

What Can We Learn from the Discussion of Personality Questionnaires for the Construction of Temperament Inventories?

Alois Angleitner and Rainer Riemann

The intention of this chapter is to analyze common and distinct procedures and problems that are prevalent in the construction and use of temperament and personality questionnaires. We start from the assumption that meaningful differences exist between the concepts of personality and temperament, and outline some consequences of these differences for the generation of questionnaire items and the subsequent item-selection process. The main focus of this chapter is the assessment of temperament in adults.

Personality and Temperament: Where Are the Differences?

The terms personality and temperament are both used to refer to temporarily relatively stable individual differences in behavior. Researchers in both fields make use of the trait concept for a convenient description of individual differences. Despite this obvious communality, and notwithstanding the fact that the border between both concepts is fuzzy, "personality" and "temperament" can be meaningfully differentiated.

A brief look at a few prominent definitions reveals both substantial agreement in the conceptualization of temperament, as well as some distinct features of the two concepts. Wundt (1903) defines temperament as the habitual emotional reactivity of an individual. In addition, physiological and biochemical factors are accentuated by McDougall (1923). Typical motor reactions of an individual as belonging to the realm of tempera-

Alois Angleitner and Rainer Riemann • Department of Psychology, University of Bielefeld, D-4800 Bielefeld, Federal Republic of Germany.

ment have been emphasized in early textbooks (Bloor, 1928; Downey, 1923). In the view of Allport (1937), personality is defined as a psychophysical system. The raw materials of personality are physique, intelligence, and temperament. These raw materials are seen as dependent on the genetic inheritance. Temperament is defined by Allport as follows:

The characteristic phenomena of an individual's emotional nature, including his susceptibility to emotional stimulation, his customary strength and speed of response, the quality of his prevailing mood, and all peculiarities of fluctuation and intensity in mood, these phenomena being regarded as dependent upon constitutional make-up, and therefore largely hereditary in origin. (Allport, 1961, p. 34)

Cattell (1946, p. 14) views temperament as a part of personality referring to temper, emotionality, and constitutional reactivity. Later he refines his view:

Temperamental traits are definable by exclusion as those traits which are unaffected by incentive or complexity. These are traits like highstrungness, speed, energy, and emotional reactivity, which common observation suggests are largely constitutional. (Cattell, 1950, p. 35)

In a more recent publication by Buss and Poley (1976, p. 76) we read:

The concept of temperament is rather straightforward and presents little difficulty—referring as it does to individual differences in “temper,” that is, characteristics which are largely stylistic ways of behaving in social situations.

Summarizing the numerous approaches to the study of temperament and the relation of temperament to personality, Strelau (1987) stresses five discriminatory features in which temperament and personality are different:

1. Determinants of development (temperament: biological, personality: social).
2. Developmental stages of shaping (temperament: childhood, personality: adulthood).
3. Referring populations (temperament: animals and man, personality: man).
4. Content-saturated behavior (temperament: absent, personality: present).
5. Central regulating function (temperament: insignificant, personality: significant).

In agreement with these classifications is also the definition of temperament given by Plomin (1981). He states that temperament involves those dimensions of personality that are largely genetic or constitutional in origin, exist in most ages and most societies, show some consistency across situations, and are relatively stable. We would conclude from this statement that temperament is most clearly detectable in formal behavioral characteristics.

How to Find the Real Temperament Traits

The fundamental question for both personality and temperament research concerns the universe of traits that are considered for study. We may find stable individual differences to a large degree. Not all differences are relevant for temperament even if

they meet the above-mentioned criteria. We may think of an example given by Goldberg: Some people are able to wiggle their ears, some are not. Two basic approaches for the selection of relevant traits can be distinguished: the inductive lexicographic strategy and the hypothetico-deductive.

The lexicographic approach is based on the "sedimentation hypothesis" formulated by Cattell in the 1940s. This hypothesis states that those individual differences are encoded in our language that are necessary and important for human conduct. Thus, the personality/temperament vocabulary represented in natural languages (personality-descriptive adjectives, nouns, and verbs) is analyzed to yield a finite set of attributes that denote socially relevant individual differences (see John, Angleitner, & Ostendorf, 1988). The lexicographic approaches by Allport and Odbert (1936), Norman (1963), Goldberg (1981, 1982) in the USA, Hofstee (1976, 1977) and Brokken (1978) in the Netherlands, and by Angleitner, Ostendorf, and John (1990) in Germany include a broad category labeled "stable temperament and character traits." The sets of adjectives and nouns assigned to this category should be checked for the selection of temperament traits using the following criteria:

1. *Nature criterion.* In the Dutch trait taxonomy studies Hofstee and his coworkers developed the person and the nature criterion for the selection of trait terms. For the nature criterion, judges were given the following instruction: "Call an adjective a personality descriptive adjective if it may replace the dots in the following sentence: He (She) is . . . by nature" (Brokken, 1978). Rusalov (personal communication) used this criterion independently formulated in the opposite direction. He asked his subjects to indicate to what degree a trait (e.g., aggression) can be changed by social influences (e.g., education, training). Trait terms judged with good agreement as high on the nature criterion and little changeable could be considered as temperament traits.

