 1971.

– A computer program for solving a set of conditional maximum likelihood equations arising in the Rasch model for questionnaires. Research Memorandum, Princeton: Educational Testing Service, 1972.

– Conditional Inference and Models for Measuring. Mentalhygiejnisk Forlag, Copenhagen 1973a.

– Conditional inference for multiple-choice questionnaires. British Journal for Mathematical and Statistical Psychology *26*, 31–44, 1973b.

– A goodness of fit test for the Rasch Model. Psychometrika *38*, 123–140, 1973c.

BOUDON, R.: Mathematische Modelle und Methoden. Ullstein, Frankfurt 1973.

BRAMEL, D., TAUB, B., BLUM, B.: On observer's reaction to the suffering of his enemy. Journal of Personality and Social Psychology *8*, 384–392, 1968.

BRANDT, U., KÖHLER, B.: Norm und Konformität. In: Handbuch der Psychologie, Vol. 7. Hogrefe, Göttingen 1971.

COLEMAN, J. S.: Introduction to Mathematical Sociology. The Free Press, Glencoe, Ill. 1964.

CRONBACH, L. J., GLESER, G. C.: Psychological Tests and Personnel Decisions. University of Illinois Press, Urbana 1965.

DANN, H. D.: Müssen Aggressionen ausgelebt werden? In: Schmidt-Mummendey, A., Schmidt, H. D., (eds.): Aggressives Verhalten. Juventa, Munich 1971.

FESHBACH, S.: The catharsis hypothesis and some consequences of interaction with aggressive and neutral play objects. Journal of Personality *24*, 339–462, 1956.

FISCHER, G. H.: Neue Entwicklungen in der psychologischen Testtheorie. In: Fischer, G. H. (ed.): Psychologische Testtheorie. Huber, Bern/Stuttgart/Wien 1968.

– Einige Gedanken über formalisierte psychologische Theorien. Psychologische Beiträge *13*, 376, 1971.

– Conditional maximum-likelihood estimation of item parameters for a linear logistic test-model. Research Bulletin No. 9, University of Vienna: Department of Psychology, 1972.
– Einführung in die Theorie Psychologischer Tests. Huber, Bern/Stuttgart/Wien 1974.
FISCHER, G. H., ALLERUP, P.: Rechentechnische Fragen zu Rasch's eindimensionalem Modell. In: Fischer, G. H. (ed.): Psychologische Testheorie. Huber, Bern/Stuttgart/Wien 1968.
FISCHER, G. H., ROPPERT, J.: Monte-Carlo-Untersuchungen an einem Intelligenzmodell mit vier Faktoren. In: Heckhausen, H. (ed.): Bericht 24. Kongress der Deutschen Gesellschaft für Psychologie. Hogrefe, Göttingen 1965.
FISCHER, G. H., SCHEIBLECHNER, H. H.: Algorithmen und Programme für das probabilistische Testmodell von Rasch. Psychologische Beiträge *12*, 23–51, 1970.
GUTTMAN, L.: The basis for Scalogram-Analysis. In: Stauffer, S. A., et al. (eds.): Measurement and Prediction. Princeton University Press, Princeton 1950.
HILKE, R., HÖLLBACHER, M., KEMPF, W. F.: Aggressives Handeln in einer Zwei-Personen-Interaktionssituation. Forschungsbericht 41 des Sonderforschungsbereichs 22. Nürnberg: Sozialwissenschaftliches Forschungszentrum, 1973.
HOLZKAMP, K.: Theorie und Experiment in der Psychologie. De Gruyter, Berlin 1964.
– Zur Problematik der Realitätsverdoppelung in der Psychologie. Psychologische Rundschau. *16*, 209–222, 1965.
– Wissenschaft als Handlung. De Gruyter, Berlin 1968.
JÜTTEMANN, G.: Was nützen Eigenschaftskonstrukte? Psychologische Rundschau *12*, 91–114, 1972.
KALVERAM, K. TH.: Über Faktorenanalyse. Archiv für Psychologie *122*, 92–118, 1970a.
– Probleme der Selektion in der Faktorenanalyse. III. Die „Invarianz" von Faktorenlösungen unter Selektion. Archiv für Psychologie *122*, 223–230, 1970b.
KAUFMANN, H.: Definitions and methodology in the study of aggression. Psychological Bulletin *64*, 351–364, 1965.
KEMPF, W. F.: Grundlagenuntersuchungen zum Thematischen Apperzeptionstest. Philosophische Dissertation Wien, 1970.
– Probabilistische Modelle experimentalpsychologischer Versuchssituationen. Psychologische Beiträge *14*, 16–37, 1972a.
– Zur Bewertung der Faktorenanalyse als psychologische Methode. Psychologische Beiträge *14*, 610–625, 1972.
– Basisprobleme der Diagnostik der Aggressivität. In: Eckensberger, L. H. (ed.): Bericht vom 28. Kongress der Deutschen Gesellschaft für Psychologie. Hogrefe, Göttingen 1974a.
– Dynamische Modelle zur Messung sozialer Verhaltensdispositionen. In: Kempf, W. F. (ed.): Probabilistische Modelle in der Sozialpsychologie. Huber, Bern/Stuttgart/Wien 1974b.
– A dynamic model for measuring individual and situative influence on social behavior. In Proceedings of the COMPSTAT-Symposium on Computational Statistics. Physica, Vienna 1974c.
– Einige dynamische Messmodelle und ihre Anwendung in der Denk- und Lernpsychologie. In: Kempf, W. F. (ed.): Drei Arbeiten über verallgemeinerte Rasch-Mo-

