Lecture Notes in
Statistics 157

Anne Boomsma
Marijtje A.J. van Duijn
Tom A.B. Snijders (Editors)

Essays on Item Response Theory

Springer-Verlag

Anne Boomsma, Marijtje A.J. van Duijn, Tom A.B. Snijders
Department of Statistics and Measurement Theory
University of Groningen
Grote Kruisstraat 2/1
9712 TS Groningen
The Netherlands

Preface

Over the past three decades, since the publishing of Lord and Novick's *Statistical Theories of Mental Test Scores* in 1968 and Fischer's *Einführung in die Theorie psychologischer Tests* in 1974, item response theory (IRT) has developed rapidly. This is demonstrated in the *Handbook of Modern Item Response Theory* (Van der Linden & Hambleton, 1997) with chapters on a wide range of topics in IRT. The study of individual responses to behavioral stimuli has clearly evolved into a major discipline of psychometric theory.

The Dutch statistician and psychometrician Ivo Molenaar has played an important role in the growth of IRT, not only in the Netherlands but worldwide. His contributions to item response modeling cover a broad area: item and person fit in both parametric and nonparametric models, for example, and polytomous nonparametric item response models including the development of the MSP program. With Gerhard Fischer he edited a book on Rasch models that was published in 1995, and in cooperation with Klaas Sijtsma he is now preparing a textbook on nonparametric item response theory. In the Netherlands in particular, he has encouraged researchers and doctoral students to advance into new areas of IRT. To honor such achievements we dedicate this volume on item response modeling to Ivo W. Molenaar.

Ivo's general attitude towards psychometrics is perhaps best characterized by the definition he gave at the European Meeting of the Psychometric Society in Lüneburg in 1999, when introducing his successor as president of that society: "Psychometrics is mathematical statistics in the service of substantive psychology." This statement reflects Ivo Molenaar's scientific efforts: after the completion of a Ph.D. in mathematical statistics and obtaining a chair in statistical analysis and measurement theory at the University of Groningen, The Netherlands, in 1971, he showed an increasing professional interest in psychometrics. In that sense Ivo's career resembles that of Georg Rasch (1901–1980) described by Andersen and Olsen in Chapter 1 of this volume. In his inaugural speech as a full professor (Molenaar, 1972), he explicitly referred to the work of Georg Rasch and Robert Mokken, exponents of parametric and nonparametric IRT, respectively. From then on he gradually used his skills in mathematics and statistics in the service of item response theory. For Ivo it was a challenge to study and further develop both classes of models with their intriguing statistical and practical problems. And in the end he became an aficionado of modern item response theory, one of his greatest scientific endeavors.

This book deals with several aspects of item response modeling. We arranged the contributions in two parts: parametric and nonparametric IRT. Within each part, the chapters are roughly ordered as (1) historic accounts and overviews, (2) new models, (3) new methods, and (4) applications and miscellaneous topics. It should be noted that some chapters, such as those of Junker and of Mellenbergh, cover both parametric and nonparametric topics.

This book honors Ivo Molenaar on the occasion of his retirement from the Department of Statistics and Measurement Theory at the University of Groningen as of September 2000. In that small department, which he guided, extended, and defended for almost 30 years, many students and colleagues were fortunate to benefit from his outstanding didactic qualities in teaching and consultation. His work is perhaps best characterized by his practical approach to the use of statistics in social science applications. The invited contributors to this volume are researchers who have cooperated more or less closely with Ivo in the area of IRT, including former doctoral students. During the past 30 years Ivo Molenaar has been a memorable actor on the psychometric stage.

This is a book that should be of special interest to psychometricians, empirical researchers, and graduate students. It gives an impression of the present state of the art of IRT models and methods. It may be seen as a complementary successor to the *Handbook of Modern Item Response Theory* (Van der Linden & Hambleton, 1997), and we hope its contents will set a direction for future developments in IRT.

We thank the reviewers for their expertise and more than useful comments on the chapters of this book. Finally, we appreciate the patience of the authors while trying to cope with our frequently demanding requests for major and minor revisions.

Groningen, September 2000

Anne Boomsma
Marijtje A.J. van Duijn
Tom A.B. Snijders

References

Fischer, G. (1974). *Einführung in die Theorie psychologischer Tests: Grundlagen und Anwendungen.* Bern: Huber.

