



CROSS-CULTURAL STUDIES ON TEMPERAMENT: THEORETICAL CONSIDERATIONS AND EMPIRICAL STUDIES BASED ON THE PAVLOVIAN TEMPERAMENT SURVEY

JAN STRELAU^{1*} and ALOIS ANGLEITNER²

¹University of Warsaw, Poland and ²University of Bielefeld, Germany

(Received 4 August 1992; received for publication 4 June 1993)

Summary—Different approaches to the study of temperament in a cross-cultural frame of reference are discussed. The procedure applied in the construction of the Revised Strelau Temperament Inventory (STI-R), which was renamed the Pavlovian Temperament Survey (PTS), is suggested as an alternative approach. This approach starts with an exhaustive and prototypical item sample (i.e. item universe) for a proposed construct in one culture. By applying identical statistical analyses in the psychometric treatment, parallel tests for mapping the constructs in different cultures can be developed. Descriptive comparisons of the properties of the PTS-versions for different languages, nations, or cultures (Australian English, Dutch, German, Greek, Hungarian, Italian, Korean, Polish, and Russian) are given.

INTRODUCTION

Cross-cultural studies in psychology generally serve two basic goals: (a) to describe the specificity of psychological phenomena and/or behaviors, depending on the cultural conditions in which individuals or groups being compared live (the emic approach); and (b) to show the similarities or identities (universals) of given psychological phenomena and/or behaviors across different cultures (the etic approach).

Findings supporting the cultural specificity of given behaviors or processes speak in favor of the environment in determining differences under study. In turn, data that show similarities or identities of given psychological phenomena in spite of cultural differences are often used as arguments in favor of genetic, or more general, biological factors in determining given behaviors. Similarity or generalizability (Van de Vijver & Poortinga, 1982) across phenomena being studied in different cultures may, however, result from etic elements not being under control of the investigator. The lack of differences in behaviors or reactions across cultures may be due to the fact that the groups being compared share such common properties of the cultural environment that are important in determining the phenomena under study. Thus, before making any conclusions regarding the significance of biological factors in determining the behavior under study, it has to be proven whether the culturally different cohorts being compared do not share other than genetically determined variables (see Poortinga & Van de Vijver, 1987) that can explain the generalizability of the phenomenon under study.

TEMPERAMENT AS A SUBJECT OF CROSS-CULTURAL RESEARCH

Among the many phenomena being examined in cross-cultural studies, temperament should also be mentioned. The question arises what inclines researchers on temperament to use a cross-cultural perspective in their studies? Among other things, the following arguments may be mentioned here:

- (1) Most temperament researchers agree that temperament is, to a given extent, biologically determined, that is, inherited. Temperament is present since early ontogenesis, also occurs in animals, and refers mainly to the formal (not content) or stylistic characteristics of behavior

*The part of the paper written by Jan Strelau has been prepared during his one-year-visit (1991/92) as a Humboldt Research Award Winner at the University of Bielefeld, Department of Psychology.

(Strelau, 1987; Strelau & Angleitner, 1991). By showing the common elements in the understanding of temperament, we are not saying that there is full agreement on the understanding or definition of temperament. Voss and Keller (1988), making methodological remarks regarding cross-cultural studies on temperament, point to the lack of universality of the concept of temperament, to the different understanding of the separate temperament dimensions (see, also, Strelau, 1991), as well as to the lack of evidence for functional significance of temperament. It has to be stated, however, that few such psychological phenomena can be found on which psychologists would agree—intelligence being the best example here. Nevertheless, it seems to us that temperament belongs to those theoretical constructs in the understanding of which there is more agreement than disagreement, at least as regards the above-mentioned features of temperament. If these features may be ascribed to temperament, as we assume, than one may expect that temperamental traits belong to those psychological characteristics that are common (universal) for human beings independent of their specific cultural environment. At the same time, the formal characteristics comprised by temperament, which—we assume—are universal, may express themselves in different, culturally specific behaviors. It is highly probable that the universality refers not only to the particular traits but also to their configuration, that is, to the structure of temperament as argued, for example, in the domain of the three superfactors of personality by Eysenck (1991; Eysenck & Eysenck, 1985; see, also, Royce & Powell, 1983) or in studies on the Big Five factors (see Digman, 1989; John, 1990; McCrae & Costa, 1987).

- (2) In many temperament theories, it is assumed that temperamental traits serve given adaptive functions and that, depending on the interaction with the broadly understood environment, different developmental and behavioral effects occur (Eliasz, 1990; Strelau, 1983; Thomas & Chess, 1977; Windle, Iwawaki & Lerner, 1987). This idea has been broadly elaborated by Super and Harkness (1986) in their concept of the 'developmental niche', which "consists of the physical and social setting children are found in" (p. 133). The interaction between both the individual temperament traits and the developmental niche results in different forms of adaptation and different values of temperament dispositions for children's development. One may expect that the same temperamental traits, depending on the cultural specificity, may reveal themselves in different situations and behaviors, may be expressions of different—considering their adaptive value—behaviors. The concepts of goodness-of-fit (Chess & Thomas, 1991; Lerner, 1983; Talwar, Nitz, Lerner & Lerner, 1991), difficult child (Bates, 1980; Carey, 1986; Thomas & Chess, 1977), and temperament risk factor (Carey & McDevitt, 1989; Strelau, 1989) may be used as examples here.
- (3) Very often the need for cross-cultural studies results from the fact that supposedly the same diagnostic instrument is used in different cultures. The cultural adaptation of intelligence tests and personality and temperament inventories may be mentioned here. Studies conducted within this trend serve mainly toward the validation of the tests themselves. Only marginally does this type of study contribute to the main goals (1 and 2) of cross-cultural research on temperament. There also exists a reverse relationship. Studies on temperament from a cross-cultural perspective are legitimate only if equivalent temperament instruments are used.