2. *Independence of Culture.* A trait term that is found in more than one language should be regarded with more confidence as a temperament trait, compared to one found in only one language. If we compare the trait terms listed in the English, German, and Dutch lists, we have an empirical base for the selection of temperament traits.

3. *Applicability to children.* Experienced judges could be used to find out whether a trait term can literally be used to denote individual differences among young children and adults. However, ratings of the applicability to animals seem less practicable.

4. *Content saturation.* In the same way, judges could be asked whether a trait term refers more to *what* a person does or to *how* the person performs certain behaviors.

We expect that these criteria are not independent of each other. Trait terms that can be applied to children and adults should be rated as little content-saturated and high on the nature criterion. In addition, these traits should be found in other languages too.

However, we are convinced that the number of temperament traits may still be large. Thus, robust higher order factors should be established in cross-cultural self- and peer-rating studies. These factors could be considered as temperament factors insofar as there is independent evidence that they meet the criteria proposed by Strelau (1987).

The taxonomy of personality-descriptive verbs, begun by Goldberg (1982) and De Raad, Mulder, Klosterman, and Hofstee (1988), seems promising for the temperament domain. For the selection of personality-descriptive verbs, De Raad *et al.* used the instructions: "If, under the circumstances, one person (verb)s more than others, does that

tell something about his/her personality?" This frame sentence could also be adapted for the temperament domain. This procedure allows one to establish a behavioral terminology that avoids situational and context specificity—a problem in act and item formulations—as well as broad generalizations—a problem in using trait adjectives or nouns.

No doubt the lexicographic approach, which starts from an exhaustive sampling of terms and requires a cumbersome reduction, is not very economical. However, if we are interested in the comprehensive study of temperament expressed in behavior, there is no alternative way.

The hypothetico-deductive approach takes its starting point from more or less elaborated temperament theories. The basic constructs of these theories may be derived from physiological research (e.g., the numerous theories based on the Pavlovian properties of the nervous system), clinical observation (e.g., Carey, 1970; Thomas & Chess, 1977), factorial studies of selected variables (e.g., Eysenck, 1944) or other sources. Take as an example Pavlov's theorizing about the concepts "strength of excitation," "strength of inhibition," "balance," and "mobility." Strelau (1983) reformulated these conceptualizations and operationalized them in behavioral terms. For the construction of the STI-R (Strelau, Angleitner, Bantelmann, & Ruch, 1990) each scale definition was partitioned in several units referring to smaller behavioral domains. For example, the mobility component includes the following five behavioral domains:

1. Reacts adequately to unexpected changes in the environment.
2. Adapts quickly to new surroundings.
3. Passes easily from one activity to another.
4. Changes mood lightly from positive to negative and the reverse, according to the meaning of the situation.
5. Prefers situations which require to perform different activities simultaneously.

These behavioral units constituted the basis for the generation of items which were nominated by experts familiar with the theory.

At the first glance the hypothetico-deductive approach seems much more efficient than the lexicographic approach for finding relevant temperament traits. However, as there are numerous theories of temperament, which are at least partially supported by empirical results, the question of the structure of temperament traits and the relative importance of postulated constructs arises again. To find universal dimensions of temperament large-scale construct validation studies are needed.

Temperament and the Item Content of Temperament Scales

Given the clarifications referring to the differences between the concepts of temperament and personality it may be argued that the items of temperament scales should differ from personality scale items in several respects. We refer here again to the features of temperament traits given by Strelau (1987) and examine their implications for the construction of questionnaire items.

1. *Biological factors play a crucial role in determining temperament.* This criterion requires focus on behaviors or characteristics of persons that are probably determined

by their biological makeup, instead of formulating items that refer to learned reactions. Typical items capture psychophysiological reactions and symptoms that are not under voluntary control, including motor reactions and their inhibition (e.g., "I move a lot in bed," DOTS-R, Windle & Lerner, 1985, 1986). The assumption that individual differences in temperament are inherited may be given more weight in using appropriate designs, for example, the comparison between monozygotic and dizygotic twins for the item selection, as was done by Buss and Plomin (1975) in their construction of the EASI. The use of heritability coefficients as a criterion for the construction of tests was already suggested by Jones (1971).