Fischer, G.H., & Molenaar, I.W. (Eds.). (1995). *Rasch models: Foundations, recent developments, and applications.* New York: Springer-Verlag.

Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Molenaar, W. (1972). *Dit is een uitdaging* [This is a challenge]. Oratie [Inaugural lecture], Rijksuniversiteit Groningen.

Van der Linden, W.J., & Hambleton, R.K. (Eds.). (1997). *Handbook of modern item response theory.* New York: Springer-Verlag.

Contents

II Nonparametric Item Response Theory

1

The Life of Georg Rasch as a Mathematician and as a Statistician

Erling B. Andersen and Lina Wøhlk Olsen[1]

ABSTRACT In this chapter an account of the life of Georg Rasch as a mathematician and as a statistician is given, with emphasis on the years 1952 to 1960 when the Rasch model was developed and justified by requirements of proper measurement. But Rasch's life before 1952 is also described, especially how he, forced by the employment situation in Denmark, had to turn to statistics after having graduated and defended his doctoral dissertation in pure mathematics. Finally, it is discussed at some length how Rasch, especially after 1960, formulated his requirements of proper measurement into the formalized concept of specific objectivity.

[1]Københavns Universitet, Økonomisk Institut, Studiestræde 6, 1455 København K., Danmark; e-mail: *erling.b.andersen@econ.ku.dk, lina@econ.ku.dk*

2
The Growing Family of Rasch Models

Jürgen Rost[1]

ABSTRACT A family of Rasch models is defined in terms of the prominent properties of all Rasch models, that is, separability, sufficiency, specific objectivity, and latent additivity. It is argued that concepts such as item homogeneity, person homogeneity, and unidimensionality do not hold for all generalizations of the Rasch model (RM). Four directions of generalizing the model are discussed: the multidimensional, the ordinal polytomous, the linear logistic, and the mixture distribution generalization. A hierarchical system of generalized Rasch models is presented. Four out of these eight models are discussed in the literature and can be applied by means of reliable computer software.

[1]Institute for Science Education, University of Kiel, Olshausenstr. 62, D-24098 Kiel, Federal Republic of Germany; e-mail: *rost@ipn.uni-kiel.de*

3

Gain Scores Revisited Under an IRT Perspective

Gerhard H. Fischer[1]

ABSTRACT For the measurement and statistical assessment of individual gain scores based on item sets that satisfy the assumptions of the Rasch, Rating Scale, or Partial Credit Models, a conditional maximum likelihood estimator, Clopper–Pearson confidence intervals, uniformly most accurate confidence intervals, and uniformly most powerful unbiased tests of the hypothesis of no change are presented. All methods are grounded on the exact conditional distribution of the gain score, given the total score for both time points, so that no asymptotic approximations are required. Typical applications of the methods are mentioned.

[1]Department of Psychology, University of Vienna, Liebiggasse 5, A-1010 Wien, Austria; e-mail: *gh.fischer@univie.ac.at*

4

Modeling Learning in Short-Term Learning Tests

Karl Christoph Klauer[1] and Hubert Sydow[2]

ABSTRACT A new model-based method for analyzing learning processes in so-called short-term learning tests is proposed and applied. Learning is measured by means of a two-dimensional item response theory model that incorporates two latent factors: ability level and learning ability. The model allows one to assess the impact of both variables and their correlation for given data sets in a manner that implicitly corrects for statistical artifacts that arise in conventional analyses. The model is applied to four short-term learning tests administered to $N = 434$ children aged five to six years. Model tests and tests for empirical validity of the model parameters succeed in establishing construct validity for one of these tests.

[1]Psychologisches Institut, Rheinische Friedrich-Wilhelms-Universität Bonn, Römerstr. 164, 53117 Bonn, FR Germany; e-mail: *christoph.klauer@uni-bonn.de*

[2]Institut für Psychologie, Humboldt-Universität Berlin, Oranienburger Str. 18, 10178 Berlin, FR Germany; e-mail: *hubert.sydow@psychologie.hu-berlin.de*

5
An IRT Model for Multiple Raters

Norman D. Verhelst[1] and Huub H.F.M. Verstralen[2]

ABSTRACT An IRT model for multiple ratings is presented. If it is assumed that the quality of a student performance has a stochastic relationship with the latent variable of interest, it is shown that the ratings of several raters are not conditionally independent given the latent variable. The model gives a full account of this dependence. Several relationships with other models appear to exist. The proposed model is a special case of a nonlinear multilevel model with three levels, but it can also be seen as a linear logistic model with relaxed assumptions (LLRA). Moreover, a linearized version of the model turns out to be a special case of a generalizability model with two crossed measurement facets (items and raters) with a single first-order interaction term (persons and items). Using this linearized model, it is shown how the estimated standard errors of the parameters are affected if the dependence between the ratings is ignored.

