An Extended Rasch Analysis of the CETSCALE -
Implications for Scale Development and Data Construction

Thomas Salzberger

Unpublished Working Paper

Dr Thomas Salzberger, is Assistant Professor at the Department of Marketing, University of Economics and Business Administration, Vienna (WU-Wien), Augasse 2-6, A-1090 Wien

phone: +43-1-31336-4406, facsimile: +43-1-31336-732,

e-mail: Thomas.Salzberger@wu-wien.ac.at

An Extended Rasch Analysis of the CETSCALE - Implications for Scale Development and Data Construction

ABSTRACT

At the end of the millennium, quantitative consumer and marketing research is looking back on an era of construct operationalization that has been predominantly characterized by the underlying paradigm of classical test theory. The latter provides a technical framework of scale development and defines quality criteria that undoubtedly have contributed to the ever increasing standards of measurement in marketing research that we have been facing during the last decades. However, as a scientific paradigm guiding empirical research, the paradigm itself does not concern us when we try to solve substantive problems. This paper attempts to explicitly analyze selected parts of the measurement paradigm, briefly discussing alternative measurement theories and illustrating the consequences by an empirical example of a CETSCALE data set. The choice of the measurement paradigm is probably the most fundamental decision in the whole research process. It goes far beyond the decision of which alternative statistical tests are applied in data analysis. Once we are concerned with substantive data analysis we usually perceive data as objective or at least given reality. However, it is the measurement paradigm that decides how the data look like. The main purpose of this paper is to direct more attention to this mostly unconscious decision for it is the crucial determinant of how we "construct" data. In the next millennium, the application of alternative measurement paradigms promises to yield interesting new insights in marketing research.

INTRODUCTION

The CETSCALE (Shimp and Sharma 1987) has obtained much popularity in consumer research since its introduction as is demonstrated by the multiplicity of applications in national as well as in cross-national marketing research (e.g. Herche 1992; Netemeyer, Durvasula and Lichtenstein 1991; Durvasula, Andrews and Netemeyer 1997; Good and Huddleston 1995; Steenkamp and Baumgartner 1998). The basic idea of the scale rests on the concept of ethnocentrism developed within sociology and anthropology almost a century ago. Ethnocentrism is seen as a personal trait governing the individual's attitude and feeling towards - generally speaking - out-groups, i.e. groups the individual does not belong to. In contrast to the out-group, the in-group, of which the individual is a member, is valued more favorably and its members enjoy an a priori higher appreciation compared to outsiders. The idea of consumer ethnocentric tendencies transfers this concept to marketing and consumer research in that it focuses on the attitude towards foreign economies and their products opposed to one's own domestic economy. In line with the basic principle of ethnocentrism, people with high consumer ethnocentric tendencies depreciate foreign economies and consider them a potential threat for their own economy. Consequently, imports, and ultimately people purchasing imported goods, are perceived to be responsible for unemployment and to reduce the local society's wealth. In contrast, consumers with low ethnocentric tendencies appreciate international trade. They see imports as stimulating the economy, reducing prices and thereby contributing to their individual wealth. Domestic products compete with imported goods without enjoying any bonus from prejudice. Both the general level of a nation's consumer ethnocentric tendencies and the level within segments of consumers relevant to a company are obviously important for corporate location policy, product mix decisions and corporate communication strategy.

As a latent construct, consumer ethnocentric tendencies require their operationalization by manifest indicators. The original development of the CETSCALE in the US followed the well established paradigm of marketing scale development set up by Churchill (1979) - starting with domain specification, generation of items, data collection, purification, assessment of reliability and validity - and is based on the measurement paradigm of classical test theory (CTT). CTT is the almost totally unrivalled measurement theory employed in marketing research and has proven very fruitful, at least from a pragmatic viewpoint. However, CTT has several shortcomings that are rarely addressed explicitly in empirical research projects - with good reason, for basic methodological considerations are beyond substantive research problems and their solution. CTT as a measurement paradigm offers a technical framework of scale development by directing item selection and providing quality criteria of scales such as reliability and validity. Thus, the basics of CTT are highly relevant for the final shape of the measurement instrument. This paper neither investigates all weaknesses and strengths of CTT, nor does it aim at deciding whether CTT is really the most appropriate theory. Rather the paper examines the consequences of the underlying measurement paradigm more closely. It will be clear later that decisions between competing measurement paradigms may never be based on empirical data analyses only let alone by a logically conclusive calculus.

