Joint International IMEKO TC1+ TC7+ TC13 Symposium
August 31st – September 2nd, 2011, Jena, Germany
urn:nbn:de:gbv:ilm1-2011imeko:2

# THE QUANTIFICATION OF LATENT VARIABLES IN THE SOCIAL SCIENCES: REQUIREMENTS FOR SCIENTIFIC MEASUREMENT AND SHORTCOMINGS OF CURRENT PROCEDURES

*Thomas Salzberger*

Institute for Marketing Management & Institute for Statistics and Mathematics,
WU Wien, Vienna, Austria

**Abstract** − In the social sciences, latent constructs play an important role. They appear as explanatory elements in structural theories, or they are of interest as the outcome of an intervention, for example a support or a preventive programme, a therapy, or a marketing activity. These constructs are typically considered to imply a quantitative latent variable that exists independently of measurement. As a matter of routine, measures of latent variables in the social sciences are treated in the same way as natural scientists handle and utilize their measures. However, in terms of what the concept of measurement is actually about, the social sciences have veered away from the rigorous concept adhered to in the natural sciences. An arbitrary definition of measurement and a multitude of procedures which are deemed appropriate for quantification have resulted in a speculative approach to measurement. Based on a return to the standard definition of measurement and a new conceptualisation of content and construct validity, the social sciences could advance their quantitative research substantially. The Rasch model for measurement plays an important role in this process.

**Keywords**: measurement in the social sciences, construct validity, Rasch model

## 1. INTRODUCTION

In the social sciences, qualitative and quantitative research are often seen as competing paradigms. In many cases they constitute separate spheres of scientific enquiry with limited crossover both in terms of exchange of ideas and findings and regarding personal overlap. Even though mixed methods research [1, 2, 3, 4] tries to counteract this shortcoming, qualitative considerations often play a limited role in quantitative research, which is generally considered more prestigious, valuable and insightful.

The role model of the natural sciences, which have thrived on quantification, has moulded the purpose of the social sciences. Medical research, education, business research etc. are all devoted to measurement and quantitative modelling. So far, success seems to prove the social sciences right. Indeed, quantitative insight allows for a more profound understanding of reality. However, this is only true provided that the concepts investigated really exist as quantitative entities and that the applied measurement procedures actually produce valid measures of latent variables.

Suggesting a quantitative latent variable represents a hypothesis built upon a substantive theory of the construct. Specifically, the hypothesis implies an ontological claim [5]. In science, empirical evidence is required to corroborate a hypothesis. However, current practice of measurement in the social sciences typically handles the issue of whether or not a latent variable exists very generously. The predominant paradigm of measurement is still based on classical test theory (CTT) [6], also known as true score theory. CTT neither explains how measurement is accomplished, nor does it address the ontological claim of a latent variable. Rather it presupposes the existence of the construct and deals with associations of scores which are presumed to be interval-scaled measures.

In terms of what the concept of measurement is actually about, the social sciences have veered away from the rigorous concept adhered to in the natural sciences. An arbitrary definition of measurement and a multitude of procedures which are deemed appropriate for quantification have resulted in a rather speculative approach to measurement. Based on a return to the standard definition of measurement and a new conceptualisation of content and construct validity, the social sciences could advance their quantitative research substantially. The Rasch model for measurement plays an important role in this process.

In the following, a brief overview of current approaches to measurement in marketing is provided. Subsequently, the meaning of measurement is discussed and conclusions are drawn.

## 2. APPROACHES TO MEASUREMENT IN MARKETING RESEARCH

In the social sciences, the vast majority of measures of latent variables are inferred from data based on some sort of questionnaires. The data set therefore consists of coded responses of persons to items. Despite being firmly rooted in CTT, measurement in marketing research has experienced a differentiation leading to a multitude of approaches. CTT is still the most popular measurement model.

### *2.1. Classical test theory*

CTT has been made popular in marketing primarily by Churchill [7], even though empirical work had been based on CTT principles before as well. The fundamental idea of CTT is the separation of the true score and the error score, which are the two components of the observed score over a number of items. Today, the congeneric model [8] is the most widely used variant of CTT. This model applies the logic of the true score and the error component to the item level and explicitly accounts for the latent variable as the factor score. Borsboom [5] therefore classifies the congeneric model as a latent variable theory to be distinguished from the original concept of true score theory, where the latent variable sits outside the model.