Taking into account the cross-cultural studies on temperament conducted until now, it can be said that our knowledge does not permit far-reaching generalizations (see Kohnstamm, 1989; Voss & Keller, 1988). In the majority of studies concerned with the universality of traits and the structure of temperament, the same diagnostic measures—in different language versions—have been used. Their main purpose was to find out whether the same temperament dimensions and a similar structure of traits or factors occur across cultures. Such studies, conducted mainly by means of psychometric tests on children and adolescents at different ages, refer mostly to Thomas and Chess' theory of temperament (see, e.g. Banks, 1989; Hsu, Soong, Stigler, Hong & Liang, 1981; Kohnstamm, 1989; Kyrios, Prior, Oberklaid & Demetriou, 1989; Prior, Kyrios & Oberklaid, 1987; Skuy, Westaway & Snell, 1986; see, also, Iwawaki, Hertzog, Hooker & Lerner, 1985; Windle, Iwawaki & Lerner, 1988). In these studies, the assumption of equivalent diagnostic measures is adopted, however, often without sufficient data supporting this supposition. Kohnstamm (1989),

summarizing a review of these studies, stated, among other things, that most of the studies are based on very small samples, not representative for the populations being compared and not fulfilling standard psychometric requirements. Voss and Keller (1988) have extended this critique by adding further weak points of these studies such as, for example, the Western (American) ethnocentrism in cross-cultural studies on temperament and atheoretical selection of procedures and cohorts. Among the studies conducted on children and adolescents' temperament, Windle *et al.*'s (1987, 1988) approach seems to be closer to the standards cross-cultural studies are expected to follow (see Berry, 1969, 1989; Brislin, 1976a, 1983; Poortinga, 1989; Super & Harkness, 1986; Triandis & Brislin, 1984).

As regards adults, cross-cultural studies on the three Eysenckian superfactors of personality (temperament)—extraversion, neuroticism, and psychoticism—are among the most popular. The main goal of these studies is to show the universality of these dimensions by documenting that the same factorial structure of these traits is obtainable across cultures. Many cross-cultural studies conducted by Sybil Eysenck together with researchers from many different countries (e.g. Barrett & Eysenck, 1984; Eysenck & Eysenck, 1982; Eysenck, 1983; Eysenck, Barrett, Spielberger, Evans & Eysenck, 1986; Eysenck & Haapsalo, 1989; Hanin, Eysenck, Eysenck & Barrett, 1991) have given such positive results.

There only exists a few cross-cultural studies on temperament that aim to show that, dependent on the specific culture under study, the same temperamental traits play different adaptive functions (see de Vries & Sameroff, 1984; Prior, Garino, Sanson & Oberklaid, 1987; Super & Harkness, 1986). The study conducted by de Vries (1987) on Masai (a pastoral tribe living on the border between Kenya and Tanzania) infants is one of the most spectacular ones. De Vries (1987) has shown that under conditions of long-lasting drought causing persistent famine, so-called 'difficult children' according to the taxonomy of Thomas and Chess (1977) had a higher chance of survival compared with 'easy children'. The intensive reactions, dominance of negative emotions and fussiness of the 'difficult' infants provoked the mothers to take more care of them and, in consequence, to feed them more often as compared with the 'easy' infants. Though several critical remarks may be mentioned regarding the methodological background of this study (very small sample, doubtful adaptation of the temperament assessment instrument), the idea of studying the adaptive functions of temperament cross-culturally is worth continuing.

The longitudinal studies conducted by Thomas and Chess (1977) have also shown that the same temperamental traits in children (e.g. high activity) lead to different educational effects depending on whether the child is reared in upper-middle-class families represented by the European-American white population or in the lower-class families of the Puerto Rican population in New York.

CROSS-CULTURAL STUDIES ON ADULT TEMPERAMENT BASED ON THE PAVLOVIAN TEMPERAMENT SURVEY (PTS)

Theoretical considerations

In cross-cultural studies based on a psychometric approach, one of the most essential issues is to construct and, as a consequence, also to adapt to other cultures a test instrument that takes into account the criteria of conceptual, metrical, and scalar equivalence (Poortinga, 1989). The efficiency in constructing, for instance, such an inventory depends on whether the constructors take into account the etic-emic relationships of the phenomena when building scales and when generating items (Berry, 1989). However, not much attention has been paid to these criteria in the construction of temperament inventories.

When constructing inventories to be used for cross-cultural studies on temperament or personality, several strategies can be applied. We would like to refer to four strategies that take into account the emic-etic relationship:

- (1) *The Przeworski-Teune (1966) method for constructing questionnaires.* The essence of this procedure consists in generating two kinds of items: core and specific items. Core items are expected to grasp etic aspects, that is, they are relevant to all cultures. In turn, specific items are different for each culture under study; they represent the emic aspects of behavior. One

of the difficulties in this approach consists in finding appropriate criteria for generating core and specific items. Helpful in separating both core and specific items may be the decentering method (Werner & Campbell, 1970). It consists in taking into account the idiosyncrasies of each language version of the inventory when generating items. Prototypicality ratings by experts from the different cultures in which the inventory may be used can also be considered as an efficient way to distinguish core items from specific ones.

