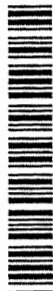


West Virginia University ILL



ILLiad TN: 568148

**Borrower:** ORE

**Call #:** BF1 .P75 - R08M18S06T07

**Lending String:**

\*WVU,WVU,BMC,FUG,FUG,ORU,IXA,AZU,IWA,G  
UA,OSU,WLU,CWR,UMC,XII

**Location:** depo

**Patron:** Mejia, Shannon Teresa

**ODYSSEY ENABLED**

**Journal Title:** Psychologische Beitrage.

**Charge**

**Maxcost:** 25.00IFM

**Volume:** 28 **Issue:** 1-2

**Month/Year:** 1986**Pages:** 255-264

**Shipping Address:**

Oregon State University  
121 Valley Library ILL  
Jefferson Way & Waldo Place  
Corvallis OR 97331-4501

**Article Author:**

**Article Title:** Nesselroade; Reliability vs. stability  
in the measurement of psychological states; An  
illustration with anxiety measures

**Fax:** (541) 737-1328

**Ariel:** 128.193.162.52

**Email:** valley.ill@oregonstate.edu

**Imprint:** Meisenheim/Glan, A. Hain.

**ILL Number:** 110203083



## Reliability vs. Stability in the measurement of psychological states: An Illustration with anxiety measures

J. R. NESSELROADE<sup>1</sup>, R. PRUCHNO<sup>2</sup>, and A. JACOBS<sup>3</sup>

### Summary, Zusammenfassung, Résumé

To illustrate the importance of distinguishing between reliability and stability concepts, the responses of 42 subjects on two state anxiety scale parallel forms, both of which were administered on four occasions of measurement, are analyzed. Separate estimates of reliability of the measuring instrument and stability of the state anxiety construct are generated by means of a confirmatory factor analysis approach. Results clearly support the usefulness of distinguishing between psychometric properties of tests and attributes of psychological processes in constructing and evaluating measuring instruments and in testing concepts.

### Realibilität und Stabilität in der Messung psychologischer *states*: Eine Illustration mit Maßen der Angst

Zur Untersuchung des Stellenwerts der Unterscheidung zwischen Reliabilitäts- und Stabilitätskonzepten wurden die Antworten von 42 Vpn auf zwei parallele Skalen der Angst als *state* untersucht. Beide Formen wurden viermal erhoben. Mit Hilfe einer konfirmatorischen Faktoranalyse wurden gesondert Schätzungen der Reliabilität des Meßinstruments und der Stabilität des Angstkonstrukts vorgenommen. Die Befunde sprechen deutlich für den Nutzen der Unterscheidung zwischen den psychometrischen Eigenschaften der Tests und den Eigenschaften der psychologischen Prozesse, wenn Meßinstrumente konstruiert und evaluiert, und wenn Konzepte untersucht werden sollen.

### Fiabilité et stabilité des mesures de «states» (états) psychologiques. Une illustration utilisant des mesures de l'angoisse

L'importance de la distinction entre les concepts de la fiabilité et la stabilité est discutée en analysant les réponses de 42 sujets à deux tests parallèles de l'angoisse

- 1 This manuscript was completed while the senior author was a guest scientist (Gastwissenschaftler) at the Max Planck Institute for Human Development and Education, Berlin, West Germany. The generous support of the MPI is acknowledged with gratitude, as is the help and advice of Alexander von Eye concerning preparation of the manuscript. Requests for reprints should be addressed to John R. Nesselroade, College of Human Development, The Pennsylvania State University, University Park, PA 16802.
- 2 Rachel Pruchno, Philadelphia Geriatric Center, Philadelphia, PA.
- 3 Alfred Jacobs, The University of Southern California, Los Angeles.

comme «state». Les sujets avaient à répondre autre fois aux deux tests. Avec l'aide d'une analyse factorielle confirmatoire, des estimations séparées de la reliabilité de l'instrument de mesure et de stabilité du construit de l'angoisse a été estimé séparément.

Il en résulte que la distinction entre les caractéristiques psychométriques des tests et les caractéristiques des concepts psychologiques est utile à la construction et l'évaluation des échelles psychométriques et à l'analyse des concepts psychologiques.