2. *Temperament traits focus on formal, stylistic characteristics of behavior.* This feature of temperament traits has obvious consequences for the construction of questionnaire items. The items should refer to how people perform behaviors rather than what behaviors are performed. Descriptions of specific contexts and social situations should be absent in temperament items. For covering the stylistic characteristics adverbs can be used, because adverbs express these characteristics most clearly (e.g., "Do you speak rapidly?" STI, Strelau, 1983).

3. *Temperament is shaped in childhood.* Behavioral characteristics that show stability over a person's life span meet this criterion best. Thus, the item stability over reasonably long time periods might be regarded as a criterion in the item-selection process. In addition, items should refer to behavior reactions that are observable in children and adults.

4. *Temperament refers to phenomena that can be observed in man and animal.* Obviously this postulate cannot be utilized directly for the construction of temperament items. However, it can be deduced that temperament questionnaires should consist of items that can be judged reliably by observers. Those items should be dismissed that refer exclusively to nonobservable internal reactions (e.g., "Can you suppress momentary moods of dejection?" STI).

5. *Central regulating functions are insignificant for temperament traits.* This criterion excludes cognitive aspects of behavior and items that refer to attitudes, values, and goals. Instead, the individuals' regulation of the stimulative value of activities and surroundings should be studied by focusing on preferences for certain types of activities and environmental conditions.

To explore the differences between temperament and personality questionnaires currently in use, the items of the following temperament questionnaires were classified according to a taxonomy proposed by Angleitner, John, and Loehr (1986): (1) the Strelau Temperament Inventory (STI, Strelau, 1983), (2) the short version of the revised STI (STI-RS, Strelau, Angleitner, Bantelmann, & Ruch, 1990), (3) the Questionnaire for the Measurement of the Structure of Temperament (QST, Rusalov, 1989), (4) the adult version of the Revised Dimensions of Temperament Survey (DOTS-R, Windle & Lerner, 1985, 1986), and (5) the Emotionality Activity Sociability Impulsivity Temperament Survey (EASI-III, Buss & Plomin, 1975).

The taxonomy consists of seven central categories: (1) descriptions of reactions, which are subdivided into overt, covert, and bodily reactions (symptoms); (2) trait attributions, subdivided into unmodified or modified by, for example, situational context specifications or frequency qualifiers; (3) wishes and interests; (4) biographical facts; (5) attitude and belief statements; (6) reactions of other persons to the subject; and (7)

bizarre item content. Three raters independently provided these classifications. The average percentage agreement between the raters was 60%. For further analysis, each item was assigned to the category that had been selected by the majority of judges. The results of this categorization are reported in Table 1, together with the classifications presented by Angleitner *et al.* (1986) for the MPI and MMQ (Eysenck, 1962, 1943), the EPI (Eysenck & Eysenck, 1964), the 16PF (Cattell & Eber, 1964), and the MMPI (Hathaway & McKinley, 1964). In the study of Angleitner *et al.* the judges were given the additional option not to subclassify an item.

In Table 1 the percentage of items assigned by the majority of the raters to each category is presented separately for each questionnaire. In the column headed "left unclassified," the percentage of those items is given that were not assigned to the same category by more than 50% of the judges. This percentage is particularly high for the EASI III. An example may illustrate the reading of Table 1: 58% of the STI items ($N = 134$) were assigned to the category descriptions of overt reactions, 27% were regarded as "covert reactions," 4% and 1% asked for unmodified and modified trait attributions, and 2% for the subjects' wishes and interests. No item was assigned to the remaining categories. Eight percent of the items were left unclassified.

A look at the temperament scales reveals that they are made up mostly by items asking for:

1. Overt reactions of the testperson that are in principle observable by the public (see Column 1a in Table 1).
2. Covert reactions or internal reactions that are private and generally not observable by others (internal sensations, feelings, cognitions; see Column 1b in Table 1). Some examples may illustrate this:

"I smile often." (DOTS-R)

"I move a lot in bed." (DOTS-R)

"Do you wake quickly and without difficulty?" (STI)

"I often feel like crying." (EASI-III)

"Do you feel bored or sleepy when performing monotonous work?" (STI)

The percentages of items assigned to these categories are much higher for the temperament inventories (including Eysenck's questionnaires) than for the 16PF and MMPI. However, substantial differences can be observed between the different temperament scales within the same questionnaire (see Angleitner *et al.*, 1986, for details). For measuring emotionality (neuroticism) the test constructors concentrate on covert reactions regarding item content (partly also on symptoms), whereas for the scales measuring activity, rhythmicity, sociability, or impulsivity, they generate items with mostly overt behavioral item content. For example, the EPI (forms A and B) extraversion scales are build by 51% of items asking for overt reactions, the neuroticism scales are made up largely of items referring to covert reactions (53%) and physical symptoms (25%).