An empirical example extends a recent application of the Rasch model, a measurement model within latent trait theory (LTT), to the CETSCALE (Salzberger, 1999). To this end, a brief introduction to LTT, some of its models, and the main differences compared to CTT are provided.

LATENT TRAIT THEORY

Latent trait theory is a comprehensive family of measurement models which traces back to personalities like Lord (1980), Birnbaum (1968), Rasch (1960/1980), and many others. While sometimes seen as a rather homogeneous group of models being in contrast to CTT, in fact, some LTT models differ from each other at least as much as LTT and CTT do. Without going into details, LTT splits up into two main groups of models. The one group goes back to normal ogive models (Lord 1952, 1968) and comprises Birnbaum’s two- and three-parameter logistic models for dichotomous items (Birnbaum 1968) (which overcome the computational burden of parameter estimation in the normal ogive model) and generalizations for polytomous items like the graded response model (Samejima 1969). The other group consists of a family of so-called Rasch models sharing important features with the basic one-parameter logistic model by Rasch (1960/1980) frequently referred to as *the* Rasch model (see figure 1).

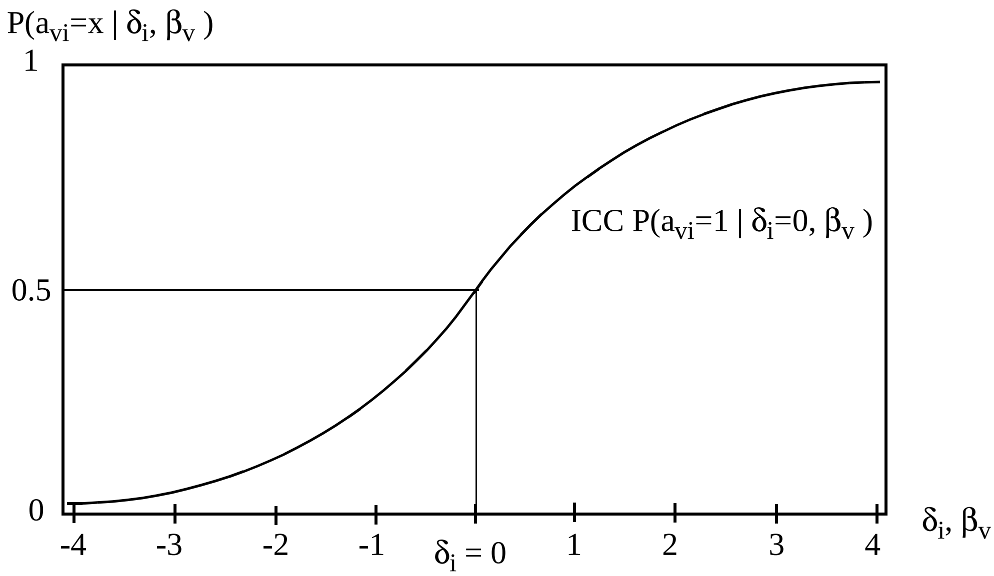
FIGURE 1
Parametrization of the Rasch Model (Rasch 1960/80, p.187).

$$P(a_{vi} = 1) = \frac{e^{\beta_v - \delta_i}}{1 + e^{\beta_v - \delta_i}} \qquad P(a_{vi} = 0) = \frac{1}{1 + e^{\beta_v - \delta_i}}$$

β_v................ person location parameter
δ_i .... item location parameter
a_{vi} .... answer of person v to item i
(0 = disagree, 1 = agree)

Since the second group is defined by "Rasch models", we might call the first "non-Rasch models" (from a Rasch perspective often referred to as "Item Response Theory"). What the two groups have in common is the concept of a latent dimension depicting the respondents' property intended to measure. Respondents are scaled onto this dimension in terms of their attitude, satisfaction, propensity to buy or whatsoever. Of course, the same applies to CTT which scales respondents to the latent dimension according to their true score. However, LTT scales not only respondents onto the scale but also items. This principle is the one which concerns us most in this paper. In LTT, the parameter characterizing the item's location on the latent dimension plays a central role. This parameter expresses the amount of the property the item stands for. The statement "*all imports should be totally banned*" stands for more consumer ethnocentric tendencies than the statement "*only those products that are unavailable in the U.S. should be imported*". This item parameter is sometimes referred to as item affectivity, however, commonly it is called item difficulty as a technical term even in attitudinal measurement. To avoid any misconception, *item location* shall be used as a neutral term in this paper. The LTT based measurement model defines the probability that a given respondent agrees with a given item characterized - in the Rasch model: only - by its location. Each item may be represented graphically by a curve, the item characteristic curve (ICC), depicting the probability of agreement depending on the respondents' location (see figure 2). Although, CTT and congeneric measurement models which may be subsumed under CTT, also provide parameters that may be termed item difficulty or affectivity, the simple proportion of people agreeing to an item, or the more sophisticated item intercept in factor analysis, these parameters suffer from their sample dependence. In contrast, the item location within the Rasch model is sample independent provided the data fit the model. Before focusing on the question what follows from the different conception of item location in LTT and CTT in terms of item generation and selection some basic LTT models are discussed.