(H. et W. Lohr)

#### 心理学的状態の測定における信頼性と安定性：不安測定による実証

信頼性と安定性という2つの概念を区別することの重要性を示すために、42名の被験者の2つの状態不安尺度に対する平行テスト形式による反応が分析された。これらのテストはともに4回実施された。測定方法の信頼性と状態不安の概念の安定性が確認的因子分析によって推定された。その結果、測定方法を構成および評価し、概念をテストする際に、テストの計量心理学的特性と心理的過程の属性間を区別することの有効性が支持された。

(山下利之 T. Yamashita)

### Introduction and Purpose

Issues involving the structure and measurement of short-term change dimensions (mood, state, etc.) have fostered a close scrutiny and reevaluation of both personality and psychometric concepts over the past two decades (e.g., CATTELL, 1973; CATTELL & SCHEIER, 1961; NESSELROADE & BARTSCH, 1977). In personality measurement, for example, the distinction between states and traits (NESSELROADE, 1983) has led to the identification of several important points. Included are: (1) the need to clarify relationships between reliability and validity concepts; (2) the importance of distinguishing between properties of measuring instruments and putative characteristics of the psychological entity or process being measured; and (3) better understanding of the roles played in theoretical formulations by psychological dimensions that, instead of involving long-term prediction (stability), show lability and thus pertain more to concurrent relationships. Research on these issues holds considerable promise for improving the construction, evaluation, and use of measures of various kinds of psychological change (e.g., developmental change, function fluctuation, oscillatory behavior, etc.).

The major purpose of this article is to review some of the pertinent substantive and psychometric matters using state anxiety as a substantive vehicle and to present data that rather clearly illustrate the central rela-

tionships and concepts. Implications of the findings for personality research will be considered.

### Intra- and interindividual differences concepts

Underlying a considerable amount of personality and ability research is the goal of imposing structure on observed individual differences. Historically, much attention has been given to the conceptualization of dimensions of relatively stable attributes (traits) and to the construction of appropriate measurement devices both to provide a framework for inferences about the nature of psychological processes and to allow valid predictions of other individual differences dimensions.

But many interesting human behavior phenomena, rather than being stable, are changeable; some over relatively short time intervals. Attempts to incorporate such change dimensions into a broad individual differences framework was an important factor in leading to the conception of psychological states (see e.g., CATTELL & SCHEIER, 1961). Subsequently, work on the identification and measurement of psychological states has resulted in the recognition of a large class of change or lability concepts. Because of its pervasiveness in major theories of personality, state anxiety has received the most attention from researchers but other dimensions of intraindividual change such as depression, effort stress, hostility, aggression, regression, etc., are among those for which a variety of measurement devices has been constructed (e.g., CATTELL, 1973; CATTELL & SCHEIER, 1971; CURRAN & CATTELL, 1976; GOTTSCHALK & GLESER, 1969; LE

devices has been constructed (e.g., CATTELL, 1973; CATTELL & SCHEIER, 1971; CURRAN & CATTELL, 1976; GOTTSCHALK & GLESER, 1969; LEBO & NESSELROADE, 1978; NESSELROADE & BARTSCH, 1977; SPIELBERGER, GORSUCH, & LUSHENE, 1969). Moreover, the list is not confined to derivatives of temperament. Motivational concepts such as drive or energetic tension level fall within the general framework of states; a state counterpart of locus of control has been investigated by ROBERTS (1979) and HORN (1972) has reported on the nature of state variation in the very bastion of putative stability – human abilities.

The essential characteristic which distinguishes state dimensions from personality traits is that the differences among individuals on states, rather than being conceptualized as more or less stable or permanent, are intrinsically changeable. State scores are expected to fluctuate, even over relatively short intervals, in response to exogenous or endogenous influences. Individual differences existing at one point in time are not expected to remain stable unless they are being maintained by persistent circumstances.