The relationship of each item to the latent variable is modeled by a linear regression with the latent variable being the cause of the manifest score [9]. The association of the manifest scores and the latent variable varies in terms of strength, that is some items are more closely related to the latent variable than others. Since all items are assumed to represent the same latent variable, the item scores need to be correlated at least moderately. The matrix of inter-item correlations is therefore used to estimate the strength of the relationship between the latent and the manifest variables by means of factor analysis. Then the factor score, which is inferred from the item scores, is used as the measure of the latent variable.

The most serious deficiency of CTT as a measurement theory lies in the fact that observed item scores are treated as interval-scaled measures. Thus, CTT is actually concerned with the behaviour of measures of the same concept rather than explaining how we arrive at measures in the first place. Factor analysis allows the researcher to represent multiple replications of measures of the same concept by a single number. On the one hand, this is desirable as the factor score is a more parsimonious summary of the data and it is more precise, as assessed by reliability [10]. On the other hand, the question whether individual items really represent measures and whether the latent variable actually exists is not even remotely addressed. This disqualifies CTT as a scientific theory of measurement.

The application of CTT to item scores that are obviously not measures but the result of an interaction of respondents and items is problematic and essentially unjustified. While the span of the latent variable is theoretically infinite and limitless, the range of the manifest score on an item is of course limited the number of response options provided. This potentially leads to floor and ceiling effects, when respondents hit the boundaries of the scale. If such effects occur, the explanatory power of item intercorrelations is seriously diminished as the assumption of normally distributed scores is almost certainly violated. In practice, researchers try to avoid floor and ceiling effects by selecting items where the mean respondent score is near the scale center and the scores' distribution is close to normal [11]. However, if this strategy proves successful, all items will necessarily capture the same range of the latent variable resulting in a very narrow bandwidth [12]. Furthermore, floor or ceiling effects may occur as soon as the scale is administered to a different sample, since the distribution of item scores depends on the distribution of the respondents. Thus reliability at the item and at the scale level is compromised as well. The sample dependence entails that reliability confounds properties of the instrument (error variance) and properties of the sample (true score variance). Hence, a low reliability need not imply a bad scale, specifically when the sample is very homogeneous, while a high reliability may be a consequence of a heterogeneous sample and/or many respondents hitting the floor and the ceiling of the item score range. The meaning of any given threshold for the reliability for a scale to be acceptable, like the often cited 0.7 [10] or any other value, is therefore limited to a particular sample. It does not transcend the application of the scale at hand.

The conceptual shortcomings of CTT impact on the interpretation of alleged respondent measures, as well. Since the metric of measures is defined by the distribution of person measures, an individual's measure or the mean of a group of respondents can only be interpreted in comparison to the reference sample. However, it would be much more informative if measures could be referenced back to the measurement instrument and qualitatively interpreted in terms of the items rather than other respondents only.

In summary, CTT presumes what it is supposed to deliver: measures of latent variables. However, CTT does not only provide us with doubtful measures. It also shapes the way measurement instruments are designed in an adverse manner. Specifically, researchers are not encouraged to elaborate their theory of the latent construct in terms of what more or less of the latent variable implies. Rather CTT favours items which are perfect replications of one another.

### *2.2. Formative models*

In CTT, causality is assumed to flow from the latent variable to the indicators, which are therefore referred to as reflective indicators [9]. In theory, there are indefinitely many potential indicators and it does not matter which items are used. It has been argued that in marketing many constructs are different inasmuch as a specific set of indicators defines the latent variable [13]. Consequently, causality is reversed and the indicators are formative rather than reflective [14, 15, 16]. For example, the perceived overall quality of a hotel consists of many elements (like its location, the quality of the room, the friendliness of the staff, etc.). As these components may or may not be correlated, the application of a measurement model based on correlations is indeed inappropriate. In any case, the elimination of one component because of its being uncorrelated with other components would be inappropriate.

According to their proponents, formative indicator models are an alternative to reflective measurement models [14] suggesting that the idea of a latent variable is compatible with either approach. However, in formative models, an important attribute of a latent variable, namely unidimensionality, does not apply. In fact, formative indicators should ideally capture different aspects as they would be redundant otherwise. CTT presumes the actual existence of a latent variable but fails to provide empirical evidence. With formative models, the question becomes quite irrelevant as the alleged latent variable is defined by its indicators. From this it follows that a formative model does not qualify for measurement of a latent variable.