Items asking for trait attributions (Columns 2a and 2b in Table 1) were not found in the STI-RS and were rare in the STI, the QST, and the DOTS-R. However, a substantial portion of the items of the EASI-III and Eysenck's questionnaires was assigned to this category. The percentage of "trait attributions" found in these questionnaires is comparable to that of personality inventories. Examples for this category are:

"Are you hot-tempered?" (STI)

Table 12.1. Relative Frequencies of Items in Each Category from Temperament and Personality Questionnaires

Questionnaire	Number of items	Description of reactions ¹					Trait attributions							Indirect references							
		1a	1b	1c	Not subclassified	2a	2b	Not subclassified	3	4	5	6	7	Left unclassified	3	4	5	6	7	Left unclassified	
		overt	covert	symptoms	subclassified	unmodified	modified	subclassified	Wishes	Biography	Attributions	Others	Bizarre		Wishes	Biography	Attributions	Others	Bizarre		
STI	134	58.21	26.87	0.0	0.0	3.73	0.75	0.0	2.24	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.21
STI-RS	84	67.86	27.38	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.76
QST	105	54.29	34.29	0.0	0.0	3.81	0.0	0.0	0.0	3.70	0.0	0.0	0.0	0.0	0.0	0.0	0.95	0.0	0.0	0.0	6.67
DOTS-R	54	70.37	24.07	0.0	0.0	3.70	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.85
EASI III	50	30.00	30.00	0.0	0.0	14.00	2.00	0.0	0.0	14.00	2.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	24.00
MPI ²	48	18.80	41.70	2.1	8.2	18.80	4.20	2.0	0.0	18.80	4.20	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.20
MMQ ²	56	19.60	23.20	17.9	8.9	10.70	7.10	7.2	0.0	10.70	7.10	7.2	0.0	1.8	0.0	0.0	0.0	0.0	0.0	0.0	3.60
EPI-A/B ²	57/57	33.35	29.80	10.5	7.95	7.00	4.40	0.85	0.0	7.00	4.40	0.85	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	5.25
16PF-A/B ²	171/171	17.80	22.55	2.65	12.60	2.95	0.30	2.60	12.55	2.95	0.30	2.60	12.55	3.2	12.3	1.5	1.5	1.5	0.0	0.0	9.05
MMPI ²	404	10.40	27.00	13.10	11.10	3.20	1.00	2.00	4.70	3.20	1.00	2.00	4.70	5.7	5.9	2.2	2.2	3.0	3.0	3.0	10.60

¹See text for description of the categories.

²Data taken from Angleitner *et al.*, 1986.

"I am very sociable." (EASI-III)

The "trait attributions" do not meet the above-mentioned criterion of public observability.

Items referring to wishes and interests (Column 3 in Table 1), to biographical facts (Column 4), attitudes and beliefs (Column 5) and other persons' reactions (Column 6) as well as bizarre items (Column 7) are nearly absent from temperament scales. Items assigned to these categories make up a substantial part of the 16PF and the MMPI.

Although not all temperament and personality questionnaires were compared, the results point at meaningful differences between the questionnaires in both domains. The high percentage of items referring to overt reactions is especially in line with the criteria derived for temperament items. The fact that authors of personality questionnaires include items in their original item pools which do not refer to specific reactions and that these items "survive" the empirical item-selection process may reflect the higher weight given to integrative functions of behavior. Systems of values and motivational constructs can be operationalized in questions that refer to a person's "wishes and interests" and "attitudes and beliefs." The social nature of personality constructs may find its expression in items referring to a person's biography and other persons' reactions.

In our reflection on the item contents of temperament scales we will now concentrate on three unsolved problems:

1. A glance at different temperament scales documents that the test constructors are probably not aware of their mixing together frequency and intensity aspects in their item formulations, for example, "I frequently get upset," "I am somewhat emotional" (EASI-III), and "I often stay still for long periods of time," "Even when I am supposed to be still, I get very fidgety after a few minutes" (DOTS-R).

