Measuring individual differences in psychological attributes:
A psychometric view of contextual effects

VESNA BUŠKO

This paper is meant to offer a psychometric perspective on the role of context in our understanding the behavioral and psychological phenomena. It is conceived as a short outline of the conceptual and methodological framework for the analyses of psychological instruments designed to measure theoretically more or less stable psychological attributes. The meaning of context within contemporary thoughts of test scores validity and validation process is briefly discussed. The contextual effects in psychological assessment are demonstrated within latent state-trait theoretical framework (Steyer et al., 1992; 1999), which follows a basic idea that measurement of any human attribute includes trait and state components, that is, depends in some systematic way on characteristics of the person, characteristics of the situation, and the interaction between person and situation.

Key words: validation process, contextual effects, latent state-trait theory, dispositional and situational sources of variance

A basis of knowledge in any scientific domain is measurement. A huge body of literature has been written thus far on the nature of measurement. There is, however, an obvious diversity in the ways that measurement has been defined in traditional quantitative science, among the behavioural sciences, or under various conceptions within philosophy of science (cf. Andrich, 1988; Brennan, 2001; McDonnald, 1999; Michell, 2005; Nunnally & Bernstein, 1994; Stemler & Birney, 2006; Zumbo, 2007). Notwithstanding the definitional and conceptual differences, a clear distinction, as stressed by Nunnally and Bernstein (1994), should be made between measurement as a process and the standards for validating measures. Whereas measurement process involves issues on quantification of attributes and/or classification of objects with regard to given attributes, measurement or test score validation involves considerations as to how well the measurement method can explain important phenomena.

Having in mind the complexity of the validation task, measurement then appears to be the major problem in psychology and related social science disciplines. At least in part, this is obviously due to the covert nature of the constructs in psychological, educational and related sciences. No matter how refined theoretical concepts and the hypothesized relationships among them were, they are testable or falsifiable to the extent that the concepts or the attributes of scientist's interest can be measured in an appropriate way. Hence, if we are to understand the concepts we study, we must be able to identify the nature and the properties of measures used to define them. This is necessarily so because when the measurement properties of the scores on the hypothesized attributes are unknown, the appropriateness of the inferences we aim to make is unknown, too.

This paper is meant to offer a psychometric perspective on the meaning and the role of contextual influences in our understanding the behavioral and psychological phenomena, either as a fundamental, theoretical inquiry or considered within applied and professional settings. The issue could probably be well discussed from the stance of each of the existing formulations of psychometric or test theory (e.g. Brennan, 2001; Embretson & Reise, 1999; McDonnald, 1999). One conceptual framework, however, seems to be particularly suited to the very topic of this volume, as it includes formal and methodological tool which enables identification of various sources of influences, including contextual or situational effects, on observed scores in any kind of measurement. The framework has been presented by Rolf Steyer and his colleagues in the early nineties and

named latent state-trait theory (LST; *see* e.g. Steyer, 2003; Steyer, Ferring & Schmitt, 1992; Steyer & Schmitt, 1990; Steyer, Schmitt & Eid, 1999). Yet, ahead of turning our attention to the main ideas of the LST theory, I wish to devote several more sentences to the concept of validity.

Validity of measurement and the role of context

There is no doubt that validity is the paramount issue in the development and evaluation of psychological measurement instruments. Following the latest version of the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 2006), it is defined as the degree to which the evidence and theory support test score interpretations derived from the intended uses of the tests. Measurement or test score validation is seen as an ongoing process wherein one provides evidence to support the appropriateness, meaningfulness, and usefulness of the specific inferences made from individual and group test scores (Kane, 2001; Kingston, 2007; Zumbo, 2007). It is clear from the above definitions that the importance of the concept, method, and processes of validation of measures used in behavioural and eventually in all other sciences cannot be overestimated since, without validation, any inferences made from a measure are potentially meaningless, inappropriate and of limited utility.

Contemporary thoughts of validity build in some conceptual novelty with regard to basic aspects of tests and testing. In view of the core subject of the present journal volume, i.e. an inquiry into contextual effects, some elements of these advancements deserve a special consideration. As underlined in the aforementioned definitions, the focus of the validation process is not the measurement tool itself but rather the inference(s) one can make from the measurement or test scores obtained from an instrument or a measurement tool. Although on the surface the differentiation between the inferences made from a measure and the measure itself doesn't need to look as substantial, it actually bears a *context* into play. Namely, it follows from this distinction that the meaning of the test scores and the validity of any interpretation derived from them is, indeed, delimited by situation, that is, by time and place where the measurement is taken and, of course, by any planned use of the scores. Similarly, it is implied by the above definitions that the task of validation process includes determining the boundaries of the inferences to be drawn from a measure or test scores. As Zumbo (2007) emphasized, "…invalidity is something that distorts the meaning of test results for some groups of examinees in some contexts for some purposes (pp. 48)."