FIGURE 2
Item Characteristic Curve ICC (For an Item With $\delta_i = 0$)



Formally, the Rasch model for dichotomous items is the most parsimonious model and it may be considered a special case of the models introduced by Birnbaum (1968). Technically, the only distinction lies in the fact that under the Rasch model the only item parameter refers to the item's location, while Birnbaum brings in two additional item parameters - an item discrimination parameter and a parameter providing a lower asymptote of the probability of agreeing or getting the item right, respectively. Consequently, by constraining all items' discrimination parameters (and setting the lower asymptote to zero) the Birnbaum model becomes the Rasch model. As Gustafsson (1980) has pointed out, varying item discrimination is tantamount to a violation of the unidimensionality assumption underlying the Rasch model. Nevertheless, from the viewpoint of non-Rasch analysis and also from conventional CTT based analysis, equal item discrimination seems to be a fairly unrealistic idea not very wise to follow. If data is regarded as being given objectively and models are to explain data best then Birnbaum models seem to be more appropriate for they feature more parameters to account for data properties. However, the Rasch model follows a different rationale. It starts with the model rather than with the data. The basic principle the Rasch model rests on is the principle of objectivity following the rationale of physical measurement. The respondents' location must not depend on specific items answered and, vice versa, the item's location must not depend on specific respondents. Rasch (1960/1980) called this principle *specific objectivity* and deduced the model that follows necessarily (see also Fischer 1995). Only under the Rasch model is the unweighted raw score for respondents and items a sufficient statistic, i.e. the specific response patterns do not provide additional information. Consequently, maximum likelihood estimation of the parameters may be conditioned on these scores and any assumptions concerning the distribution of the respondents are no longer necessary (see, e.g., Molenaar 1995 for parameter estimation techniques). Thus, equal item discrimination is not just for simplicity or parsimoniousness. With unequal item discrimination the sufficient statistics of respondents are weighted by the item discrimination parameter and conditional maximum likelihood estimation is no longer possible. The ICCs intersect and the order of items in terms of the probability of agreeing changes with the respondents' location, also referred to as Lord's paradox - a fact that actually seems to be even more problematic than equal item discrimination as such.

Consequently, the Rasch model and its disciples and the Birnbaum-like models and their supporters are separated by more than just a parameter or two. In fact, the two groups pursue different viewpoints of how data and models are rletaed to each other. For Rasch disciples the model comes first and the data have to fit the model, otherwise the data do not constitute measurement at all. In contrast, for Birnbaum supporters data come first and, in general, it is the models that are to blame whenever they do not fit the data. No wonder that the two groups hardly ever concur and regularly come to different conclusions. While a Rasch analysis discards an item which shows over- or underdiscrimination for whatever reason, a Birnbaum analysis would account for this phenomenon as long as item fit is satisfactory (i.e. for a given range of respondent locations, the difference between the probability of agreement due to the model and the actual proportion of respondents agreeing is small). Thus, it depends on the measurement model which items are retained in the measurement instrument and, because of that, how the data will look like.

Currently, LTT applications are very rare in marketing research (Singh 1996, Singh et al. 1990, Balasubramian and Kamakura 1989). So it seems to be higly relevant to investigate how CTT and LTT differ in terms of item selection. CTT models the relationship between the manifest variable (i.e. the observed item score) and the latent variable (i.e. the true score) by a regression. As mentioned before, with the intercept of this regression analysis, CTT provides a parameter that somewhat parallels the item location parameter in LTT. On the other hand, LTT aims at covering a wide range of items in terms of their location parameters in order to constitute a dimension of measurement rather than a point of measurement. By that, the measurement instrument is effective over a wide range and measurement errors are small. In contrast, the intercept parameters hardly ever play a role in item selection under CTT. For good reason because the larger the difference of the intercept terms of two items gets the lower gets the correlation of the two items due to floor and ceiling effects - and the lower gets reliability. Consequently, the classical approach of scale development leads to the selection of items more or less equal in terms of item location (see also Singh 1996). Table 1 illustrates the effects of discrete item scores rather than continuous numbers (playing a minor role) and bounding scores to a limited number of answer categories based on randomly generated variables (simulated data set).