It may be the case that some temperament traits, for such as the various constructs referring to a person's activity, are more connected with frequency assumptions, others more with intensity assumptions, such as emotionality, impulsivity, or arousability. Following the arguments by Strelau (1987) in defining temperament, it seems reasonable that an intensity conception is theoretically more adequate than a frequency conception. People may differ in their actions dependent on their different living circumstances, but still they may be similar in the intensity by which they perform their actions: They are called phlegmatic or choleric.

In the future it may be worthwhile to compare items based on frequency assumptions with items measuring the intensity of reactions. The missing correspondence between the postulated physiological mechanism and the proposed psychometric scales (Strelau, 1990) may find a partial explanation in the fact that most psychophysiological measures capture the intensity of reactions, whereas frequency conceptualizations are preferred in the psychometric tradition. No one would *a priori* expect much convergence between these different conceptualizations.

A look at the descriptions and items of temperament scales reveals that the authors are somewhat inclined to use an abilities conception. Examples of such items are: "Do you have difficulties in adapting to a new daily schedule?" (STI); "Are you able to work while waiting for guests?" (STI); and "I get bored easily" (EASI-III). The abilities conception of personality was proposed from independent perspectives by Fiske and Butler (1963) and Wallace (1966). The main focus of the abilities conception are individ-

ual differences in persons' capabilities to perform behaviors required by the situation rather than in their typical or average behaviors. It may be worthwhile to consider the abilities conception in temperament research more explicitly, not only for the generation of questionnaire items, but also for the construction of laboratory tests of behavioral performance.

2. The vast majority of authors claim in their writings that stylistic components of behavior are central for temperament constructs. How should these stylistic components of behavior be incorporated into the item formulations? In general, we see two solutions:

1. The items for temperament should be formulated in a broader way, covering different content domains. However, this may lead to very abstract formulations involving much interpretation on behalf of the answering subject. Examples are:

"Are you capable of adapting your conduct to the behavior of others in a group when necessary?" (STI)

"My life is fast-paced." (EASI-III)

"I move towards new situations." (DOTS-R)

2. A second possibility would be to start from the opposite direction: Items should be generated that refer to specific actions in a variety of different contexts, using verbs instead of "I like," "I tend," "Do you easily," "Are you able," "Can you," etc. This means that one concentrates on the actions that a hot-tempered person is likely to show (e.g., "I raise my voice in a debate," instead of "I have trouble controlling my impulses") and gathers these reactions across many different situations. The strategy developed by Buss and Craik (1980) in their act-frequency approach may be used for eliciting relevant actions for different temperament traits.

3. A serious problem arises if "parallel" versions of a questionnaire are constructed, with the intention to measure the same temperament dimensions in young children and adults. No generally accepted procedure exists that ensures the equivalence of items across age groups. What are the behavioral expressions for impulsivity that are equivalent in meaning for different age periods, especially in children? In general, authors of temperament questionnaires have developed two strategies for solving this issue: (1) For each age group different items are formulated, which ask for appropriate behavioral reactions. The equivalence of these items across age groups is accepted by face validity. For example, the item "My child splashes hard in the bath and plays actively," is seen to be equivalent to the item "Child seems to have difficulty in sitting still, may wriggle a lot or get out of seat" (NYLS-scales, Thomas & Chess, 1977); and (2) either the scales or the items are constructed in such a way that they can be used in a broad range of age groups with minor modifications (EASI-III; DOTs-R).

Steps in the Construction of a Temperament Scale

Suppose we know the relevant temperament traits for which we would like to develop an inventory. The question arises, what should a test constructor consider in developing a successful instrument? Several issues should be considered:

1. *Use of a rational scale construction strategy.* A strategy similar to the procedure of eliciting behavioral acts used by Buss and Craik (1980, 1983) may be adopted. We

may ask subjects (parents, teachers, etc.) if they know somebody whom they will call, for instance, "active," and then ask how this person typically behaves to elicit this impression. However, to avoid the problem of specificity of context, these items would most likely have to be reformulated.

Another strategy is to bring together experts of these temperament concepts for which scales should be developed (preferably from different nations) and to ask them to generate items. It is important to define the behavioral domains very precisely beforehand and to find agreement concerning some basic rules about what constitutes a good item. The suggestions given by Jackson (1970) are recommended.

The items should be (1) short and clearly understandable, (2) free from extreme levels of social desirability, (3) diverse in content as to cover the universe of human conduct, (4) applicable to adults in different cultures and not biased toward particular populations, as for example, college students or males, (5) logically related to the construct under consideration and at the same time not overlapping with similar but irrelevant constructs, and (6) balanced in their keying.