Despite the fact that current *level* on state dimensions is subject to change and thus may not constitute a useful predictor of future *level*, states may have some stable individual difference aspects that are potentially valuable as long-term predictors of other individual differences. For example, the upper and lower limits within which an individual's state scores fall, periodicity, latency of state response, magnitude of response to a particular stimulus event, etc., may differ from individual to individual in a stable, predictable manner. Consistent with this line of reasoning, some researchers have elected to define *trait* anxiety in relation to characteristics of *state* anxiety patterns such as average state level (e.g., PATTERSON & BECHTEL, 1977).

### Psychometric aspects of state measurement

The potential for short-term changes intrinsic to the state dimension concept requires the application of somewhat different criteria of instrument construction and evaluation than is the case for trait structures. Of course, the basic concepts of reliability and validity remain applicable but some explicit recognition of how these concepts are to be made operational is necessary.

Validation activities may vary from testing long-term predictions to the examination of concurrent relationships and the purposes for such investigations range from essentially atheoretic checks on predictors to elaborate attempts at construct validation. Clearly, the objective of prediction within an individual differences perspective is met most efficiently when individual's scores are stable over lengthy periods of time. If a predictor is a labile dimension and individuals must be remeasured to determine their current status each time predictions are to be made, the attribute loses much of its attractiveness as a predictor, at least from the standpoint of simple approaches to prediction. Although such intra-individual change dimensions may represent something of a "nightmare" to the psychometrist, to the personality theorist, the clinician, etc., they are pervasive and important features of individuals. Devices constructed for the measurement of such changeable attributes cannot, at the same time, both suppress the manifestation of lability and lay claim to valid measurement of the concept. Rather, evidence of lability is a central aspect of construct validity for such scales (NESSELRODE & BARTSCH, 1977).

Reliability, depending on an author's purpose, is depicted as having from one (e.g. LABOVITZ & HAGEDORN, 1971) to seven (e.g. CATTELL, 1964) salient aspects which bear on the description and evaluation of psychological measures. Obviously, a premium is to be placed on high reliability of measuring instruments in so far as one is referring to how

accurately, precisely, and consensually an attribute is being measured at a given point in time. But that notion of reliability is to be distinguished from stability. Indices that are computed from multiple occasions of measurement (e.g., test-retest stability coefficients) can confound unreliability of measurement and psychological change. Without a high level of within-occasion reliability, the interpretation of test-retest stability coefficients becomes ambiguous because they may reflect measurement error, process lability, or both. The importance of the distinction between reliability and stability pertains also to more trait-like measures in as much as they exhibit some fluctuation over time. Recent developments, some of which will be considered later, provide quite sophisticated approaches to separating and estimating both components (WHEATON, MUTHÈN, ALWIN, & SUMMERS, 1977).

This brief overview leads to one of our major conclusions. Reliability is a characteristic of the measuring instrument or measurement process whereas stability is, or ought to be considered to be, a characteristic of the phenomenon or dimension being measured. Therefore, it is appropriate to judge the quality of a measuring instrument against reliability of measurement criteria but not to assess it in relation to how high test-retest correlations are unless the phenomenon being measured is somehow known to be a highly stable one. In the case of state measurement, theory dictates that under many circumstances the test-retest correlations or stability coefficients ought to be low. In such circumstances high stability coefficients would constitute *prima facie* evidence against the validity (sensitivity) of the measuring instruments (NESSELROADE & BARTSCH, 1977). We will next examine a set of data which illustrates clearly the distinction between reliability as a characteristic of the measurement procedure and stability as a characteristic of the concept being measured.

### Method and Procedure

The design of the study involved three critical features. First, the focus was on dimensions of change – in this particular case, state anxiety – which, from a theory point of view, were not expected to show high levels of stability (test-retest correlation). Second, the participants were measured repeatedly (four occasions), so that more direct inferences about the nature of stability and change could be made than are possible from cross-sectional data. Third, multiple indicators (in this case “parallel” test forms) of the state anxiety construct were used to improve both the assessment of measurement reliability and to permit its separation from stability of the anxiety construct.

Form A and Form B of a scale<sup>4</sup> initially developed by R. B. CATTELL and J. R. NESSELROADE was used to measure state anxiety. Each form consists of 20 self-report items with a six point response format (strongly agree to strongly disagree). Scale scores represent simple summations of item scores.