Formative models summarize multiple measures [5, 17, 18] but they do not constitute measurement. Similar to CTT, measurement, which takes place at the level of the individual indicators, is simply presumed. The fact that formative models merely summarize measures is also reflected by the problems of parameter estimation. Since a formative model as such is unidentified, formative indictors could only be added using equal or unequal weights that have to be specified by the researcher. Such an approach would not involve any parameter estimation at all. The parameters specifying the relationship between the formative indicators and their summary variable can only be estimated empirically if at least two dependent variables exist that are causally influenced by the summary variable [13]. Consequently, a formative model can be used when it is an antecedent to two dependent latent variables which are measured by reflective indicator models. However, then the dependent variables determine the path coefficients in the formative model [19]. What seems odd when considering the formative model a measurement model, makes perfect sense when interpreting the formative model as a structural model. A formative model aims at predicting dependent variables. The summary variable represents a more stringent structural theory as it mediates the causal relationship between the formative indicators and the dependent variables.

An alternative approach to model identification is the MIMIC model [9]. In this case, the two dependent variables are not latent but considered reflective indicators of the variable to be measured by the formative model. In other words, there are both formative and reflective indicators. Although typically discussed in the context of formative measurement models, the MIMIC model more closely resembles CTT. The latent variable is identified by the reflective indicators, while the formative indicators are antecedent variables, or covariates of the latent variable. However, the shortcomings of CTT apply in this case as well.

In summary, the view that formative models represent measurement models is misleading as it obscures fundamental differences between reflective and formative models with regard to the concept of a latent variable. Summary variables, or indices, are an instrument aimed at summarizing variables in order to simplify complex assessments. In this regard, they can be useful.

### *2.3. Single-item measurement*

If a latent variable is measured by a single item, the manifest variable is equated with the latent concept. While single item measures are now widely considered obsolete, Bergkvist and Rossiter [20] try to support the use of single-item measures. The authors compare a single-item measure with a three-item scale in the context of advertising. They demonstrate that a single-item measure has about equal predictive power than the multiple-item measure of the same construct. However, this argument has at least two major weaknesses. First, predictive power tells very little about the validity of a measure [16]. We have to provide evidence on construct validity (that is that measurement has been achieved and that the scale measures the suggested construct) first before assessing external relationships. Second, it is not surprising that a three-item measure with highly similar items has about equal predictive power as the single item. The items in the multiple-item scale are redundant and their reliability is inflated. In fact, what the authors show is that a poorly designed multiple-item measure does not perform better than a single item, a situation known as the attenuation paradox [6]. Single-item measures entail all theoretical problems of CTT-based measures. In addition, the assessment of measurement error (as it is defined in the context of CTT) is impaired.

### *2.4. The C-OAR-SE approach*

C-OAR-SE is an acronym which stands for construct definition - object representation - attribute classification - rater-entity identification - selection of item-type and answer scale - enumeration and scoring rule [21]. This approach dismisses statistical or psychometric analyses altogether and places emphasis only on content validity. The latter implies a semantic correspondence of the definition of the construct and the item wording. Interestingly, the C-OAR-SE approach does not follow a constructivist point of view. It rather claims to maintain a realist position as regards the variables measured. However, this claim is completely disconnected from any sort of empirical evidence.

### *2.5. Item response theory and Rasch modeling*

While CTT treats the data as consisting of measures, item response theory (IRT) [22, 23] regards data being composed of the outcome of interactions of respondents and items. Thus, IRT focuses on individual responses to particular items rather than aggregate statistics. IRT models specify a respondent location parameter, that is the measure we are ultimately interested in, and a set of item parameters, typically a location and a discrimination parameter. The item location parameter specifies the amount of the property the item stands for. A logistic, s-shaped function models the relationship of the respondent location and the response probability given item properties.

IRT models differ in terms of the parameterization of the items. The Birnbaum model [24] features a discrimination parameter for each item. In contrast, the Rasch model [25] requires item discrimination to be equal across items. This difference has important theoretical and philosophical consequences. Specifically, the Rasch model of measurement features unique properties with specific objectivity [26] as its defining characteristic. Thus, Rasch models constitute a separate class of models to be distinguished from general IRT models [27]. While general IRT seeks for an optimal description of the data, in Rasch measurement, the model takes precedence over the data. In case of data misfitting the Rasch model, IRT proponents resort to a more general model. By contrast, to advocates of the Rasch model, misfit of data to the Rasch model indicates serious problems rejecting the hypothesis of measurement.