Yet, it seems worth reminding at this point that, whatever the proposed interpretation(s) or intended use(s) of the test scores were, validity is always a matter of degree. Also, when studying the extent to which a test score interpretation is conceptually and/or empirically sound, one is dealing with potentials to generalize from test responses to the proposed conceptualizations of these responses. This process of collecting evidence in support of inferences made from the observed test scores to the theoretical constructs is the core of the validation practice (Kane, 2001; Nunnally & Bernstein, 1994; Zumbo, 2007). Alternatively, different formulations of psychometric test theory offer a rationale for drawing inferences from the observed scores to the true scores or latent variables (e.g. Andrich, 1988; Brennan, 2001; Embretson, & Reise, 1999; McDonnald, 1999; Steyer, 2003). While the formerly described evaluation of test score substantive interpretations and related inferences is a matter of validity, the latter concerns the estimates of and the inferences on the test score accuracy or the precision of measurement, and is basically a question of reliability. This distinction between the *construct* and the *true score* or the *latent variable* should therefore be clearly recognized as it has obvious implications on the use of the test scores, their interpretation, and the consequences of inferences made from test scores.

Although the estimations of portions of variance in test scores due to measurement error and latent variable, respectively, is not the ultimate goal of psychological measurement, and reliability is just one consideration in the responsible use of tests, it is the prerequisite to validity, as the quantities resulting from reliability analyses naturally limit the inferences about the meaning of test scores. It might be said, then, that the task of validity theory is to put forward a framework for evaluating inferences from the observed scores to the hypothetical constructs of interest, by use of information on the latent variables.

A psychometric account of contextual influences

Even though there is a general acknowledgment of the contextual effects in psychological assessment among both theoreticians and researchers, this source of variation in measures of intentionally stable psychological attributes has long been considered merely at a conceptual or descriptive level. If psychological assessment of personal characteristics, especially those presumed to be rather enduring, such as traits, aptitudes, or capabilities, is affected by contextual or situational factors, this fact should be kept in mind when designing validation studies on operationalizations of such constructs, including their practical utility.

Previously announced latent state-trait theoretical framework follows a basic idea that any measure of human behavior, cognition, or emotion, includes trait and state components, that is, depends in some systematic way on characteristics of the person (traits), characteristics of the situation, and the interaction between person and situation (Steyer et al., 1992; 1999). The stance resembles the perspective of modern interactionism (Endler & Parker, 1992; Magnusson, 1990) on the existence of stable interindividual differences and intraindividual changes in psychological attributes. Stable dispositional as well as systematic albeit instable situational or contextual factors together create a *psychological state* which varies across time points to the extent that the situation to which the person is exposed changes. As it will be shown in the following paragraphs,

the theory respects and explicitly takes into account the fact which is indeed already highlighted at the beginning of the Preface of this Special Issue. Akin to Reber's dictionary account, these authors underline that measurement of any human attribute does not occur in *situational vacuum*. On the contrary, consistent individual differences assessed at some time point may be due to *psychological* situations which are different for different individuals at the same time point. In other words, psychological instruments always measure individual differences in *states*, which are attributable to differences in *trait* that the instrument is intended to measure, differences in (subjective) *situation* in which individuals are, and the person-situation *interactions*.

Unlike analyzing contextual effects on the measures of traits by use of experimental manipulation of situational characteristics, LST theory uses structural equation models with latent variables that enable analyses of contribution of particular sources of variance in observed variables in non-experimental, correlational studies. Situation specific sources of variance thus become integral components of the models instead of being reduced or eliminated, as for instance suggested by Epstein (1990), by aggregation across different situations. Hence, within latent variable analyses, all the main LST theoretical concepts, that is, states, traits, situation and/or interaction effects as well as measurement error turn into explicit components of the model (Steyer & Schmitt, 1990).

The LST theory is developed as a generalization of the classical test theory (CTT; McDonnald, 1999; Nunnally & Bernstein, 1994; Steyer, 2003; Steyer et al., 1992; 1999) where each test score Y is defined as a simple linear combination of the true score t and the measurement error e. In terms of probability theory, the true score variable $t_i := E(Y_i|U)$, i=1,...,m, is the conditional expectation of the test score variable Y_i given a person or an observational unit $U=u$. Measurement error variables e_i, i=1,..., m, are simply defined by the difference $e_i := Y_i - t_i$.