To reduce the item pool to a manageable size, the items may be judged for prototypicality following Rosch and Mervis's (1975) procedures. Only items that are regarded as good examples for the behavioral domain in question should be considered for the empirical scale construction. The application of such strategies limits the personal idiosyncrasies which may enter if one researcher alone generates the whole item pool. It allows a more representative sampling of the items.

2. *Checking the syntactic characteristics of the items.* The empirical fact that, for example, about 30–40% of the items of Eysenck's inventories are judged as not immediately understandable (Angleitner *et al.*, 1986) should alert us to the linguistic characteristics of items. It has been documented that syntactic properties are related to the psychometric item and scale characteristics (see Angleitner *et al.*, 1986).

3. *Consider the item stability and the homogeneity issue.* It has been argued above that item stability should be given an important rank for the selection of items. The homogeneity/heterogeneity of items for a proposed scale has to be considered in advance. Theoretical arguments have to be considered to decide whether the trait under study covers a narrow domain of behaviors or is conceptualized more broadly and includes a variety of facets. Keeping the number of items constant, the internal consistency of a broader scale should be lower compared to a narrow scale. Therefore, lower internal consistencies of a scale may be acceptable. However, the values should not fall short of a critical border of about .70. In line with Loevinger's (1957) emphasis on the substantial validity issue, we do not recommend the *blind* application of item-analytical strategies for finding homogeneous scales. Really homogeneous scales are yielded if items are repeated or phrased tautologically.

In case the temperament traits are postulated to be orthogonal, the use of factor-analytic strategies to warrant unidimensionality is preferable (Windle, 1988). However, if the concepts are theoretically correlated, the application of factor analysis is not justified. Take the example of the correlated Pavlovian concepts "strength of excitation" and "mobility." In constructing scales for these concepts, the authors of the STI-R relied on the multitrait-multimethod approach formulated by Campbell and Fiske (1959). In the STI-R, only items that correlated higher with their own scale compared with the related or unrelated scales in question have been accepted (Strelau *et al.*, 1990). The procedure

is similar to the differential reliability index as proposed and used by Jackson (1967) for developing the Personality Research Form (PRF).

4. *Consider response distortions in the item generation and in the construction process.* In hundreds of studies it has been shown that personality scale values may be distorted by several factors. These factors have been called tendencies to say yes or no, faking good or bad, social desirability, etc. In principle, response distortions cannot be avoided, but it seems possible to reduce them. For instance, the yes-saying tendency may be reduced by balancing the item-keying. Furthermore, if the subjects are informed about such distortions in the instructions, the chance to avoid response distortions may be enhanced.

Numerous strategies to control social desirability have been proposed. However, until now no procedure could be established that clearly helps to increase the validity of a scale. In the construction of the STI-R this problem was treated in the following way: All the items were rated by 20 judges concerning their social desirability value. The items with the most extreme mean ratings were selected for a social desirability scale. However, although the items selected for the content scales correlated higher with their own scales than with the social desirability scale, it turned out that the content scales showed high correlations with this new social desirability scale (SD). Furthermore, this SD scale did not correlate substantially with other social desirability or control scales (EPQ-Lie scale, PRF-Infrequency scale). This result shows that the initial item pool for the STI-R was already very high in content saturation. Nevertheless, rating the items' social desirability values is recommended for reformulating or eliminating items with extreme values. This brings more balance and neutralization in social desirability.

In his reexamination of response tendencies Paulhus (1986) made a distinction between self-deception and impression-management. He underlines that "items asking about overt behavior were minimally subject to self-deception" (p. 153). He also concludes that "impression management may also be usefully controlled by asking about behaviors with neutral desirability value" (p. 153). It follows that the application of these procedures helps to reduce the possible influence of response distortions.

5. *Consider convergent and discriminant validity in construction.* Temperament scales contain mainly items that express overt behaviors. This implies that they are good candidates for using samples of observers (peers, spouses, friends, colleagues, parents, relatives, etc.). Self- and observer ratings on the item as well as on the scale or facet levels can easily be carried out, resulting in a multitrait-multimethod design (Campbell & Fiske, 1959). In general, researchers continue to document the validity of their scales in pointing to the convergent validities by correlating scales with conceptually similar scales. But very seldom do authors consider some possible item-overlap. Likewise, a clear theoretical deduction regarding for which concepts zero or moderate correlations are expected is missing in almost all studies. This seems to be true for personality as well as temperament inventories.