TABLE 1
Effects of Bounding Manifest Variables on the Item Intercorrelations

Variable type	Variables correlated	Correlation coefficient	Scale reliability (Cronbach's alpha)
continuous (as generated)	var1 - var2	.490	.77
	var1 - var3	.518	
	var2 - var3	.573	
discrete (rounded to the nearest integer)	var1 - var2	.454	.75
	var1 - var3	.503	
	var2 - var3	.558	
bounded (to categories 1 to 5)	var1 - var2	.383	.58
	var1 - var3	.268	
	var2 - var3	.382	
Generation of variables	true score variable = random normal (0,1) var1 = 2 + true score variable + random error normal (0,1) var2 = 3.5 + true score variable + random error normal (0,1) var3 = 5 + true score variable + random error normal (0,1)		

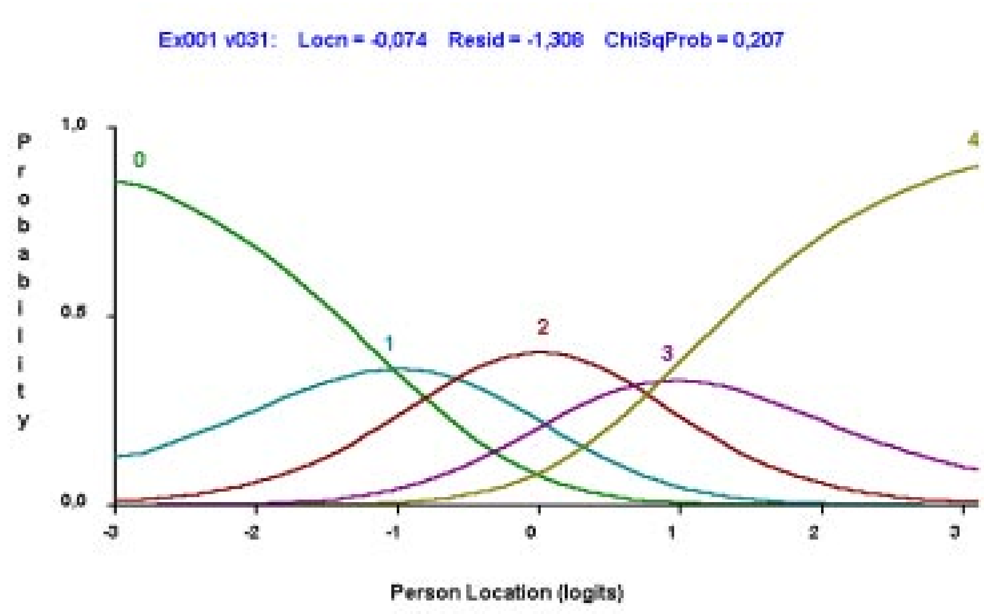
When generating potential items in scale development, we anticipate the necessity of items with roughly the same location. We try to cover as many facets as possible of a construct that we want to measure but we refrain from formulating items that obviously differ in terms of the amount of the property they stand for. The result, i.e. high reliability, makes all the effort worthwhile. Compared to physical measurement of people's height by different yardsticks, however, this strategy would mean that we develop yardsticks of all possible sorts, but all of more or less the same size.

EMPIRICAL EXAMPLE

Turning back to the CETSCALE, an empirical example illustrates the application of the Rasch approach though for items generated in line with CTT. Fortunately, the data set, collected in Austria, comprises not only the 17 items of the final CETSCALE, but also 83 items more belonging to the pool of items during initial development of the CETSCALE. Since CTT and Rasch analysis follow a different rationale of item selection, it is not only possible but very likely that the Rasch analysis ends up with different items in the scale.