### *2.6. Conclusion*

The approaches to measurement in marketing differ widely. The predominant approach of CTT presumes measurement and treats observed scores as measures. Formative models are concerned with summarizing measures. Single-item measures as well as measures based on the C-OAR-SE approach may address content validity but defy the scientific requirement of empirical evidence of construct validity. IRT models properly account for the nature of data. However, general IRT and the Rasch model hold very different views as to the precedence of the data and the model, respectively. The question arises, how is it possible that a range of vastly different approaches can all purport to yield interval scale measures of latent variables? The answer lies in the definition of measurement in the social sciences.

## 3. THE MEANING OF MEASUREMENT

### *3.1. Definition of measurement*

Traditionally, measurement is the process of determining the magnitude of a quantity relative to a unit of measurement [28, 29, 30]. However, the social sciences have adopted the definition of measurement by S.S. Stevens [31], according to which measurement is accomplished by assigning numerals to objects. Although this definition has turned measurement on its head, few social scientists seem to be concerned about the consequences. Only occasionally do some scholars acknowledge that this notion of measurement is non-committal and arbitrary [32, 33]. It entails the risk that apparently successful quantification is highly speculative and, basically, unscientific.

In the social sciences, measurement is used as an umbrella term for different sorts of "number-generating" procedures. It is not even confined to quantification but comprises classification ("nominal scale measurement") and order ("ordinal scale measurement") as well. Summaries of measures (index formation, "formative measurement") are also categorised as measurement [14]. In contrast, the scientific concept of quantification followed by the natural sciences only comprises measurement and counting. Consequently, in the natural sciences measurement is a well-defined type of quantification, whereas in the social sciences quantification is a loosely-circumscribed type of measurement. It might be argued that it is essentially only a matter of terminology. However, this would mean to seriously underestimate the consequences of the different notion of measurement. Stevens' definition paved the way for measurement which does not even need to be a type of quantification. Any consistently applied rule for assigning numerals to objects yields some sort of measurement. The researcher has to argue what sort of measurement (classification, order or quantification) has been accomplished. Specifically, quantification does not seem to imply a substantially more rigorous concept than mere order or classification. This has led to the naive belief that statistical processing of correlations or covariances gives rise to measurement.

### *3.2. Assessing validity*

Certainly researchers are aware of the requirement that their measures represent non-numerical entities. However, this is not tested empirically but only suggested (and believed). Consequences are manifold. Latent variables are said to be measured without any convincing evidence for their existence as quantitative properties. Measures of latent variables are interpreted as linear, interval-scaled magnitudes while in fact they might be non-linear and distorted. Procedures aiming at determining construct validity fail to test the data for representing the structure of quantity. In fact, approaches to test for convergent, discriminant or factorial validity implicitly presume that something has been measured. With external validity this is explicitly the case. All these analyses merely investigate whether the purported measures behave in a way that corresponds with structural theories linking various constructs. However, these approaches essentially fail to address the essence of construct validity.

One might object that the assessment of validity is not confined to construct validity (for which insufficient procedures are currently employed in the social sciences) but also comprises content validity. Some researchers, discontented with the status quo of construct validity assessment, even resort to content validity as the sole element of validity that matters. It goes without saying that proper instrument validation has to account for content validity. However, the challenge of measurement comprises a theoretical part as well as an empirical part, provided we endorse the notion of entity realism. Content validity is essentially a part of the theoretical realm, more precisely the domain of the substantial theory of the construct to be measured. Content validity is deductive in nature. It is concerned with the translation of a conceptual definition of the hypothesised latent variable into concrete items. Therefore, the assessment of content validity typically examines whether all relevant facets are represented in the item pool. However, as regards content validity, traditional scale development usually fails to consider the most obvious element of measurement, namely the variation in the magnitude of the latent variable. In any case, content validity is part of the theory and not part of the evidence. Content validity ensures that the hypothesis of a latent variable is spelled out properly and consistently. Thus, it is a required but not a sufficient condition. In contrast, construct validity provides the empirical evidence that the suggested latent variable is tenable. As such, it involves inductive reasoning. Without content validity, construct validity would be of very little, if any, use. Conversely, content validity without construct validity is merely an unproven theoretical consideration. Thus, proper measurement in the social sciences requires a strong substantial theory of the suggested latent variable, tantamount to content validity, and empirical evidence provided by construct validity assessment, which shows that the theoretical and the empirical sphere are linked [33 34]. For the sake of completeness, external validity is not considered a necessary element of the validation process. It is rather concerned with the usefulness of a latent variable as a predictor or an antecedent to other constructs.