The application of these guidelines, which are the result of our experience with personality scales, will, it is hoped, improve the construction of temperament scales. In asking for specific reactions, temperament inventories are relying more on direct modes of self-description. Such direct modes have already been shown to possess more satisfactory item-response statistics (Angleitner *et al.*, 1986). It follows that temperament traits should be good candidates for self-report measures.

The Cross-Cultural Adaptation of Temperament Scales

Temperament traits should be measurable in different cultures. The applicability of items in diverse cultures and nations should be considered in the initial item-generation process. By using experts, a fairly representative sample of items for a temperament trait in question can be developed. This initial item pool may then be considered as the item universe, and translated by several independent translators. The translators should be instructed about the concepts, which should be measured by this item set. Bilingual samples of subjects may be used for checking the equivalence of item meanings. Furthermore, it seems important to us that, for the empirical construction process, the procedures are comparable and standardized. The aim of such a cross-cultural research cooperation would be the development of comparable, multi-language temperament scales that are relatively culture-free.

Conclusions

In this chapter we have argued that the conceptual differences between "temperament" and "personality" should find their expression in the item content and construction strategies of temperament scales. Despite the fact that individual differences in temperament are regarded as biologically determined—and, therefore, that psychophysiological assessment may seem to be more appropriate—sufficient variation in behavior is observable to warrant the measurement of temperament via self-reports. We see five major and largely unresolved issues that require further theoretical and empirical efforts:

1. The scope and structure of the universe of temperament traits measured via self-reports is as yet not studied systematically.
2. Robust correlations of self-report data and psychophysiological measures have to be established.
3. Cross-cultural research programs are rare and scattered.
4. The equivalence of temperament scales across age groups needs further clarification.
5. Item formulations that refer to the frequency of performing certain behaviors, the intensity of behavioral expressions, and the capability of showing a behavior, when it is required by the situation, should be disentangled.

References

- Allport, G. W. *Personality: A psychological interpretation*. New York: Holt, 1937.
- Allport, G. W. *Pattern and growth in personality*. London: Holt, Rinehart & Winston, 1961.
- Allport, G. W., & Odbert, H. S. Trait-names: A psycho-lexical study. *Psychological Monographs*, 1936, 47, No. 211.
- Angleitner, A., John, O. P., & Loehr, F. J. It's what you ask and how you ask it: An itemmetric analysis of personality questionnaires. In A. Angleitner & J. S. Wiggins (Eds.), *Personality assessment via questionnaires: Current issues in theory and measurement*. Berlin: Springer, 1986.