The participants were 42 undergraduate students enrolled in a clinically-oriented psychology class at an eastern U.S. university. Subjects had elected to participate in personal growth, training group (T-group) meetings involving discussion and self-disclosure, and agreed to provide self-reported information to monitor their moods. For the latter, students chose from a variety of options available for earning points towards their course grades. Subjects completed a survey of mood and emotion (the SMA, JACOBS, 1971) and both forms of the state anxiety scale two times in the week preceding and two times in the week following the T-group sessions for a total of four measurement occasions.

### Analysis and Results

The correlations among scales over the four occasions of measurement are presented in Table 1. With respect to traditional criteria the outcome is rather clear. Reliability estimates represented by the correlations between Forms A and B at a given occasion of measurement are quite high (+.91, +.93, +.96, +.95 for occasions 1, 2, 3, and 4, respectively).

The stability coefficients, in marked contrast to the reliability coefficients, are very small. Some are slightly negative. The largest, +.32, is between occasions one and four (Form B), but it is not statistically significant. Clearly, the relationship pattern is one of high reliability of measurement (correlation between parallel forms) with low stability of the phenomena (test-retest correlations).

To provide more precise estimates of these relationships, a confirmatory factor analysis as depicted in Figure 1 was undertaken. The coefficients estimated in the fitting procedure are presented in Figure 1. The stability coefficients vary around zero as shown. The non-zero loadings

4 Originally developed from item analysis of a pool of several dozens of items, the scale was used in a series of studies (BARTSCH & NESSELROADE, 1973; LACHMAN, 1981; NESSELROADE & CABLE, 1974). Subsequently, the majority of items have been incorporated into the anxiety subscale of the 8-State Questionnaire (CURRAN & CATTELL, 1976). To provide the reader with a flavor of the items comprising the state anxiety scales, here are some examples: "Just thinking of some hard job waiting for me makes me jittery."; "Today, I have no worries or problems."; "If one more person told me what to do today, I'm afraid I'd blow up."

Table 1:  
Within- and Between-Occasion Intercorrelations of State Anxiety Measures

Anxiety Measure	1	2	3	4	5	6	7	8
1. Form A Occ. 1	—							
2. Form B Occ. 1	.91	—						
3. Form A Occ. 2	-.18	-.09	—					
4. Form B Occ. 2	-.16	-.05	.93	—				
5. Form A Occ. 3	.01	.02	-.07	-.10	—			
6. Form B Occ. 3	.06	.04	-.05	-.09	.96	—		
7. Form A Occ. 4	.30	.34	.19	.11	.06	.04	—	
8. Form B Occ. 4	.27	.32	.14	.08	.15	.12	.95	—

in this particular case are also estimates of the correlations between variables and factors (factor structure elements) which, in classical test theory terms, makes them indices of reliability (GULLIKSEN, 1950). Thus, the squares of the loadings are estimates of *reliability coefficients*. So defined, the reliability coefficients of forms A and B are presented (in parentheses) beside the corresponding factor loadings in Figure 1.

### Discussion

In summary, a model which sharply discriminated between *reliability* as a characteristic of measures and *degree of stability* as a characteristic of the phenomenon being measured was fitted to four repeated measurements of a state anxiety battery. Reliabilities of the measures were determined to be uniformly high. In contrast, stability coefficients varied around zero. These data demonstrate that highly labile characteristics can be measured quite reliably and that in such instances one should not use test-retest correlations as estimates of measurement reliabilities.