- (2) *The Triandis (1978) procedure in a simplified form.* The main idea of this method consists in generating scale constructs that are expected to be etic, that is, universal across cultures. In turn, items representing the separate scales are culturally specific, that is, emic. The items are generated in the language of each culture by members of the given culture cooperating with the scale constructor. Factor analysis of the emic items is the main way to prove whether the instrument constructed for the separate language (culture) versions refers to the set of etic constructs hypothesized by the author of the instrument. In spite of the emic items, cross-cultural comparisons are possible due to the etic scale concepts on which the instrument is built.
- (3) *The factor analytic procedure.* Another possible strategy is to apply factor analytic procedures to a sufficiently broad and representative item pool in different cultures for the constructs in question. If the factors that appear explain comparable amounts of variance and, furthermore, show comparable weight matrices, the high loading items in each culture can then be considered to constitute the core items forming the scale for measuring the construct in each culture (see Eysenck & Eysenck, 1985, pp. 102–111). However, not all items can be expected to constitute core items. Therefore, this approach is similar to the construction of parallel tests of a given construct in different cultures. The usefulness of the factor analytic procedure and the methods for factor comparisons for intercultural studies is best demonstrated by Eysenck and Eysenck (1982), Barrett and Eysenck (1984), and Eysenck, Barrett and Eysenck (1985). These studies are impressive, and they clearly show the universality of these dimensions. However, this procedure works best as long as the respective constructs are considered to be orthogonal to each other.
- (4) *The procedure applied in the construction of the Strelau Temperament Inventory-Revised (STI-R, Strelau, Angleitner, Bantelmann & Ruch, 1990a), which was renamed the Pavlovian Temperament Survey (PTS).**

This procedure differs from the other approaches in two ways: First, much more emphasis is given to the operationalization of a construct and the item generation process. Second, this procedure is also applicable to constructs that are related and therefore cannot be assumed to be orthogonal. This strategy incorporates basic ideas formulated by Loewinger (1957) as well as some ideas taken from Campbell and Fiske's (1959) conceptualization of convergent and divergent validation by the multitrait-multimethod matrix. The starting point of the PTS procedure consists in operationalizing the theoretical constructs on which the scales are built. The scale constructs measured by means of the PTS inventory (strength of excitation, strength of inhibition, mobility of nervous processes) fulfil all criteria typical for temperament traits; thus for reasons presented earlier, they are assumed to be etic, that is, universal across cultures.

The operationalization of the constructs consists in generating definitional components of the separate constructs. These components, which refer to the formal characteristics of behavior as the theoretical scale constructs do, are also supposed to be cross-culturally comparable. The definitional components of the scales constitute the basis for generating

*To avoid misunderstanding and confusion we decided to re-name this inventory into the Pavlovian Temperament Survey, (PTS). This name has been chosen for two reasons: (1) it makes clear at once what kind of temperamental traits this inventory measures; (2) as far as we know (see e.g. the Buros' catalogue) the abbreviation PTS is not known among temperament and personality inventories. Thus in further publications as well as in this paper when using the label Pavlovian Temperament Survey (PTS) it has to be understood that it refers to the questionnaire known until now as the Strelau Temperament Inventory-Revised (STI-R). This means also that the two basic papers (Strelau *et al.*, 1990a; Ruch *et al.*, 1991) in which the construction and psychometric characteristics of the STI-R are presented have to be read as referring to the Pavlovian Temperament Survey. For the measurement of temperamental traits according to Strelau's regulative theory of temperament (RTT, Strelau, 1983) a newly constructed questionnaire is used—Formal Characteristics of Behavior-Temperament Inventory (FCB-TI, Strelau & Zawadzki, in press).

items. Since temperamental traits reveal themselves in behaviors that may be culturally specific, a broad list of items has been generated by the authors. The item pool is considered as the universe of items, and the items fulfil the requirement to be rated as prototypical for the concepts. The items are common for all language versions and they refer to as many kinds of behavior and situations as possible in which temperamental characteristics may express themselves.

The emic approach in constructing the PTS consists in selecting from the item pool—by means of elaborated psychometric procedures—those items that, in a given culture (language version), are most representative for the etic constructs. It means that for each language (culture) version of the inventory, the number and kind of items is or may be different. The fact that the number of items of which the inventory is composed may vary requires standardization procedures for cross-cultural comparisons. Due to the fact that the theoretical constructs and their definitional components are etic, cross-cultural comparisons are possible.

From the very beginning of the construction of the PTS, this inventory was intended to be used for cross-cultural studies. The strategy applied and the steps in constructing the PTS are described in Strelau *et al.* (1990a) and Ruch, Angleitner and Strelau (1991). Thus we are limiting ourselves to a very general description required to understand the cross-cultural approach of our study.

THE CONSTRUCTION OF THE PTS

Taking as a starting point the Pavlovian (Pavlov, 1951–52) concept of the central nervous system (CNS) properties, we developed an inventory aimed at measuring three traits: strength of excitation, strength of inhibition, and mobility of the CNS. After a theoretical analysis of the constructs studied, we generated definitional components of the separate CNS properties—seven for strength of excitation (SE), five for both strength of inhibition (SI) and mobility of the CNS (MO). One component of each of the scales is given below to exemplify this procedure: SE: tendency to undertake activity (actions) in highly stimulating conditions; SI: ability to restrain from behaviors that are not expected or not desired for social reasons; MO: ability to pass easily from one activity to another. These selected components illustrate the intention of the authors to construct definitional components of the temperament scales in such a way as to be universal across cultures. The formal aspect of behavior comprised by the components occurs in all kinds of behavior, irrespective of whether the activity in which temperament characteristics are expressed is culturally bound or not. So, for example, eagerness to hunt dangerous animals (behavior often to be met in African cultures) and desire to parachute (behavior likely to occur in Western cultures) have—in spite of the extreme difference in their content—a common denominator: both refer to activity under highly stimulating conditions.

For each of the three scales operationalized by the definitional components, items have been generated by the four authors. Each of the authors has written at least five items for each of the 17 components. Altogether 377 items were written—152 for the SE scale, 113 for SI scale, and 112 for the MO scale. Taking into account the psychometric personality scale construction strategies as developed by Angleitner, John and Loehr (1986) and Jackson (1970), the following basic rules were obeyed: The items should be: (a) short and clearly understandable; (b) free from extreme levels of social desirability, (c) diverse in content so as to cover the whole universe of human conduct; (d) applicable to adults in different cultures and not biased toward particular populations; (e) logically related to the construct under consideration and, at the same time, not convergent with similar but irrelevant constructs; and (f) balanced in their keying. The four authors working as independent judges have selected among the 377 items only those for which full agreement among the four judges was reached regarding the six criteria given above. This procedure resulted in a 252-item pool.