The attempts at modeling latent states and traits can basically be regarded as a study of situation and/or interaction effects, similar to estimations of measurement error in the models of classical test theory (Lord & Novick, 1968; Nunnally & Bernstein, 1994). While measurement error variance can be estimated by multiple measures or indicators of the same latent variable, the amount of situation and/or interaction effects within LST models can be estimated by multiple (repeated) measures of the same latent trait, invariant in time, with an appropriate time lag between measurements.[1] Within LST models, as an extension of CTT models, the true score variable is named *latent state variable* and defined as

---

[1] In view of the main objective of this paper, just basic concepts of the LST theory are outlined here. For a thorough account of the formal definitions of the LST variables, their properties, and basic LST parameters and coefficients, see e.g. Steyer (2003) and Steyer et al. (1992; 1999).

$$t_{ik} \text{ (or } h_{ik}) := E(Y_{ik} \mid U, S_k), \qquad (1)$$

that is, the expectation of the distribution of variable Y_{ik} conditional on the person u in the situation k. Measurement error variable is defined again by the differences between observed and true scores on variable i measured at occasion k:

$$e_{ik} := Y_{ik} - E(Y_{ik} \mid U, S_k). \qquad (2)$$

Unlike latent state variable t_{ik}, which, as already stated, represents the *person* in a *situation* as regards observable Y_{ik}, *latent trait variable* ξ_{ik} contains scores describing persons regardless of situations in which they are acquired. Thus, each person is assigned one and only one ξ_{ik} score,

$$\xi_{ik} := E(Y_{ik} \mid U), \qquad (3)$$

that is, conditional expected value of the distribution of Y_{ik} for a given unit, i.e. person U=u. It can be said that ξ_{ik} represents a component of the true score variable responsible for stability of scores across different situations or measurement points. This appears to be evident as it is consistent with the concept of trait. The second component of t_{ik} is particularly interesting considering the main topic of the journal issue as it stands for situational and/or interaction effects on scores on observable variable. The component is named *occasion-specific residual* or *latent state residual* and symbolized by ζ_{ik}:

$$\zeta_{ik} := E(Y_{ik} \mid U, S_k) - E(Y_{ik} \mid U). \qquad (4)$$

In this way, the effects of situation in which the measurement takes place, including the effects attributable to interactions between person and situation, are explicitly defined by the theoretical variable ζ_{ik}. As shown by the above equation (4), these occasion-specific effects are, like measurement error variables, simply defined as a difference variable, that is, as the difference in scores between latent state variable t_{ik} and latent trait variable ξ_{ik}.

Finally, the illustration of basic concepts of the latent state-trait theory should be completed by several important parameters of the theory. As the authors stress, the definitions of the above stated latent variables imply additivity of respective portions of variance, so that (see Steyer et al., 1999, Deinzer et al., 1995):

$$Var(Y_{ik}) = Var(t_{ik}) + Var(e_{ik}) = \\ Var(\xi_{ik}) + Var(\zeta_{ik}) + Var(e_{ik}). \qquad (5)$$

The above equation makes possible to define relevant parameters of the LST theory: coefficient of *consistency* is thus defined as the proportion of variance of observed variable Y_{ik} that is attributable to the latent trait ξ_{ik}; *occasion specificity* is the proportion of variance of observed variable Y_{ik} that is due to situational and/or interaction effects; and the *reliability* coefficient can be defined as the proportion of variance of observed variable Y_{ik} explained by latent state variable t_{ik}.

Concluding remarks

This paper aimed to present a psychometric approach to the study of contextual effects in the assessment of psycho-

logical attributes. As shown in this brief account, the LST framework provides a psychometric theory background and a methodology for acquiring estimates of the reliability, trans-situational consistency, and occasion specificity coefficients pertaining to measurement instruments used in correlational, non-experimental studies. The theory offers mathematical definitions of its key concepts of traits and states and, along with well known constructs of CTT, explicitly takes into account the existence of *situation*. It sets thus a thorough methodological framework for estimation of potential role of situational factors in any sort of psychological measurement.

Empirical demonstrations of the utility of the LST theory are beyond the scope of this paper, however, there is abundant research evidence in support of its basic assumptions saying that observable psychological measures are not independent from contextual or situational attributes, and those resulting from person-situation interaction. This proves to be true even in cases when scores are obtained by instruments designed to measure relatively stable individual dispositions (e.g. Buško & Kulenović, 2004; Courvoisier, Eid, & Nussbeck, 2007; Deinzer et al., 1995; Kulenović & Buško, 2005; Le, Schmidt, & Putka, 2009; Steyer et al., 1992; 1999; Steyer & Partchev, 2001). Implications of such findings appear to be critical as within the scope of validation studies so with regard to any kind of administration of psychological measurement instruments in applied settings.