There are several implications to be drawn from these data. First, the data point to the need for, and potential gains from developing and using more sensitive measures of attributes expected to change. Program evaluation work as well as basic research on developmental process and other kinds of change stand to benefit greatly from a more enlightened approach to the development and appraisal of measuring instruments. Moreover, attending to concerns about sensitivity to change would lead to better "probes" for monitoring the effectiveness of experimental treatments as



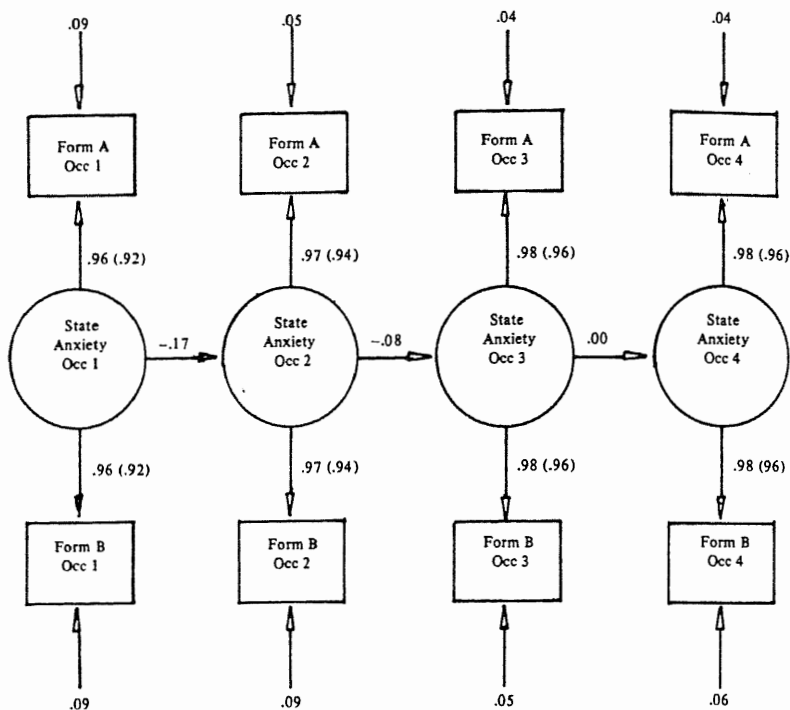


Fig. 1: Conceptual model and coefficients fitted to data ( $\chi^2 = 23.87$ ,  $df = 21$ ;  $p = .30$ ). From top to bottom within each occasion, numbers are: (1) unique variance-Form A; (2) loading (and estimated reliability)-Form A; (3) loading (and estimated reliability)-Form B; and (4) unique variance-Form B. The three numbers between pairs of occasions are estimates of stability of the latent variable, state anxiety.

operational expressions of psychological variables (e.g., evoking fear or anxiety) and, from a validity perspective, help to meet LABOUVIE'S (1980) injunction that strict psychometric criteria be applied to independent as well as dependent variables.

Second, with regard to psychometric concerns, the degree of convergence between expectation based on theory and data analytic outcome lends considerable strength to the viability of state conceptions and their measurement. High reliability with low test-retest stability stands in clear opposition to the psychometric orientation represented in many scales and scale construction procedures. This is not to downgrade the notion of stable individual differences by any means, but rather to support several

decades of conceptual and empirical work on the structuring and measurement of psychological change.

Finally, the construction and systematic use of change sensitive measures such as the state anxiety battery discussed above will further the development of personality theory. For instance, state measurement may help to clarify the nature of person x situation interactions (see e.g., MAGNUSSON & ENDLER, 1977; SPIELBERGER, 1977). If, in the study of person x situation interactions, one is looking across situations for differential behaviors that can be mapped to measured individual differences it seems important to separate measures of changeable and stable individual differences in representing the person. Two extroverts, for instance, can independently have high or low levels of state anxiety entering, or in, a given situation. Knowledge of trait-like scores is only a part of the information about individual differences that one needs to know in order to assess person x situation linkage.

Obviously, dimensions of individual differences which show little or no stability over time are not candidates for use in traditional prediction schemes. However, understanding the nature of various kinds of intraindividual changes including "steady state" variability and temporary and more permanent changes in steady state functioning, determining the antecedents for such changes, finding their temporally concurrent correlates, and exploring the extent and nature of sequential correlates offer some exciting prospects for researchers.