As the CETSCALE items are multicategorical with seven verbalized categories ranging from fully agree to fully disagree, the dichotomous Rasch model may not be applied. However, the Rasch model may be generalized for polytomous items in a straightforward way without losing its key property, i.e. specific objectivity. The two most important models are the rating scale model (Andrich 1978) and the partial credit model (Masters 1982, Andrich 1988b). Both models may be derived by applying the dichotomous Rasch model repeatedly to adjacent categories of polytomous items. This will be clear if we regard dichotomous items: the item location may be seen as a threshold between disagreement and agreement as the most likely answer. Likewise a threshold between fully disagree and partly disagree may be modeled, another between partly disagree and somewhat disagree, and so on. Consequently k answer categories call for k-1 threshold parameters. In the case of dichotomous items, the principle of local stochastic independence, which is closely related to unidimensionality (see, e.g., Gustafsson 1980), requires any covariation of two items to be solely due to the underlying dimension and consequently to vanish for fixed respondents' location. Therefore, the probability of agreeing to several items is the product of the probabilities for each item. The same is true for thresholds within a polytomous item. Responding with somewhat disagree means passing the threshold between fully and partly disagree and the threshold between partly and somewhat disagree. This principle leads to the Rasch model for polytomous items. The probability of choosing a category may be depicted graphically by item category characteristic curves (CCC) (see figure 3).

FIGURE 3
Category Characteristic Curves for a CETSCALE Item
(After Collapsing Two Sets of Adjacent Categories)



In contrast to the graded response model (Samejima 1969) as a generalization of the dichotomous Birnbaum model, polytomous Rasch models estimate the threshold parameters independently from each other. As a consequence, the empirical threshold parameters may reflect or may not reflect the order that is hypothesized when setting up a polytomous answer scale (cf. Andrich 1995a, 1995b). If the threshold between fully disagree and partly disagree is larger than the one between partly and somewhat disagree then these thresholds are reversed and it is “easier“ to pass the second threshold than the first. Such reversed thresholds indicate that the scale does not really work as intended and has no ordinal properties. In this case, adjacent categories should be collapsed, i.e. the scoring function assigns the same numbers to adjacent categories. However, further data have to be collected in order to cross-validate the new scale format.

FIGURE 4
Extended Logistic Model (ELM), General Polytomous Rasch Model
(cf. Andrich, 1988b, p.366)

$$P(a_{vi} = 0 \mid \beta_v, \tau_{ij}, j = 1 \ldots m) = \frac{e^{0 + 0 \cdot (\beta_v - \delta_i)}}{\Upsilon} = \frac{1}{\Upsilon}$$

$$P(a_{vi} = x \mid \beta_v, \tau_{ij}, j = 1 \ldots m, 0 < x \le m) = \frac{e^{\left(\sum_{j=1}^{x} -\tau_{ij}\right) + x \cdot (\beta_v - \delta_i)}}{\Upsilon}$$

$$\Upsilon = 1 + \sum_{k=1}^{m} e^{\left(\sum_{j=1}^{k} -\tau_{ij}\right) + k \cdot (\beta_v - \delta_i)}$$

β_v................person v location parameter
δ_i ....item i location parameter
τ_{ij} ....threshold j of item j parameter
m.................maximum score, number of categories - 1
a_{vi}................answer of person v to item i (item score)

The partial credit model (see figure 4) is the most general model since it does not put any constraints on the thresholds whereas the rating scale model requires the distances between thresholds to be the same across all items in a scale. In attitudinal measurement, we usually assume that frequently used Likert scales with uniformly verbalized answer scales provide scales that are the same for all items, i.e. the difference between, e.g., somewhat agree and partly agree should be the same for all items. It should be noted that within CTT models we even require the scales to have interval scale properties, a fairly unrealistic though very pragmatic assumption. Consequently, we prefer the rating scale model for the analysis of the CETSCALE.

Polytomous Rasch models still rest on the prerequisite of unidimensionality. Although all basic LTT models (i.e. the Rasch model, the Birnbaum model, the rating scale and partial credit model, the graded response model, etc.) require unidimensional data, it is not correct to assume that LTT models necessarily are unidimensional models. Rasch (1961) already pointed out that the person parameter may be generalized to a vector, i.e. more than one dimension determines the response behavior. Today, software packages are capable of estimating multidimensional

Rasch models by conditional maximum likelihood procedures which fully separate item and person parameters (MULTIRA by Carstensen and Rost 1997). However, the concept of the CETSCALE assumes a unidimensional construct that's why it would make little sense to apply multidimensional models unless a further development of the scale is intended.