REFERENCES

AERA, APA, NCME (2006). *Standardi za pedagoško i psihološko testiranje [Standards for Educational and Psychological Testing]*. Jastrebasko: Naklada Slap.

Andrich, D. (1988). *Rasch Models for Measurement*. Newbury Park: Sage - Quantitative Applications in Social Sciences Series.

Brennan, R. L. (2001). *Generalizability theory*. New York: Springer-Verlag.

Buško, V., & Kulenović, A. (2004). Multi-Construct Latent State-Trait Model of Coping with Low-Control Situations. Paper presented at 24th SMABS Conference, Jena, Germany, 18-21.07.2004. http://www.smabs.org/.

Courvoisier, D. S., Eid, M., & Nussbeck, F. W. (2007). Mixture distribution latent state-trait analysis: Basic ideas and applications. *Psychological Methods, 12 (1),* 80-104.

Deinzer, R., Steyer, R., Eid, M., Notz, P., Schwenkmezger, P., Ostendorf, F., & Neubauer, A. (1995). Situational effects in trait assessment: the FPI, NEOFFI, and EPI questionnaires. *European Journal of Personality, 9,* 1-23.

Embretson, S. E., & Reise, S. P. (1999). *Item response theory for psychologists*. Mahwah: LEA.

Endler, N. S., & Parker, J. D. A. (1992). Interactionalism revisited: Reflections on the continuing crisis in the personality area. *European Journal of Psychology, 6,* 177-198.

Epstein, S. (1990). Comment on the effects of aggregation across and within occasions on consistency, specificity, and reliability. *Methodika, 4,* 95-100.

Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement, 38, (4),* 319-342.

Kingston, N. (2007). Future Challenges to Psychometrics: Validity, Validity, Validity (pp.1111-1122). In C. H. Rao & S. Sinharay (Eds), *Handbook of Statistics on Psychometrics* Vol 26. Oxford: Elsevier.

Kulenović, A., & Buško, V. (2005). Latentne osobine i stanja: O jednoj generalizaciji klasične teorije testova [Latent states and traits: On a generalization of the classical test theory]. *Suvremena psihologija, 8 (2),* 133-146.

Le, H., Schmidt, F. L., & Putka, D. J. (2009). The Multifaceted Nature of Measurement Artifacts and Its Implications for Estimating Construct-Level Relationships. *Organizational Research Methods, 12 (1),* 165-200.

Lord, F. C. & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Massachusetts: Addison-Wesley Publishing Company.

Magnusson, D. (1990). Personality development from an interactional perspective. In L. A. Pervin (Ed), *Handbook of personality: Theory and research* (pp. 193-222). New York: Guilford Press.

McDonald, R. P. (1999). *Test Theory: A unified treatment*. Mahwah: LEA.

Michell, J. (2005). Measurement: Overview. In B. S. Everitt & D. C. Howell (Eds), *Encyclopedia of Statistics in Behavioral Sciences* (pp. 1176-1183). West Sussex: Willey.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York: McGraw-Hill.

Steyer, R. (2003). *Wahrscheinlichkeit und Regression*. Berlin: Springer-Verlag.

Steyer, R., Ferring, D., & Schmitt, M. J. (1992). States and traits in psychological assessment. *European Journal of Psychological Assessment, 8 (2),* 79-98.

Steyer, R., & Partchev, I. (2001). Latent state-trait modeling with logistic item response models. In R. Cudeck, S. du Toit, & D. Sorbom (Eds), *Structural Equation Modeling: Present and Future* (pp. 481-520). Lincolnwood: Scientific Software International, Inc.

Steyer, R., & Schmitt, M. J. (1990). The effects of aggregation across and within occasions on consistency, specificity, and reliability. *Methodika, 4,* 58-94.

Steyer, R., Schmitt, M. J., & Eid, M. (1999). Latent state-trait theory and research in personality and individual differences. *European Journal of Personality, 13,* 389-408.

Stemler, S. E., & Birney, D. P. (2006). Measurement. In N. J. Salkind (Ed), *Encyclopedia of Measurement and Statistics*. Newbury Park: SAGE Publications.

Zumbo, B. D. (2007). Validity: Foundational Issues and Statistical Methodology. In C. H. Rao, & S. Sinharay (Eds), *Handbook of Statistics on Psychometrics,* Vol. 26 (pp. 45-80). Oxford: Elsevier.