- Angleitner, A., Ostendorf, F., & John, O. P. Towards a taxonomy of personality-descriptors in German: A psycho-lexical study. *European Journal of Personality*, 1990, 4, 89-118.
- Bloor, C. *Temperament*. London: Methuen, 1928.
- Brokken, F. B. *The language of personality*. Unpublished doctoral dissertation, University of Groningen, 1979.
- Buss, A. H., & Plomin, R. *A temperament theory of personality development*. New York: Wiley, 1975.
- Buss, A. H., & Poley, W. *Individual differences: Traits and factors*. New York: Gardner, 1976.
- Buss, D. M., & Craik, K. H. The frequency concept of disposition: Dominance and prototypically dominant acts. *Journal of Personality*, 1980, 48, 379-392.
- Buss, D. M., & Craik, K. H. The act frequency approach to personality. *Psychological Review*, 1983, 90, 105-126.
- Campbell, D. T., & Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 1959, 56, 81-105.
- Carey, W. B. A simplified method for measuring infant temperament. *Journal of Pediatrics*, 1970, 77, 188-194.
- Cattell, R. B. *Description and measurement of personality*. New York: World Book, 1946.
- Cattell, R. B. *Personality: A systematic theoretical and factual study*. New York: McGraw-Hill, 1950.
- Cattell, R. B., Eber, H. W. *The Sixteen Personality Factor Questionnaire*. Champaign, IL: Institute for Personality and Ability Testing, 1964.
- De Raad, B., Mulder, E., Kloosterman, K., & Hofstee, W. K. B. Personality-descriptive verbs. *European Journal of Personality*, 1988, 2, 81-96.
- Downey, J. E. *The will temperament and its testing*. Yonkers, NY: World Book, 1923.
- Eysenck, H. J. Types of personality: A factorial study of seven hundred neurotics. *Journal of Mental Science*, 1944, 90, 851-861.
- Eysenck, H. J. *The Maudsley Personality Inventory*. (Manual prepared by R.R. Knapp). San Diego: Educational and Industrial Testing Service, 1962.
- Eysenck, H. J. *MMQ*. Göttingen: Hogrefe, 1964.
- Eysenck, H. J., & Eysenck, S. B. G. *Manual of the Eysenck Personality Inventory*. London: University of London Press, 1964.
- Fiske, D. W., & Butler, J. M. The experimental conditions for measuring individual differences. *Educational and Psychological Measurement*, 1963, 23, 249-266.
- Goldberg, L. R. Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), *Review of Personality and Social Psychology*. Vol. 2. Beverly Hills: Sage, 1981.
- Goldberg, L. R. From Ace to Zombie: Some explorations in the language of personality. In C. D. Spielberger, & J. N. Butcher (Eds.), *Advances in Personality Assessment*. Vol. 1. Hillsdale, NJ: Erlbaum, 1982.
- Hathaway, S. R., & McKinley, J. C. *Manual for the Minnesota Multiphasic Personality Inventory*. New York: Psychological Corporation, 1943.
- Hofstee, W. K. B. *Dutch traits: The first stages of the Groningen taxonomy study of personality descriptive adjectives*, University of Groningen, The Netherlands, 1976.
- Hofstee, W. K. B. *Preliminary steps in structuring the set of 1.204 Dutch personality-descriptive adjectives*. Heymanns Bulletins, HB-77-302 IN, University of Groningen, The Netherlands, 1977.
- Jackson, D. N. *Personality Research Form. Manual*. Goshen, NY: Research Psychologists Press, 1967.
- Jackson, D. N. A sequential system for personality scale development. In C. D. Spielberger (Ed.), *Current topics in clinical community psychology*. Vol. 2. New York: Academic Press, 1970.
- John, O. P., Angleitner, A., & Ostendorf, F. The lexical approach to personality: A historical review of trait taxonomic research. *European Journal of Personality*, 1988, 2, 171-203.
- Jones, M. B. Heritability as a criterion in the construction of psychological tests. *Psychological Bulletin*, 1971, 75, 92-96.
- Loevinger, J. Objective tests as instruments of psychological theory. *Psychological Reports*, 1957, 3, 635-694.
- McDougall, W. *An outline of psychology*. New York: Scribners, 1923.

- Norman, W. T. Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, 1963, 66, 574-583.
- Paulhus, D. L. Self-deception and impression management in test responses. In A. Angleitner & J. S. Wiggins (Eds.), *Personality assessment via questionnaires. Current issues in theory and measurement*. Berlin: Springer, 1986.
- Plomin, R. Heredity and temperament: A comparison of twin data for self-report questionnaires, parental ratings, and objectively assessed behavior. In L. Gedda, P. Parisi, & W. E. Nance (Eds.), *Progress in clinical and biological research: Twin research III, Part B. Intelligence, personality, and development* Vol. 69B. New York: Alan R. Liss, 1981.
- Rosch, E., & Mervis, C. B. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 1975, 7, 563-605.
- Rusalov, V. M. Object-related and communicative aspects of human temperament: A new questionnaire of the structure of temperament. *Personality and Individual Differences*, 1989, 10, 817-827.
- Strelau, J. *Temperament—personality—activity*. London: Academic Press, 1983.
- Strelau, J. The concept of temperament in personality research. *European Journal of Personality*, 1987, 1, 107-117.
- Strelau, J. *Are psychophysiological scores good candidates for diagnosing temperament/personality traits and for a demonstration of the construct validity of psychometrically measured traits?* Paper submitted for publication, 1990.
- Strelau, J., Angleitner, A., Bantelmann, J., & Ruch, W. The Strelau temperament inventory revised (STI-R): Theoretical considerations and scale development. *European Journal of Personality*, 1990, 4, 209-235.
- Thomas, A., & Chess, S. *Temperament and development*. New York: Brunner/Mazel, 1977.
- Wallace, J. An abilities conception of personality: Some implications for personality measurement. *American Psychologist*, 1966, 21, 132-138.
- Windle, M., & Lerner, R. M. *Revised Dimensions of Temperament Survey-Adult*. Unpublished manuscript, 1985.
- Windle, M., & Lerner, R. M. Reassessing the dimensions of temperamental individuality across the life span: The Revised Dimensions of Temperament Survey (DOTS-R). *Journal of Adolescent Research*, 1986, 1, 213-230.
- Windle, M. Psychometric strategies of measures of temperament: A methodological critique. *International Journal of Behavioral Development*, 1988, 11, 171-201.
- Wundt, W. *Grundzüge der physiologischen Psychologie* (5th ed.). Vol. 3. Leipzig: W. Engelmann, 1903.