The CETSCALE Data Set

Data has been collected in Austria (n=974 listwise nonmissing respondents, self administered interviews) based on a translated version of the whole set of 100 items that remained in the item pool after a judgmental panel screening of originally 180 items generated to develop the CETSCALE (Shimp and Sharma 1987, Sinkovics 1999). The items' seven-point scale provides categories labelled as follows: fully disagree, partly disagree, somewhat disagree, neither disagree nor agree, somewhat agree, partly agree, and fully agree.

Rasch Based Analysis

The analysis started with a conventional factor analysis (principal axis factoring) in order to ensure unidimensionality. As a cutoff criterion a factor loading of .3 has been chosen which is rather small compared to CTT standards. The reason is that the correlation of the item and the factor may be reduced due to bounding effects already mentioned especially if the item is extraordinarily easy or hard to endorse. The remaining 65 items have been Rasch analyzed by applying the rating scale model, however, the algorithm did not converge. Therefore the data have been reanalyzed (using WINMIRA32 by von Davier 1999) in the first place by applying the partial credit model which allows for specific scales for each item. This step represents a rough screening of the items in terms of their basic suitability. Items showing significant misfit (assessed by the item q-index provided by WINMIRA32) do not qualify for being analyzed using the rating scale model. After several steps of discarding misfitting items and recalibrating the remaining ones, 25 items were retained. Subsequently, these items have been analyzed using the rating scale model implemented in RUMM 2.7 (Sheridan, Andrich and Luo 1997). First, a scale of ordered categories has been established. The original seven-point Likert scale turned out to lack fully ordinal properties, i.e. several thresholds were reversed. Consequently, adjacent categories had to be collapsed. After collapsing the adjacent categories of *partly agree / somewhat agree* and *neither disagree nor agree* / *somewhat disagree* and recalibrating the item parameters, the remaining five thresholds were properly ordered. Thus, subsequent analyses were based on a five point rating scale.

On each step of parameter estimation the worst significantly misfitting item (alpha =.001) in terms of a chi-square test of fit provided by RUMM 2.7, which compares model predicted probability and actual response behavior, has been deleted. Ultimately, a scale containing ten items fitting the model was derived. Six of these items already proved to fit the model when having analyzed the 17 final CETSCALE items reported by Salzberger (1999). The established rating scale is very similar, i.e. the threshold distances are almost identical. The same applies to the item location parameters as the mean of the thresholds of each item (see table 3).

However, the striking outcome of the current analysis is the fact that widening the base of the analysis from 17 to 100 items resulted in a mere increase of four additional items fitting the model. At first glance even more surprising is the small increase in the range of item locations from 0.942 to 1.188. While ten items might in principle suffice for most applications, the real problem still lies in the small range of item locations increasing the measurement error for respondents who do not fall into this small area and limiting the suitability of the scale to specific populations. (The measurement error for a specific person depends on the item information which reaches a maximum when person and item location coincide.) It should be noted that from the viewpoint of CTT, this fact does not represent a severe problem at all. In fact, the whole item pool has proved to be designed for CTT based analyses. Consequently, following the LTT approach of measurement, whether being based on a Rasch or non-Rasch model, means more than (re-)analyze a data set which creation has been guided by a different measurement paradigm, i.e. CTT. The measurement theory adhered to has a significant impact on the items generated and, eventually, on the data collected. In other words, data are, at least in part, constructed by the measurement paradigm chosen.

TABLE 2
Threshold Parameters

Analysis	number of items (rating scale categories)	Thresholds between adjacent categories (as deviations form the overall item location)					
			adjacent categories collapsed		adjacent categories collapsed		
		fully disagree / partly disagree	partly disagree / somewhat disagree	somewhat disagree / neither disagree nor agree	neither disagree nor agree / somewhat agree	somewhat agree / partly agree	partly agree / fully agree
current analysis of 100 items	10 (5)	-0.882	-0.592		0.680		0.793
analysis of 17 CETSCALE items only (Salzberger, 1999)	6 (5)	-1.021	-0.572		0.697		0.896