[1]National Institute for Educational Measurement (CITO), P.O. Box 1034, 6801 MG Arnhem, The Netherlands; Faculty of Educational Science and Technology, University of Twente, The Netherlands; e-mail: *norman.verhelst@cito.nl*

[2]National Institute for Educational Measurement (CITO), P.O. Box 1034, 6801 MG Arnhem, The Netherlands; e-mail: *huub.verstralen@cito.nl*

6

Conditional Independence and Differential Item Functioning in the Two-Parameter Logistic Model

Herbert Hoijtink[1]

ABSTRACT The two-parameter logistic model is among other things characterized by conditional independence and the absence of differential item functioning. Two fit statistics are proposed that can be used to investigate conditional independence and differential item functioning globally and at the item level. The statistics are evaluated using their posterior predictive distribution.

[1]Department of Methodology and Statistics, University of Utrecht, P.O. Box 80140, 3508 TC Utrecht, The Netherlands; e-mail: *h.hoijtink@fss.uu.nl*

7

Differential Item Functioning Depending on General Covariates

Cees A.W. Glas[1]

ABSTRACT Item response theory (IRT) is a powerful tool for the detection of differential item functioning (DIF). It is shown that the class of IRT models with manifest predictors is a comprehensive framework for the detection of DIF. These models also support the investigation of the causes of DIF. In principle, the responses to every item in a test can be subject to DIF, and traditional IRT-based detection methods require one or more estimation runs for every single item. Therefore, Glas (1998) proposed an alternative procedure that can be performed using only a single estimate of the item parameters. This procedure is based on the Lagrange multiplier test or the equivalent Rao efficient score test. In this chapter, the procedure is generalized in various directions, the most important one being the possibility of conditioning on general covariates. A small simulation study is presented to give an impression of the power of the test. In an example using real data it is shown how the method can be applied to the identification of main and interaction effects in DIF.

[1]Department of Educational Measurement and Data Analysis, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands; e-mail: *glas@edte.utwente.nl*

8

Statistical Tests for Differential Test Functioning in Rasch's Model for Speed Tests

Margo G.H. Jansen[1] and Cees A.W. Glas[2]

ABSTRACT We consider a latent trait model developed by Rasch for the response time on a set of pure speed tests, which is based on the assumption that the test response times are approximately gamma distributed with known index parameters and scale parameters depending on subject ability and test difficulty parameters. In this chapter, the principle of Lagrange multiplier tests is used to evaluate differential test functioning and subgroup invariance of the test parameters. Two numerical illustrations are given.

[1]Department of Education, University of Groningen, Grote Rozenstraat 38, 9712 TJ Groningen, The Netherlands; e-mail: *g.g.h.jansen@ppsw.rug.nl*

[2]Department of Educational Measurement and Data Analysis, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands; e-mail: *glas@edte.utwente.nl*

9
Expected Response Functions

Charles Lewis[1]

ABSTRACT Item response theory makes use of what are sometimes referred to as item response functions. Such a function is actually a probability density function (pdf) for the response of an individual to a test item, given the values of certain parameters, classified as item parameters and person parameters (or abilities). In testing, there is typically a calibration phase, in which item parameters are estimated and abilities are ignored. This is followed by an application phase, in which abilities are estimated while conditioning on the estimated values of the item parameters. A Bayesian alternative to treating the item parameters as known quantities involves replacing item response functions with another class of pdfs, referred to as expected response functions (ERFs). The latter take the uncertainty regarding item parameters into account for purposes of estimating abilities. This chapter provides a formal description of ERFs and briefly illustrates their application to the Rasch model for binary item responses.