delle. IPN-Arbeitsbericht 10. Institut für die Pädagogik der Naturwissenschaften, Kiel 1974d.
– A dynamic test model and its use in the microevaluation of instructional material. In: Spada, H., Kempf, W. F. (eds.): Structural Models of Thinking and Learning. Huber, Bern/Stuttgart/Wien 1976 (in press).
KEMPF, W. F., HAMPAPA, P.: The numerical solution of a set of conditional estimation equations arising in a dynamic test model. In: Kempf, W. F., Hampapa, P., Mach, G. (eds.): Conditional maximum likelihood estimation for a dynamic test model. Arbeitsbericht 13, Institute for Science Education at the University of Kiel, 1975.
KEMPF, W. F., MACH, G.: A Fortran program for CML-estimation in a dynamic test model. In: Kempf, W. F., Hampapa, P., Mach, G. (eds.): Conditional maximum likelihood estimation for a dynamic test model. Arbeitsbericht 13, Institute for Science Education at the University of Kiel, 1975.
KRISTOF, W.: Einige Skalenfragen. In: Fischer, G. H. (ed.): Psychologische Testtheorie. Huber, Bern/Stuttgart/Wien 1968.
LAUTMANN, R.: Wert und Norm. Begriffsanalysen für die Soziologie. In: Beiträge zur sozialen Forschung 5. Westdeutscher Verlag, Opladen 1971.
LUCE, R. D.: Individual Choice Behavior. Wiley, New York 1959.
MERZ, F.: Über die Erfassung aggressiver Einstellungen mit Hilfe von Fragebogen. Psychologische Beiträge *5*, 402–415, 1960.
MICKO, H. C.: Eine Verallgemeinerung des Messmodells von Rasch mit einer Anwendung auf die Psychophysik der Reaktionen. Psychologische Beiträge *12*, 4–22, 1970.
NEYMANN, J., SCOTT, E. L.: Consistent estimates based on partially consistent observations. Econometrica *16*, 1–32, 1948.
PFANZAGL, J.: Theory of Measurement. Physika Verlag, Würzburg 1968.
RASCH, G.: Probabilistic Models for Some Intelligence and Attainment Tests. Nielson & Lydiche, Copenhagen 1960.
– On general laws and the meaning of measurement in psychology. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability *5*, 321–333, 1961.
– Kolloquium über Messmodelle. Mimeo 1965.
– An individualistic approach to item analysis. In: Lazarsfeld, P. F., Henry, N. W. (eds.): Readings in Mathematical Social Science. Science Research Associates, Chicago 1966a.
– An individual-centered approach to item analysis with two categories of answers. Proceedings of the NUFFIC International Summer Session in Science at „Het Oude Hof", The Hague 1966b.
– An item analysis which takes individual differences into account. British Journal of Mathematical and Statistical Psychology *19*, 49–57, 1966c.
ROHRACHER, H.: Einführung in die Psychologie. Urban & Schwarzenberg, Wien 1963.
SCANDURA, J. M.: Structural Learning I: The Theory and Empirical Research. Gordon & Breach, New York 1973.
SCHEIBLECHNER, H.: The separation of individual and system-influences on behavior in social contexts. Acta Psychologica *35*, 442–460, 1971a.

– A simple algorithm for CML-parameter-estimation in Rasch's probabilistic measurement model with two or more categories of answers. Research Bulletin No. 5. Department of Psychology, University of Vienna 1971b.

SELG, H.: Diagnostik der Aggressivität. Hogrefe, Göttingen 1968.

SUPPES, P., ZINNES, J. L.: Basic measurement theory. In: Luce, R. D., Bush, R. R., Galanter, E. (eds.): Handbook of Mathematical Psychology. Vol. I. Wiley, New York 1963.

TRAXEL, W.: Einführung in die Methodik der Psychologie. Huber, Bern/Stuttgart/Wien 1964.