TABLE 3
Item Location Parameters

item [code] (original formulation in the US)	item location (mean of all thresholds)	
	current analysis n=974	Salzberger (1999) n=1105
We should buy American first - it´s the American way. [v016]	-0.433	not included
It is morally unwise to purchase foreign makes of merchandise. [v019]	0.530	not included
I favor the purchase of American-made products as a matter of principle and patriotism. [v027]	-0.331	not included
Purchasing foreign made products is un-American. [v028]	0.755	0.711
American people should always buy American-made products instead of imports. [v031]	-0.006	-0.074

item [code] (original formulation in the US)	item location (mean of all thresholds)	
	current analysis n=974	Salzberger (1999) n=1105
Only those products that are unavailable in the U.S. should be imported. [v048]	-0.173	not included
We should purchase products manufactured in America instead of letting other countries get rich of us. [v079]	0.023	-0.049
It is always best to purchase American products. [v081]	-0.149	-0.231
We should buy from foreign countries only those products that we cannot obtain within our own country. [v090]	-0.104	-0.167
Americans should not buy foreign cars, because this hurts American businesses and causes unemployment. [v093]	-0.113	-0.190
range	1.188	0.942

IMPLICATIONS AND CONCLUSIONS

Classical test theory has guided consumer and marketing research for decades as a paradigm of measurement. According to Kuhn (1996) any research within a paradigm is termed normal science. Within normal science, data seem to represent objective reality which models have to account for. However, once the paradigm itself is called in question, it immediately becomes clear that it is the early stages of scale development (i.e. domain specification and item generation) which are affected fundamentally by the measurement paradigm. While CTT calls for highly correlated items to maximize reliability and best describe populations, LTT approaches suggest batteries of items with widely varying difficulty parameters in order to establish a dimension of a construct intended to measure. Within LTT, Rasch models and non-Rasch models differ significantly in that the first feature specific objectivity while the latter fail to provide person and item parameter separation. From the viewpoint of (CTT based) normal science, it may be argued that LTT models have to prove that they account for data properties better than CTT does. Otherwise, it would make little sense to replace a pragmatic, convenient theory by - at first glance - a more complicated, unfamiliar one. If data were given objectively and if models were to fit data rather than vice versa, this approach might be feasible. However, it is the measurement theory that determines how the data will look like and it depends on the researcher's decision whether s/he wants to blame the model or the data in case of data model misfit. Any research carried out as normal science may never provide evidence about which paradigm is the better one for "the choice is not and cannot be determined merely by evaluative procedures characteristic of normal science" (Kuhn, 1996, p.94). Nevertheless paradigms do change sometimes and "[w]hat were ducks in the scientist's world before the revolution are rabbits afterwards" (Kuhn, 1996, p.111). Reflecting the explicit and implicit assumptions we usually build on, will lead us to different viewpoints. The application of LTT models to existing marketing scales is a good starting point for further dissemination. However, in the long run LTT should guide us from the beginning of scale development. Then, LTT will reveal its full potential to lift measurement in consumer and marketing research to a higher level and provide a better foundation for managerial decision making.

REFERENCES

Andrich, David. 1978. "A Rating Formulation for Ordered Response Categories." *Psychometrika* 43 (4):561-573.

----. 1988a. *Rasch Models for Measurement*, Sage University Paper Series on Quantitative Applications in the Social Sciences 68. Beverly Hills: Sage.

----. 1988b. "A General Form of Rasch's Extended Logistic Model for Partial Credit Scoring." *Applied Measurement in Education* 1 (4): 363-378.

----. 1995a. "Models for Measurement, Precision and the Non-Dichotomization of Graded Responses." *Psychometrika* 60 (1): 7-26.

----. 1995b. "Further Remarks on the Non-Dichotomization of Graded Responses." *Psychometrika* 60 (1): 37-46.

Balasubramian, Siva. K. and Wagner A. Kamakura. 1989. "Measuring Consumer Attitudes Toward the Marketplace With Tailored Interviews." *Journal of Marketing Research* 26 (3):311-326.

Birnbaum, Allan. 1968. "Some Latent Trait Models and Their Use in Inferring an Examinee's Ability." In: *Statistical Theories of Mental Test Scores*, Chapters 17-20, Eds. Frederic Lord and Melvin R. Novick. Reading (Mass.):Addison-Wesley.

Carstensen C.H. and Jürgen Rost. 1997. *MULTIRA*. Kiel: Institute for Science Education at the University of Kiel (IPN).

Churchill, Gilbert A. 1979. “A Paradigm for Developing Better Measures of Marketing Constructs.“ *Journal of Marketing Research* 26: 64-73.

Durvasula, Srinivas, Craig J. Andrews and Richard G. Netemeyer. 1997. “A Cross-Cultural Comparison of Consumer Ethnocentrism in the United States and Russia.“ *Journal of International Consumer Marketing* 9 (4): 73-93.