The 252-item pool is considered as a starting point for the construction of the PTS for all language forms and cultures. At the very beginning, two original language forms of the 252 item pool were constructed by the authors—the German version and the Polish version. These two forms constituted the basis for the translation into the following languages: English, German, Spanish,

Table 1. Characteristics of the samples studied by means of the PTS

| Sample | No. of Ss | Sex | | Age (years) | | |
|------------|--------------|------|--------|-------------|-----------|-------|
| | | Male | Female | <i>M</i> | <i>SD</i> | Range |
| Australian | 501 | 183 | 318 | 25 | 10.16 | 14-75 |
| Dutch | 336 | 55 | 281 | 36 | 13.72 | 22-80 |
| German 1 | 510 | 170 | 340 | 30 | 13.40 | 14-81 |
| German 2 | 74 | 31 | 43 | — | — | 17-68 |
| Greek | 408 | 198 | 210 | 44 | 17.26 | 17-82 |
| Hungarian | 281 | 138 | 143 | 26 | 8.92 | 17-50 |
| Italian | 288 | 119 | 169 | 42 | 18.65 | 16-85 |
| Korean 1 | 480 | 216 | 264 | 30 | 9.36 | 18-54 |
| Korean 2 | 210 | 138 | 72 | 21 | 1.96 | 18-27 |
| Polish 1 | 402 | 143 | 259 | 29 | 10.93 | 15-63 |
| Polish 2 | 163 | 98 | 65 | 26 | 10.94 | 15-69 |
| Russian | 290 | 66 | 224 | 28 | 6.40 | 16-59 |

The German 1, Polish 1 and Russian PTS versions are based on a yes/no answer format. This is also valid for the next tables.

French, and Russian. Each translation was done by two independent, professional translators. Researchers interested in the construction of the PTS in their own country are given the possibility to make use, during the translation procedure, of all of the above-mentioned language forms of the 252-item pool. The comparison of different language forms is one of the most efficient ways of adapting inventories to specific cultures (see Brislin, 1976b; Drwal, 1990).* Hitherto, the PTS has been constructed for the following countries: Australia, Greece, Germany, Hungary, Italy, Poland, Russia, South Korea, and The Netherlands. The construction of the PTS in the following countries is in progress: Japan, Romania, and the U.S.A.

At the beginning, the PTS existed in two rating versions: with a 2-point rating scale (yes/no format) and with a 4-point rating scale (fully agree, moderately agree, moderately disagree, strongly disagree). Since Ss reported some difficulty in answering the yes/no format of the PTS (see Strelau *et al.*, 1990a), the 4-point rating scale was used in all cross-cultural studies except the Russian project.

CROSS-CULTURAL PTS STUDIES: RESULTS

Research regarding the construction of the PTS in different countries is beginning. Therefore, only preliminary results illustrating our approach consisting in satisfactory statistical analyses will be presented in this paper. We will limit ourselves to some data based on the construction of the PTS in Australia, Germany, Greece, Hungary, Italy, Poland, Russia, South Korea, and The Netherlands. The item analytic procedures for constructing the PTS were handled as outlined in Strelau *et al.* (1990a). A special computer program was developed. All data analyses were done at the University of Bielefeld using this program. For each definitional component, at least 2, preferably 3 or 4 items have been retained. All of them fulfil the following two criteria: (a) they must show sufficiently high item-scale correlation, and (b) they should correlate higher with their own scale compared to the correlations of the items with the other scales. Thus, not the factor analysis procedure for constructing the separate scales, but some basic ideas from the multi-trait-multimethod matrix were used in our study. The sample characteristics of the groups being compared are given in Table 1.

The number of items included in the separate scales of the 9 language versions of the PTS varies (see Table 2). They range from 17 (Italy) to 24 (Germany) for the SE scale, from 14 (South Korea) to 24 (Germany) for the SI scale, and from 17 (Italy) to 24 (Germany) for the MO scale. The total number of items for the separate versions of the PTS varies from 51 (Italy) to 72 (Germany).

Table 2 also gives information about the reliability scores of the PTS scales separately for the different language versions. The reliability coefficients as measured by Cronbach's Alpha are consistent across countries for all three scales. They vary from 0.80 to 0.88 for the SE scale, from 0.68 to 0.82 for SI and from 0.79 to 0.91 for MO, with somewhat lower values for SI.

*Angleitner and Strelau have prepared a Program for the Cross-Cultural Approach and Co-Operation in Constructing the Pavlovian Temperament Survey (PTS), which describes all steps and procedures to be undertaken by researchers willing to cooperate in constructing the PTS for their own language and country.

Table 2. Reliability coefficients (Alpha Cronbach) for the PTS scales and number of items (No.) in different country samples

| Sample | SE | | PTS scales SI | | MO | | Total No. of items |
|------------|-----|---------|------------------|---------|-----|---------|--------------------------|
| | No. | Alpha-C | No. | Alpha-C | No. | Alpha-C | |
| Australian | 21 | 0.85 | 21 | 0.77 | 21 | 0.88 | 63 |
| Dutch | 20 | 0.88 | 20 | 0.78 | 20 | 0.91 | 60 |
| German 1 | 24 | 0.84 | 24 | 0.80 | 24 | 0.88 | 72 |
| Greek | 20 | 0.86 | 20 | 0.80 | 20 | 0.88 | 60 |
| Hungarian | 21 | 0.88 | 21 | 0.82 | 21 | 0.88 | 63 |
| Italian | 17 | 0.82 | 17 | 0.73 | 17 | 0.83 | 51 |
| Korean 1 | 23 | 0.84 | 14 | 0.74 | 20 | 0.85 | 57 |
| Korean 2 | 23 | 0.80 | 14 | 0.68 | 20 | 0.79 | 57 |
| Polish 1 | 19 | 0.85 | 19 | 0.75 | 19 | 0.83 | 57 |
| Polish 2 | 19 | 0.84 | 19 | 0.75 | 19 | 0.85 | 57 |
| Russian | 21 | 0.83 | 21 | 0.77 | 21 | 0.88 | 63 |

Since all PTS versions stem from the same 252-item pool, one may expect that they share at least some items in common. If so, such items could be regarded as core items as suggested by Przeworski and Teune (1966, p. 10). Table 3 gives the number of inventories among the 9 PTS language versions being compared for which common items exist. As can be seen from the table, there are only 3 items—belonging to the MO scale only—that are shared by all 9 PTS versions. These are the following: 102, I can easily switch back and forth between very different activities; 144, it takes time for me to switch from sadness to joy; and 177, I can easily adjust to sudden changes in my job routine.