[1]Educational Testing Service, Rosedale Road, Mailstop 03-T, Princeton, NJ 08541, USA; e-mail: *clewis@ets.org*

10

A Logistic IRT Model for Decreasing and Increasing Item Characteristic Curves

Edwin L. Klinkenberg[1]

ABSTRACT In item response theory, dominance relations are modeled by cumulative models, and proximity relations by unfolding models. Usually cumulative models are used for the measurement of latent abilities and unfolding models for the measurement of latent attitudes and preferences. The distinction between both types of measurement models is best represented by the shape of the item characteristic curve which is monotone increasing in the cumulative model and single-peaked in the unfolding model. A boundary case is a situation in which some items are monotone decreasing and others are monotone increasing. In this chapter, an extension of the one-parameter logistic model is proposed for that situation. It is shown that the parameters of the proposed model can be estimated by a simple data transformation. The model is illustrated using an empirical data set.

[1]Department of Psychiatry and Neuropsychology, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands; e-mail: *e.klinkenberg@np.unimaas.nl*

11

Using Parameter Expansion to Improve the Performance of the EM Algorithm for Multidimensional IRT Population-Survey Models

Donald B. Rubin[1] and Neal Thomas[2]

ABSTRACT Several large-scale educational surveys use item response theory (IRT) models to summarize complex cognitive responses and relate them to educational and demographic variables. The IRT models are often multidimensional with prespecified traits, and maximum likelihood estimates (MLEs) are found using an EM algorithm, which is typically very slow to converge. We show that the slow convergence is due primarily to missing information about the correlations between latent traits relative to the information that would be present if these traits were observed ("complete data"). We show how to accelerate convergence using parameter extended EM (PX-EM), a recent modification of the EM algorithm. The PX-EM modification is simple to implement for the IRT survey model and reduces the number of required iterations by approximately one-half without adding any appreciable computation to each EM-iteration.

[1]Datametrics Research, Inc. and Harvard University, 1 Oxford Street, Cambridge, MA 02138, USA; e-mail: *rubin@stat.harvard.edu*

[2]Datametrics Research, Inc. and Bristol-Myers Squibb, 61 Dreamlake Drive, Madison, CT 06443, USA; e-mail: *snthomas99@yahoo.com*

12

Cross-Validating Item Parameter Estimation in Adaptive Testing

Wim J. van der Linden and Cees A.W. Glas[1]

ABSTRACT From a statistical point of view, adaptive testing belongs to the class of problems of selecting a best subset of random variables. As in any other instance of this problem, adaptive testing is likely to capitalize on estimation error. An empirical example of the phenomenon is discussed, and a strategy for cross-validating results from adaptive testing is proposed. Results from a simulation study on the strategy are presented.

[1]Both authors are at the Department of Educational Measurement and Data Analysis, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands; e-mail: *vanderlinden@edte.utwente.nl* and *glas@edte.utwente.nl*

13

Imputation of Missing Scale Data with Item Response Models

Mark Huisman and Ivo W. Molenaar[1]

ABSTRACT Confronted with incomplete data due to nonresponse, a researcher may want to impute missing values to estimate latent properties of respondents. In this chapter the results of a simulation study are presented, investigating the performance of several imputation techniques. Some imputation procedures are based on item response theory (IRT) models, which can also be used to estimate latent abilities directly from the incomplete data by using incomplete testing designs when data are missing by design. This strategy has some serious disadvantages in the case of item nonresponse, because nonresponse is assumed to be ignorable and computational problems arise in scales with many items. In a second simulation study, the performance of some imputation techniques is compared to the incomplete design strategy, in the case of item nonresponse. The latter strategy results in slightly better ability estimates, but imputation is almost as good, especially when it is based on IRT models.

[1]Both authors are at the Department of Statistics and Measurement Theory, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands; e-mail: *m.huisman@ppsw.rug.nl* and *w.molenaar@ppsw.rug.nl*

14

On the Interplay Between Nonparametric and Parametric IRT, with Some Thoughts About the Future

Brian Junker [1]

ABSTRACT This chapter reviews some of the important research in nonparametric and parametric item response theory (IRT) today, and considers some current measurement challenges in education and cognitive psychology. This leads to assessment models that do not look very much like today's IRT models, but for which the tools and conceptual framework of nonparametric and parametric IRT are still quite well suited.