### References

- Baltes, P. B. & Nesselroade, J. R.: The developmental analysis of individual differences on multiple measures. In J. R. Nesselroade & H. W. Reese (Eds.), *Life-span development a psychology: Methodological issues*. New York: Academic Press, 1973.
- Cattell, R. B.: Validity and reliability: A proposed more basic set of concepts. *Journal of Educational Psychology*, 1964, 55, 1-22.
- Cattell, R. B.: *Personality and mood by questionnaire*. San Francisco: Jossey-Bass, 1973.
- Cattell, R. B., & Scheier, I. H.: *The meaning and measurement of neuroticism and anxiety*. New York: Ronald Press, 1961.
- Curran, J. P., & Cattell, R. B.: *Handbook for the 8-State Questionnaire*. Champaign, Illinois: Institute for Personality and Ability Testing, 1976.
- Gottschalk, L. A. & Gleser, G. C.: *The measurement of psychological states through the content analysis of verbal behavior*. Berkeley, CA: University of California Press, 1969.
- Gulliksen, H.: *Theory of mental tests*. New York: Wiley, 1950.

- Horn, J. L.: State, trait, and change dimensions of intelligence. *British Journal of Educational Psychology*, 1972, 42, 159–185.
- Jacobs, A.: Mood, emotion, affect. In A. Jacobs & L. Sachs (Eds.), *The psychology of private events*. New York: Academic Press, 1971.
- Labouvie, E. W.: Identity versus equivalence of psychological measures and constructs. In L. W. Poon (Ed.), *Aging in the 1980s: Psychological issues*. Washington, D.C.: American Psychological Association, 1980.
- Labovitz, S., & Hagedorn, R.: *Introduction to social research*. New York: McGraw-Hill, 1971.
- Lachman, M.: Personality and intelligence in old age: Fluid and crystallized abilities in relation to two-year change in locus of control and perceived intellectual efficacy. Unpublished Dissertation, The Pennsylvania State University, 1981.
- Lebo, M. A., & Nesselroade, J. R.: Intraindividual differences dimensions of mood change during pregnancy identified in five P-technique factor analysis. *Journal of Research in Personality*, 1978, 12, 205–224.
- Magnusson, D., & Endler, N. S. (Eds.): *Personality at the crossroads: Current issues in interactional psychology*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.
- Nesselroade, J. R.: Some implications of the trait-state distinction for the study of aging: Still labile after all these years. Paper presented at the annual meeting of the American Psychological Association, Anaheim, CA, August, 1983.
- Nesselroade, J. R., & Bartsch, T. W.: Multivariate perspectives on the construct validity of the trait-state distinction. In R. B. Cattell & R. M. Dreger (Eds.), *Handbook of modern personality theory*. Washington, D.C.: Hemisphere, 1977.
- Nesselroade, J. R., & Cable, D. G.: "Sometimes, it's okay to factor difference scores" – the separation of trait and state anxiety. *Multivariate Behavioral Research*, 1974, 9, 273–282.
- Patterson, G. R., & Bechtel, G. C.: Formulating situational environment in relation to states and traits. In R. B. Cattell & R. M. Dreger (Eds.), *Handbook of modern personality theory*. Washington, D.C.: Hemisphere Publishing Corporation, 1977.
- Roberts, M. R.: An investigation of husband-wife interaction patterns using P-technique factor analysis. Unpublished dissertation. The Pennsylvania State University, 1979.
- Sörbom, D., & Jöreskog, K. G.: *Cofamm: Confirmatory factor analysis with model modification*. Chicago, Illinois: International Educational Services, 1976.
- Spielberger, C. D.: State-trait anxiety and interactional psychology. In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads: Current issues in interactional psychology*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.
- Spielberger, C. D., Gorsuch, R. L., & Lushene, R.: *The State-trait Anxiety Inventory (STAI) test manual, form x*. Palo Alto, CA: Consulting Psychologists Press, 1969.
- Spielberger, C. D., Lushene, R. E., & McAdoo, W. G.: Theory and measurement of anxiety states. In R. B. Cattell & R. M. Dreger (Eds.), *Handbook of modern personality theory*. Washington, D.C.: Hemisphere, 1977.
- Wheaton, B., Muthén, B., Alwin, D. F., & Summers, G. F.: Assessing reliability and stability in panel models. In D. R. Heise (Ed.), *Sociological methodology*, 1977. San Francisco, CA: Jossey-Bass, 1977.