If, however, we take as the criterion of the so-called core items these ones that may be found in at least 5 inventories of the 9 PTS versions being compared, the number of common items grows to 50–17 for SE, 13 for SI, and 20 for MO. The smallest number of common items has been found within the SI scale, especially when a large number of PTS versions is taken into account (6 to 9 language versions). Curiously enough, it turned out that among the 252 items that served as the starting point for the construction of the PTS, 94 items have not been included in 1 of the 9 language versions. This finding, far from being accidental, needs a separate analysis and will not be dealt with in this paper.

The matching of items per scale in the different language versions presented in percentages is given in Table 4. The percentages of common items are calculated using the following formula for the respective scales:

$$\frac{\text{Number of common items} \times 2}{\text{Number of items of Language 1} + \text{Number of items of Language 2}} \times 100$$

The table therefore reads as follows: the German SE scale has 51.2% of items in common with the Polish version, 38.3% with the Korean version, 53.7% with the Italian version, 57.8% with the Russian version, and so on. Taking into account the highest and lowest percentage scores as a measure of the dispersion of items in common, it is possible to get an impression of the range of communality between the separate PTS forms.

Table 3. The number of inventories among the 9 PTS language versions being compared for which common items exist

| No. of PTS versions | Number of common items for PTS scales | | | Total No. of items |
|---------------------------|---------------------------------------|---------|--------|--------------------------|
| | SE | SI | MO | |
| 9 | 0 | 0 | 3 | 3 |
| 8 | 7 (7) | 0 (0) | 4 (7) | 11 (14) |
| 7 | 1 (8) | 3 (3) | 3 (10) | 7 (21) |
| 6 | 5 (13) | 3 (6) | 5 (15) | 13 (34) |
| 5 | 4 (17) | 8 (14) | 5 (20) | 17 (51) |
| 4 | 3 (20) | 8 (22) | 2 (22) | 13 (64) |
| 3 | 5 (25) | 8 (30) | 5 (27) | 18 (84) |
| 2 | 15 (40) | 11 (41) | 8 (35) | 34 (118) |

In parentheses the cumulative number of items is given. It comprises also the items being common for the larger number of PTS versions.

Table 4. Percentages of matching items of the different language versions of the PTS

| | German | | | Polish | | | Korean | | | Italian | | | Russian | | | Hungarian | | | Dutch | | | Australian | | | Greek | | | |
|------------|--------|------|------|--------|------|------|--------|------|------|---------|------|------|---------|------|------|-----------|------|------|-------|------|------|------------|------|------|-------|----|----|---|
| | SE | SI | MO | SE | SI | MO | SE | SI | MO | SE | SI | MO | SE | SI | MO | SE | SI | MO | SE | SI | MO | SE | SI | MO | SE | SI | MO | |
| German | 51.2 | 55.8 | 74.4 | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — |
| Polish | 38.3 | 15.8 | 50.0 | 38.1 | 36.4 | 51.3 | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — |
| Korean | 53.7 | 34.1 | 58.5 | 50.0 | 33.3 | 61.1 | 40.0 | 32.3 | 59.5 | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — |
| Italian | 57.8 | 35.6 | 62.2 | 50.0 | 40.0 | 70.0 | 54.5 | 45.7 | 53.7 | 47.4 | 31.6 | 63.2 | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — |
| Russian | 62.2 | 53.3 | 48.9 | 55.0 | 40.0 | 50.0 | 36.4 | 28.6 | 48.8 | 63.2 | 52.6 | 52.6 | 52.4 | 42.9 | 52.4 | — | — | — | — | — | — | — | — | — | — | — | — | — |
| Hungarian | 59.1 | 50.0 | 68.2 | 51.3 | 41.0 | 66.7 | 60.5 | 23.5 | 60.0 | 59.5 | 43.2 | 70.3 | 63.4 | 29.3 | 73.2 | 48.8 | 43.9 | 53.7 | — | — | — | — | — | — | — | — | — | — |
| Dutch | 57.8 | 48.9 | 66.7 | 40.0 | 50.0 | 50.0 | 40.9 | 22.9 | 58.5 | 52.6 | 36.8 | 57.9 | 52.4 | 38.1 | 57.1 | 42.9 | 47.6 | 47.6 | 53.7 | 43.9 | 63.4 | — | — | — | — | — | — | — |
| Australian | 50.0 | 54.5 | 59.1 | 46.2 | 30.8 | 56.4 | 37.2 | 11.8 | 70.0 | 54.1 | 27.0 | 81.1 | 68.3 | 29.3 | 68.3 | 63.4 | 43.9 | 58.5 | 55.0 | 45.0 | 80.0 | 34.1 | 39.0 | 58.5 | — | — | — | — |
| Greek | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — |

The number of items in each language form is given below the scale abbreviations.

Without further statistical tests, two results seem clear: The percentage of matching items for two language forms oscillates around 50% and is somewhat lower for the SI scale items in comparison to the SE and MO scale items.

The fact that few items are in common when several inventories are compared does not contradict our hypothesis. As mentioned before, we expected the number and kind of items included into the scales of the different language versions to be different since they are, in essence, emic.

The numbers of items differ for the different PTS versions. This means, among other things, that scalar equivalence of the PTS cannot be expected (Van de Vijver & Poortinga, 1982) unless some standardized transformation is arranged.