[1]Department of Statistics, Carnegie Mellon University, 232 Baker Hall, Pittsburgh, PA 15213, USA; e-mail *brian@stat.cmu.edu*

15

Reversibility Revisited and Other Comparisons of Three Types of Polytomous IRT Models

Bas T. Hemker[1]

ABSTRACT Molenaar (1983) was the first to distinguish and to compare three different types of polytomous item response theory (IRT) models: cumulative, continuation, and partial credit models. In his research, and later in research by others, the three types of models were compared on various criteria. One of the criteria introduced in IRT modeling is the property of reversibility. In this chapter, the results of a number of comparison studies are summarized. Next, the property of reversibility is further investigated. Finally, the relationship between the three types of models and psychological reality is discussed.

[1]National Institute for Educational Measurement (CITO), P.O. Box 1034, 6801 MG Arnhem, The Netherlands; e-mail: *bas.hemker@cito.nl*

16

Progress in NIRT Analysis of Polytomous Item Scores: Dilemmas and Practical Solutions

Klaas Sijtsma and L. Andries van der Ark[1]

ABSTRACT This chapter discusses several open problems in nonparametric polytomous item response theory: (1) theoretically, the latent trait θ is not stochastically ordered by the observed total score X_+; (2) the models do not imply an invariant item ordering; and (3) the regression of an item score on the total score X_+ or on the restscore R is not a monotone nondecreasing function and, as a result, it cannot be used for investigating the monotonicity of the item step response function. Tentative solutions for these problems are discussed. The computer program MSP for nonparametric IRT analysis is based on models that neither imply the stochastic ordering property nor an invariant item ordering. Also, MSP uses item restscore regression for investigating item step response functions. It is discussed whether computer programs may be based (temporarily) on models that lack desirable properties and use methods that are not (yet) supported by sound psychometric theory.

[1]Both authors are at the Department MTO, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands; e-mail: *k.sijtsma@kub.nl* and *a.vdark@kub.nl*

17

Two-Level Nonparametric Scaling for Dichotomous Data

Tom A.B. Snijders[1]

ABSTRACT This chapter considers a two-level design where the objects to be scaled are the higher-level units. Nested within each object are lower-level units, called subjects, and a set of dichotomous items is administered to each subject. The subjects are regarded as strictly parallel tests for the objects. Examples are the scaling of teachers on the basis of their pupils' responses, or of neighborhoods on the basis of responses by inhabitants. A two-level version of the nonparametric scaling method first proposed by Mokken (1971) is elaborated upon. The probabilities of positive responses to the items are assumed to be increasing functions of the value on a latent trait. The latent trait value for each subject is composed of an object-dependent value and a subject-dependent deviation from this value. The consistency of responses within, but also between objects is expressed by two-level versions of Loevinger's H-coefficients. The availability of parallel tests is used to calculate a reliability coefficient.

[1]Department of Statistics and Measurement Theory, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands; e-mail: *t.a.b.snijders@ppsw.rug.nl*

18

The Circles of Our Minds: A Nonparametric IRT Model for the Circumplex

Robert J. Mokken,[1] Wijbrandt H. van Schuur,[2] and Ard Jan Leeferink[3]

ABSTRACT A nonparametric item response theory (IRT) model for the circumplex is introduced, based on Mokken's (1971) nonparametric IRT model for cumulative scaling, and Van Schuur's (1984) nonparametric IRT model for unfolding. Some examples of circumplex representations in the social sciences are given. Model fit is based first on an extension of Loevinger's coefficient of homogeneity using quadruples as elementary units of analysis. Diagnostics for the probabilistic circumplex are suggested. Assignment of (ordinal) scale values is based on the notion of item steps, as developed by Molenaar (1983). The model presented here is based on dichotomous *pick any/k* data. Suggestions for extension to rank m/k data and to polytomous data are discussed.