Consequently, successful measurement in the social sciences requires content validity which takes account of variation in terms of the latent variable. Suggested items should be linked to different amounts of the property to be measured and thus allow for hypotheses in terms of their order. Construct validity has to test whether the structure of quantity is present in the data and whether the expected order of the items is mirrored in the responses. The latter establishes a link between content and construct validity. Since construct validity is concerned with a well-defined problem, its assessment has to be based on methods which adequately address this problem. The Rasch model [25, 35] for measurement specifies the requirements data have to meet in order to infer measures of a latent variable from manifest responses [37, 38]. It is important to stress that the measurement model takes precedence over the data. This is in sharp contrast to the traditional understanding in the social sciences, according to which a measurement model has to account for the data. Due to the arbitrary definition of measurement in the social sciences, a plethora of models and procedures seem to warrant measurement. This also accommodates the appreciation for methodological pluralisms. A statistical description, or summary, of the data may be useful for some purposes. However, if the analysis fails to address the requirements of measurement, quantification remains a speculation at best.

## 3. CONCLUSION

For more than half a century, the social sciences have walked their own way as to what measurement means and how it can be achieved. Measurement has become ubiquitous in the social sciences but this has happened at the expense of scientific rigour. Most measures remain speculation and very little is revealed about the structure of latent variables. Fruitless discussions about reflective versus formative indicators add little to the advancement of measurement in the social sciences. Only the realisation that measurement implies a rigorous scientific concept which defies arbitrary definition will help the social sciences to enhance the measurement of latent variables.

## REFERENCES

[1] C.A. Yauch, H.J. Steudel, “Complementary Use of Qualitative and Quantitative Cultural Assessment Methods”, Organizational Research Methods, vol. 6 (4), pp. 465-481, 2003.

[2] B.R. Johnson, A.J. Onwuegbuzie, “Mixed Methods Research: a Research Paradigm whose Time has Come”, Educational Researcher, vol. 33 (7), pp. 14-26, 2004.

[3] L. Hurmerinta-Peltomäki, N. Nummela, “Mixed Methods in International Business Research: a Value-added Perspective”, Management International Review, vol. 46 (4), pp. 439-459, 2006.

[4] Foscht, T., T. Angerer and B. Swoboda, “Mixed Methods, A Systematization of Research Designs” [“Mixed Methods, Systematisierung von Untersuchungsdesigns”] In R. Buber and H.H. Holzmüller (eds), *Qualitative Market Research, Concepts, Methods, Analyses* [*Qualitative Marktforschung, Konzepte, Methoden, Analysen*] Wiesbaden: Gabler, pp. 247-259, 2007.

[5] D. Borsboom, *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics*, Cambridge: Cambridge University Press, 2005.

[6] F.M. Lord, M.R. Novick (eds), *Statistical Theories of Mental Test Scores*, Reading, MA: Addison-Wesley, 1968.

[7] G.A. Churchill, “A Paradigm for Developing Better Measures of Marketing Constructs”, Journal of Marketing Research, vol. XVI (February), pp. 64-73, 1979.

[8] K.G. Jöreskog, ”Statistical Analyses of Sets of Congeneric Tests”, Psychometrika, vol. 36, pp. 109-133, 1971.

[9] J.R. Edwards, R.P. Bagozzi, “On the Nature and Direction of Relationships Between Constructs and Measures”, Psychological Methods, vol. 5 (2), pp. 155-174, 2000.

[10] J.C. Nunnally, *Psychometric Theory*, 2. ed, New York: McGraw-Hill, 1978.

[11] R. Likert, “Technique for the Measurement of Attitudes”, Archives of Psychology, vol. 22 (140), pp. 1-55, 1932.

[12] J. Singh, “Tackling Measurement Problems with Item Response Theory: Principles, Characteristics, and Assessment, with an Illustrative Example”, Journal of Business Research, vol. 57 (2), pp. 184-208, 2004.

[13] C.B. Jarvis, S.B. MacKenzie, P.M. Podsakoff, "A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research", Journal of Consumer Research, vol. 30 (September), pp. 199-218, 2003.

[14] A. Diamantopoulos, H.M. Winklhofer, "Index Construction with Formative Indicators: an Alternative to Scale Development", Journal of Marketing Research, vol. 38 (May), pp. 269-277, 2001.

[15] S.B. MacKenzie, "The Dangers of Poor Construct Conceptualization", Journal of the Academy of Marketing Science, vol. 31 (3), pp. 323-326, 2003.