Since the means and the SDs of the different PTS language forms are not comparable across countries, it might be interesting to compare the relative variability of the scores obtained on the basis of the 9 inventories. The variability index ($V = 100 SD/M$) may serve as a useful measure to indicate the range of individual differences within the scales of the separate PTS versions. As can be seen from Table 5, the variability scores do not differ essentially across countries except for the Greek version of the PTS for which the variability index is higher in the domain of the SI and MO scales as compared with other countries.

For the other countries, the variability index is, in general, the smallest for the SI scale. At least two conclusions may be drawn from these data: (a) the scales of the separate language versions probably do not differ essentially in the sensitivity with which they can measure individual differences in temperamental traits; (b) the cultures being compared probably do not differ very much in the range of individual differences in these temperamental traits.

Cross-cultural equivalence of the PTS requires, among other things, that relationships between the PTS scales are similar for each language version. This is one of the requirements for metrical equivalence. Table 6 presents the relevant comparison data. In general, they seem to be rather consistent in that they show that the highest correlations, without exception, are between the SE and MO scales (from 0.30 to 0.64) and the lowest are between SI and MO (from 0.03 to 0.34). For all countries, the correlations between the SE and SI scales are between 0.11 and 0.46. To test the homogeneity of the scale correlations in the different language forms, a Chi-square procedure for testing the correlational patterns taken from Bortz (1985, pp. 265–266) was applied. This procedure yielded highly significant differences in the correlational patterns for the different scales ($SE/SI = 37.08$, $SE/MO = 69.54$, $MO/SI = 59.01$, $df = 10$, $P < 0.01$). A more detailed testing showed that the lack of homogeneity was mostly due to the Korean 1 and the Russian form. If the correlations between SE and MO of these two language forms are eliminated, the correlational pattern for the correlations between SE and MO show homogeneity ($SE/MO = 14.10$, $df = 8$, $P > 0.01$, NS). For the correlational pattern between SE and SI, the elimination of the Hungarian form yields homogeneity ($SE/SI = 20.53$, $df = 9$, $P > 0.01$, NS). Furthermore, the correlations between SI and MO result in a homogeneous pattern if the Italian, Russian, and Korean 1 data are excluded ($SI/MO = 10.95$, $df = 7$, $P > 0.01$, NS). The fact, that it is mostly the Korean 1 sample that leads to a deviation from the homogeneous pattern, suggests that this deviation may be seen

Table 5. Variability index for the PTS scales of the 9 language versions (with the 4-point rating scale only)

| Sample | PTS scales | | |
|------------|------------|-------|-------|
| | SE | SI | MO |
| Australian | 12.33 | 12.11 | 14.71 |
| Dutch | 16.85 | 11.02 | 13.80 |
| German 2 | 19.61 | 15.45 | 17.48 |
| Greek | 17.13 | 20.37 | 23.15 |
| Hungarian | 15.94 | 11.64 | 14.93 |
| Italian | 20.02 | 16.47 | 17.76 |
| Korean 1 | 13.37 | 12.46 | 12.42 |
| Korean 2 | 12.75 | 12.08 | 12.15 |
| Polish 2 | 16.27 | 13.40 | 14.53 |

Table 6. Intercorrelations between scales for the different language versions of PTS

| Sample | PTS scales | | |
|------------|------------|-------|-------|
| | SE/SI | SE/MO | SI/MO |
| Australian | 0.20 | 0.62 | 0.26 |
| Dutch | 0.35 | 0.64 | 0.25 |
| German 1 | 0.30 | 0.55 | 0.34 |
| Greek | 0.11 | 0.53 | 0.25 |
| Hungarian | 0.46 | 0.58 | 0.24 |
| Italian | 0.32 | 0.55 | 0.08 |
| Korean 1 | 0.22 | 0.30 | 0.00 |
| Korean 2 | 0.24 | 0.48 | 0.11 |
| Polish 1 | 0.24 | 0.57 | 0.32 |
| Polish 2 | 0.34 | 0.64 | 0.31 |
| Russian | 0.18 | 0.43 | 0.03 |

All correlation coefficients above 0.15 are statistically significant or highly significant.

as some specificity of the Korean 1 sample. The Korean 2 sample shows comparable results in their correlational pattern.

However, these discrepancies between the correlations being compared were unexpected. Furthermore, these differences in the correlational pattern occurred in spite of identical procedures in the scale development.

DISCUSSION

The aim of our study is to acquaint researchers in the field of temperament and personality with our alternative approach to constructing temperament scales for different nations and cultures. At this stage of research, we are not interested in comparing the different nations and cultures in their scale values. However, one could accomplish such a study by using an ipsatization strategy, for instance, by dividing the scale values of a language form through the number of items of the scale in their respective languages. Bodunov (1993) has recently introduced a more advanced strategy using confirmatory factor modeling to demonstrate satisfactory levels of convergent and discriminant validity for the scales and components of the 252-item pool of the PTS by comparing Russian and German samples. This research confirms that the original German item universe for the PTS is also applicable for Russia. It may be expected that the shorter form of the Russian PTS is also an equivalent parallel test for measuring essentially the same constructs in Russia. The somewhat contradictory findings in the correlational patterns of the short PTS-scales in the Russian sample compared to other language forms may also be caused by some sample specificity and does not rule out the possibility that the Russian test forms may be considered equivalent.

However, the differences in the correlational patterns between the scales of the different language forms could also point to possible different adaptive functions these traits may possess in different cultures.

Taking the frame of reference adopted by Poortinga (1989), we may say that the PTS shows some invariance as a measurement scale. The most important fact is the identity of concepts. Identity of the item content is not necessary. Poortinga points out that "the probability that two test versions differing in content will form invariant scales is very low" (1989, p. 744). Our approach suggests that by defining the content domain in an appropriate way one may find reliable scales for the measurement of temperament in different cultures via self-report.