Fischer, Gerhard H. 1995. “Derivations of the Rasch Model.“ In: *Rasch Models, Foundations Recent Developments, and Applications*. Eds. Gerhard H. Fischer and Ivo W. Molenaar. New York: Springer, 15-38.

Good, Linda and Patricia Huddleston. 1995. “Ethnocentrism of Polish and Russian Consumers: Are Feelings and Intentions related?“ *International Marketing Review* 12 (5): 35-48.

Gustafsson, Jan-Eric. 1980. “Testing and Obtaining Fit of Data to the Rasch Model.“ *British Journal of Mathematical and Statistical Psychology* 32: 205-233.

Herche, Joel. 1992. “A Note on the Predictive Validity of the CETSCALE.“ *Journal of the Academy of Marketing Science* 20(3): 261-264.

Kuhn, Thomas. 1996. *The Structure of Scientific Revolutions* Third Edition, Chicago: The University of Chicago Press.

Lord, Frederic M. 1952. *A Theory of Test Scores. Psychometric Monograph* 7. Psychometric Corporation.

----. 1980. *Applications of Item Response Theory to Practical Testing Problems.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

---- and Melvin R. Novick. 1968. *Statistical Theories of Mental Test Scores.* Reading (Mass): Addison-Wesley.

Masters, Geofferey N. 1982. “A Rasch Model for Partial Credit Scoring.“ *Psychometrika* 47 (2): 149-174.

Molenaar, Ivo W. 1995. “Estimation of Item Parameters.“ In: *Rasch Models, Foundations Recent Developments, and Applications*. Eds. Gerhard H. Fischer and Ivo W. Molenaar. New York: Springer, 39-51.

Netemeyer, Richard G., Srinivas Durvasula and Donald R. Lichtenstein. 1991. “A Cross-National Assessment of the Reliability and Validity of the CETSCALE.“ *Journal of Marketing Research* 28 (3): 320-327.

Rasch, Georg. 1960/1980. *Probabilistic Models for Some Intelligence and Attainment Tests*, Chicago: MESA Press. Reprint of the original publication in 1960 by the Danish Institute for Educational Research.

Rasch, Georg. 1961. “On General Laws and the Meaning of Measurement in Psychology.“ *Proceedings of the Fourth Berkely Symposium on Mathematical Statistics and Probability.* Held at the Statistical Laboratory University of California, June 20 - July 30, 1960, Volume IV.. Ed. Jerzy Neyman. Berkely:University of California Press, 321-333.

Salzberger, Thomas. 1999. “How the Rasch Model May Shift Our Perspective of Measurement in Marketing Research.“ Paper presented at the 1999 Australia and New Zealand Marketing Academy Conference (ANZMAC). Sydney.

Samejima, Fumiko. 1969. "Estimation of Latent Ability Using a Response Pattern of Graded Responses." *Psychometric Monograph* 17, Iowa City (IA): Psychometric Society.

Sheridan, Barry, David Andrich and Guanzhong Luo. 1997. *User's Guide to RUMM Rasch Unidimensional Measurement Models*. Perth: RUMM Laboratory.

Shimp, Terence A. and Subhash Sharma. 1987. "Consumer Ethnocentrism: Construction and Validation of the CETSCALE." *Journal of Marketing Research* 24 (3): 280-289.

Singh, Jagdip. 1996. "A Latent Trait Theory Approach to Measurement Issues in Marketing Research: Principles, Relevance and Application" In *Proceedings of the EMAC Annual Conference, Budapest University of Economic Sciences,* Vol. 1. Eds. József Berács, András Bauer and Judith Simon. 1045-1064.

---- , Roy D. Howell and Gary K. Rhoads. 1990. "Adaptive Designs for Likert-Type Data: An Approach for Implementing Marketing Surveys." *Journal of Marketing Research* 27 (3), 304-321.

Sinkovics, Rudolf R. 1999. *Ethnozentrismus und Konsumentenverhalten*. Wiesbaden: Deutscher Universitätsverlag.

Steenkamp, Jan-Benedict E.M. and Hans Baumgartner. 1998. "Assessing Measurement Invariance in Cross-National Consumer Research." *Journal of Consumer Research* 25: 78-90.

von Davier, Matthias. 1999. *WINMIRA 32 User Manual, A Program System for Analyses with the Rasch Model, with the Latent Class Analysis and with the Mixed Rasch Models*. Kiel: Institute for Science Education at the University of Kiel (IPN).