[16] J. Rossiter, "The C-OAR-SE Procedure for Scale Development in Marketing", International Journal of Research in Marketing, vol. 19, pp. 305–335, 2002.

[17] A.J. Stenner, D.S. Burdick, M.H. Stone, "Formative and reflective models: Can a Rasch analysis tell the difference?", Rasch Measurement Transactions, vol. 22 (1), pp. 1152-3, 2008.

[18] A.J. Stenner, M.H. Stone, D.S. Burdick, "Indexing vs. Measuring", Rasch Measurement Transactions, vol. 22 (4), pp. 1176-7, 2009.

[19] R.D. Howell, E. Breivik, J.B. Wilcox, "Reconsidering Formative Measurement", Psychological Methods, vol. 12 (2), pp. 205-218, 2008.

[20] L. Bergkvist, J. Rossiter, "The predictive validity of multiple-item versus single-item measures of the same constructs", Journal of Marketing Research; vol. XLIV (May), pp. 175–184, 2007.

[21] J. Rossiter, *Measurement for the social sciences. The C-OAR-SE method and why it must replace psychometrics*, New York: Springer, 2010.

[22] R.K. Hambleton, H. Swaminathan, J.H. Rogers, *Fundamentals of Item Response Theory*, Newbury Park: Sage Publications, 1991.

[23] S.E. Embretson, S.P. Reise, *Item Response Theory for Psychologists*, Mahwah, NJ: Lawrence Erlbaum Associates, 2000.

[24] A. Birnbaum, "Some Latent Trait Models and Their Use in Inferring an Examinee's Ability". In F.M. Lord and M.R. Novick (eds), Statistical Theories of Mental Test Scores, Reading, MA: Addison-Wesley, Chapters 17-20, 1968.

[25] G. Rasch, *Probabilistic Models for Some Intelligence and Attainment Tests*, Copenhagen: Danish Institute for Educational Research, expanded edition (1980) with foreword and afterword by B.D. Wright, The University of Chicago Press, Chicago, 1960.

[26] G. Rasch, "On Specific Objectivity: an Attempt at Formalizing the Request for Generality and Validity of Scientific Statements", Danish Yearbook of Philosophy, vol. 14, pp. 58-93, 1977.

[27] D. Andrich, "Controversy and the Rasch model: A characteristic of incompatible paradigms?", Medical Care, vol. 42, pp. 7-16, 2004.

[28] J. Michell, *An Introduction to the Logic of Psychological Measurement*, Erlbaum, Hillsdale, 1990.

[29] J. Michell, "Quantitative Science and the Definition of Measurement in Psychology', British Journal of Psychology, vol. 88, pp. 355-383, 1997.

[30] J. Michell, *Measurement in Psychology – a Critical History of a Methodological Concept*, Cambridge University Press, Cambridge, 1999.

[31] S.S. Stevens, "Mathematics, Measurement, and Psychophysics", in S.S. Stevens (ed), *Handbook of Experimental Psychology*, Wiley, New York, NY, pp. 1-49, 1951.

[32] M. Balnaves, P. Caputi, "Introduction to Quantitative Research Methods, An Investigative Approach", Sage, London, 2001.

[33] T. Salzberger, *Measurement in Marketing Research - an Alternative Framework*, Edward Elgar, Cheltenham, UK, Northampton, MA, USA, 2009.

[34] M. Wilson, *Constructing Measures, An Item Response Modeling Approach*, Lawrence Erlbaum Associates, Mahwah, NJ, 2005.

[35] D. Andrich, "Rasch Models for Measurement, Series: Quantitative Applications in the Social Sciences", Sage University Paper #68, Sage Publications, Newbury Park, 1988.

[36] E.W. Wolfe, E.V. Smith Jr., "Understanding Rasch Measurement: Instrument Development Tools and Activities for Measure Validation using Rasch Models: Part I – Instrument Development Tools", Journal of Applied Measurement, vol. 8, no. 1, pp. 97-123, 2007.

[37] E.W. Wolfe, E.V. Smith Jr., "Understanding Rasch Measurement: Instrument Development Tools and Activities for Measure Validation using Rasch Models: Part II – Validation Activities", Journal of Applied Measurement, vol. 8, no. 2, pp. 204-234, 2007.

**Author**: Dr Thomas Salzberger, WU Wien (Institute for Marketing Management & Institute for Statistics and Mathematics), Augasse 2-6, 1090 Wien, Austria; T +43-1-31336 4609, F +43-1-31336 732, E Thomas.Salzberger@wu.ac.at