REFERENCES

- Angleitner, A., John, O. P. & Loehr, F. J. (1986). It's what you ask and how you ask it: An itemmetric analysis of personality questionnaires. In Angleitner, A. & Wiggins, J. S. (Eds), *Personality assessment via questionnaires* (pp. 61–107). New York: Springer-Verlag.
- Banks, E. (1989). Temperament and individuality. A study of Malay children. *American Journal of Orthopsychiatry*, *59*, 390–397.
- Barrett, P. & Eysenck, S. (1984). The assessment of personality factors across 25 countries. *Personality and Individual Differences*, *5*, 615–632.
- Bates, J. E. (1980). The concept of difficult temperament. *Merrill-Palmer Quarterly*, *26*, 299–319.
- Berry, J. W. (1969). On cross-cultural comparability. *International Journal of Psychology*, *4*, 119–128.
- Berry, J. W. (1989). Imposed etics-emics-derived etics: The operationalization of a compelling idea. *International Journal of Psychology*, *24*, 721–735.
- Bodunov, M. V. (1993). Factor structure of the Pavlovian Temperament Survey in a Russian population: Comparison and preliminary findings. *Personality and Individual Differences*, *4*, 557–563.
- Bortz, J. (1985). *Lehrbuch der Statistik für Sozialwissenschaftler*. 2. Auflage. Berlin: Springer.
- Brislin, R. W. (1976a). Comparative research methodology: Cross-cultural studies. *International Journal of Psychology*, *11*, 215–229.
- Brislin, R. W. (Ed.). (1976b). *Translation: Applications and research*. New York: Wiley/Halsted.
- Brislin, R. W. (1983). Cross-cultural research in psychology. *Annual Review of Psychology*, *34*, 363–400.
- Campbell, J. B. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 546–553.
- Carey, W. B. (1986). The difficult child. *Pediatrics in Review*, *8*, 39–45.
- Carey, W. B. & McDevitt, S. C. (Eds) (1989). *Clinical and educational applications of temperament research*. Lisse: Swets and Zeitlinger.
- Chess, S. & Thomas, A. (1991). Temperament and the concept of goodness of fit. In Strelau, J. & Angleitner, A. (Eds), *Explorations in temperament: International perspectives on theory and measurement* (pp. 15–28). New York: Plenum Press.
- Digman, J. M. (1989). Five robust trait dimensions: Development, stability, and utility. *Journal of Personality*, *57*, 195–214.

- Drwal, R. L. (1990). Problemy kulturowej adaptacji kwestionariuszy osobowosci [Problems in the cultural adaptation of personality questionnaires]. In Ciechanowicz, A. (Ed.), *Kulturowa adaptacja testow* [Cultural adaptation of tests] (pp. 115–138). Warszawa: Polish Psychological Society and Faculty of Psychology, University of Warsaw.
- Eliasz, A. (1990). Broadening the concept of temperament: From dispositions to hypothetical construct. *European Journal of Personality*, 4, 287–302.
- Eysenck, H. J. (1991). Dimensions of personality: 16, 5 or 3?—Criteria for a taxonomic paradigm. *Personality and Individual Differences*, 12, 773–790.
- Eysenck, H. J. & Eysenck, M. W. (1985). *Personality and individual differences A natural science approach*. New York: Plenum Press.
- Eysenck, H. J. & Eysenck, S. B. G. (1982). Recent advances in the cross-cultural study of personality. In Spielberger, C. D. & Butcher, J. N. (Eds), *Advances in personality assessment* (Vol. 2, pp. 41–69). Hillsdale, NJ: Erlbaum.
- Eysenck, H. J., Barrett, P. T. & Eysenck, S. B. G. (1985). Indices of factor comparisons for homologous and non-homologous personality scales in 24 different countries. *Personality and Individual Differences*, 6, 503–504.
- Eysenck, S. B. G. (1983). One approach to cross-cultural studies of personality. *Australian Journal of Psychology*, 35, 381–391.
- Eysenck, S. B. G. & Haapsalo, J. (1989). Cross-cultural comparisons of personality: Finland and England. *Personality and Individual Differences*, 10, 121–125.
- Eysenck, S. B. G., Barrett, P., Spielberger, C., Evans, F. J. & Eysenck, H. J. (1986). Cross-cultural comparisons of personality dimensions: England and America. *Personality and Individual Differences*, 7, 209–214.
- Hanin, Y., Eysenck, S. B. G., Eysenck, H. J. & Barrett, P. (1991). A cross-cultural study of personality: Russia and England. *Personality and Individual Differences*, 12, 265–271.
- Hu, C. C., Soong W. T., Stigler, J. W., Hong, C. C. & Liang, C. C. (1981). The temperamental characteristics of Chinese babies. *Child Development*, 52, 1337–1340.
- Iwawaki, S., Hertzog, C., Hooker, K. & Lerner, R. (1985). The structure of temperament among Japanese and American young adults. *International Journal of Behavioral Development*, 8, 217–237.
- Jackson, D. N. (1970). A sequential system for personality scale development. In Spielberger, C. D. (Ed.), *Current topics in clinical and community psychology* (Vol. 2, pp. 61–96). New York: Academic Press.
- John, O. P. (1990). The “Big Five” factor taxonomy: Dimensions of personality in the natural language and in questionnaires. In Pervin, L. A. (Ed.), *Handbook of personality theory and research* (pp. 66–100). New York: Guilford Press.
- Kohnstamm, G. A. (1989). Temperament in childhood: Cross-cultural and sex differences. In Kohnstamm, G. A., Bates, J. E. & Rothbart, M. K. (Eds), *Temperament in childhood* (pp. 483–508). Chichester: Wiley.
- Kyrios, M., Prior, M., Oberklaid, F. & Demetriou, A. (1989). Cross-cultural studies of temperament: Temperament in Greek infants. *International Journal of Psychology*, 24, 585–603.
- Lerner, J. V. (1983). The role of temperament in psychosocial adaptation in early adolescence: A test of a “goodness of fit” model. *Journal of Genetic Psychology*, 143, 149–157.
- Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, 3, 635–694.
- McCrae, R. R. & Costa, P. T. Jr (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52, 81–90.
- Pavlov I. P. (1951–52). *Complete works* (2nd Edn). Moscow and Leningrad: SSSR Academy of Sciences (in Russian).
- Poortinga, Y. H. (1989). Equivalence of cross-cultural data: An overview of basic issues. *International Journal of Psychology*, 24, 737–756.
- Poortinga, Y. H. & Van de Vijver (1987). Explaining cross-cultural differences: Bias analysis and beyond. *Journal of Cross-Cultural Psychology*, 18, 259–282.
- Prior, M., Kyrios, M. & Oberklaid, F. (1987). Temperament in Australian, Chinese, and Greek infants: Some issues and directions for future research. *Journal of Cross-Cultural Psychology*, 17, 455–474.
- Prior, M., Garino, E., Sanson, A. & Oberklaid, F. (1987). Ethnic influences on “difficult” temperament and behavioural problems in infants. *Australian Journal of Psychology*, 39, 163–171.
- Przeworski, A. & Teune, H. (1966). Equivalence in cross-national research. *Public Opinion Quarterly*, 30, 33–43.
- Royce, J. R. & Powell, S. (1983). *Theory of personality and individual differences: Factors, systems, and processes*. Englewood Cliffs, NJ: Prentice-Hall.
- Ruch, W., Angleitner, A. & Strelau, J. (1991). The Strelau Temperament Inventory—Revised (STI-R): Validity studies. *European Journal of Personality*, 5, 287–308.
- Skuy, M., Westaway, M. & Snell, D. (1986). Applicability of the temperament construct to black South African children. *American Journal of Orthopsychiatry*, 56, 622–624.
- Strelau, J. (1983). *Temperament, personality, activity*. London: Academic Press.
- Strelau, J. (1987). The concept of temperament in personality research. *European Journal of Personality*, 1, 107–117.
- Strelau, J. (1989). Temperament risk factors in children and adolescents as studied in Eastern Europe. In Carey, W. B. & McDewitt, S. C. (Eds), *Clinical and educational applications of temperament research* (pp. 65–77). Lisse: Swets and Zeitlinger.
- Strelau, J. (1991). Renaissance in research on temperament: Where to? In Strelau, J. & Angleitner, A. (Eds), *Explorations in temperament: International perspectives on theory and measurement* (pp. 336–358). New York: Plenum Press.
- Strelau, J. & Angleitner, A. (1991). Temperament research: Some divergences and similarities. In Strelau, J. & Angleitner, A. (Eds), *Explorations in temperament: International perspectives on theory and measurement* (pp. 1–12). New York: Plenum Press.
- Strelau, J. & Zawadzki, B. (in press). The Formal Characteristics of Behavior—Temperament Inventory (FCB-TI): Theoretical assumptions and scale construction. *European Journal of Personality*, 7.
- Strelau, J., Angleitner, A., Bantelmann, J. & Ruch, W. (1990a). The Strelau Temperament Inventory—Revised (STI-R): Theoretical considerations and scale development. *European Journal of Personality*, 4, 209–235.
- Strelau J., Angleitner, A. & Ruch, W. (1990b). Strelau Temperament Inventory (STI): General review and studies based on German samples. In Butcher, J. N. & Spielberger, C. D. (Eds), *Advances in personality assessment* (Vol. 8, pp. 187–241). Hillsdale, NJ: Erlbaum.

- Super, C. M. & Harkness, S. (1986). In Plomin, R. & Dunn, J. (Eds), *The study of temperament: Changes, continuities and challenges* (pp. 131–149). Hillsdale, NJ: Erlbaum.
- Talwar, R., Nitz, K., Lerner, J. V. & Lerner, R. M. (1991). The functional significance of organismic individuality: The sample case of temperament. In Strelau, J. & Angleitner, A. (Eds), *Explorations in temperament: International perspectives on theory and measurement* (pp. 29–42). New York: Plenum Press.
- Thomas, A. & Chess, S. (1977). *Temperament and development*. New York: Brunner/Mazel.
- Triandis, H. C. (1978). Some universals of social behavior. *Personality and Social Psychology Bulletin*, 4, 1–16.
- Triandis, H. C. & Brislin, R. W. (1984). Cross-cultural psychology. *American Psychologist*, 39, 1006–1016.
- Van de Vijver & Poortinga, Y. H. (1982). Cross-cultural generalization and universality. *Journal of Cross-Cultural Psychology*, 13, 387–408.
- Voss, H. G. & Keller, H. (1988). Cross-cultural research on temperament: A critical evaluation. Unpublished manuscript, Technische Hochschule Darmstadt & Universitaet Osnabrueck, Germany.
- de Vries, M. W. (1987) Cry babies, culture, and catastrophe: Infant temperament among the Masai. In Scheper-Hughes, N. (Ed.) *Child survival* (pp. 165–185). Reidel.
- de Vries, M. W. & Sameroff, A. J. (1984). Culture and temperament: Influences on infant temperament in three East African societies. *American Journal of Orthopsychiatry*, 54, 83–96.
- Werner, O. & Campbell, D. T. (1970). Translating, working through interpreters, and the problem of decentering. In Naroll, R. & Cohen, R. (Eds), *A handbook of method in cultural anthropology* (pp. 398–420). New York: American Museum of Natural History.
- Windle, M., Iwawaki, S. & Lerner, R. M. (1987). Cross-cultural comparability of temperament among Japanese and American early- and late-adolescents. *Journal of Adolescent Research*, 2, 423–446.
- Windle, M., Iwawaki, S. & Lerner, R. M. (1988). Cross-cultural comparability of temperament among Japanese and American preschool children. *International Journal of Psychology*, 23, 547–567.