[1]Faculty of Social and Behavioral Sciences, University of Amsterdam. Private address: Franklinstraat 18, 1171 BL Badhoevedorp, The Netherlands; e-mail: *mokken@ccsom.uva.nl*

[2]Department of Sociology, University of Groningen, Grote Rozenstraat 31, 9712 TG Groningen, The Netherlands; e-mail: *h.van.schuur@ppsw.rug.nl*

[3]Department of Policy Research of the City of Groningen, Postbus 1125, 9701 BC Groningen, The Netherlands; e-mail: *j.g.y.leeferink@sozawe.groningen.nl*

19

Using Resampling Methods to Produce an Improved DIMTEST Procedure

William Stout,[1] Amy Goodwin Froelich,[2] and Furong Gao[3]

ABSTRACT Following in the nonparametric item response theory tradition, DIMTEST (Stout, 1987) is an asymptotically justified nonparametric procedure that provides a test of hypothesis of unidimensionality of a test data set. This chapter introduces a new bias correction method for the DIMTEST procedure based on the statistical principle of resampling. A simulation study shows this new version of DIMTEST has a Type I error rate close to the nominal rate of $\alpha = 0.05$ in most cases and very high power to detect multidimensionality in a variety of realistic multidimensional models. The result with this new bias correction method is an improved DIMTEST procedure with much wider applicability and good statistical performance.

[1]Department of Statistics, University of Illinois at Urbana-Champaign, 725 S. Wright Street, Champaign, IL 61820, USA; e-mail: *stout@stat.uiuc.edu*

[2]Department of Statistics, University of Illinois at Urbana-Champaign, 725 S. Wright Street, Champaign, IL 61820, USA; e-mail: *amyf@stat.uiuc.edu*

[3]CTB/McGraw-Hill Research Department, 20 Ryan Ranch Road, Monterey, CA 93940, USA; e-mail: *FGao@ctb.com*

20

Person Fit Across Subgroups: An Achievement Testing Example

Rob R. Meijer and
Edith M.L.A. van Krimpen-Stoop[1]

ABSTRACT Item response theory (IRT) models are used to describe answering behavior on tests and examinations. Although items may fit an IRT model, some persons may produce misfitting item score patterns, for example, as a result of cheating or lack of motivation. Several statistics have been proposed to detect deviant item score patterns. Misfitting item score patterns may be related to group characteristics such as gender or race. Investigating misfitting item score patterns across different groups is strongly related to differential item functioning (DIF). In this study the usefulness of person fit to compare item score patterns for different groups was investigated. In particular, the effect of misspecification of a model due to DIF on person fit was explored. Empirical data of a math test were analyzed with respect to misfitting item score patterns and DIF for men and women and blacks and whites. Results indicated that there were small differences between subgroups with respect to the number of misfitting item score patterns. Also, the influence of DIF on the fit of a score pattern was small for both gender and ethnic groups. The results imply that person-fit analysis is not very sensitive to model misspecification on the item level.

[1]Both authors are at the Department of Educational Measurement and Data Analysis, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands; e-mail: *meijer@edte.utwente.nl* and *krimpen@edte.utwente.nl*

21

Single-Peaked or Monotone Tracelines? On the Choice of an IRT Model for Scaling Data

Wendy J. Post,[1] Marijtje A.J. van Duijn,[2] and Berna van Baarsen[3]

ABSTRACT It is investigated how a motivated choice can be made for the analysis of an item set using either a cumulative item response model with monotone tracelines, modeling dominance relations, or a unimodal item response model with single-peaked tracelines, modeling proximity relations. The focus is on item sets consisting of positively and negatively formulated items (with respect to the latent trait to be measured), where the common practice is to reverse one type of item. The differences between the cumulative and unimodal model are studied theoretically, in terms of item location and item order, and empirically, in a reanalysis of a sample of De Jong Gierveld loneliness scale data. Item locations, and, in the case of the unimodal model, also subject locations are shown to be important determinants of the differences. For the loneliness scale data the analysis with the unimodal model is preferred over the cumulative model. An outline of a recommended strategy for an IRT analysis of scaling data is given.

[1]Department of Medical Statistics, Leiden University Medical Center, P.O. Box 9604, 2300 RC Leiden, The Netherlands; e-mail: *w.j.post@mta.azg.nl*

[2]Department of Statistics and Measurement Theory, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands; e-mail: *m.a.j.van.duijn@ppsw.rug.nl*

[3]Department of Philosophy and Medical Ethics, Free University, Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands; email: *b.vbaarsen.metamedica@med.vu.nl*

22

Outline of a Faceted Theory of Item Response Data

Gideon J. Mellenbergh[1]

ABSTRACT Coombs' (1964) *Theory of Data* is a classification system of behavioral data. In his system, item response data are individual–comparison or individual–stimulus differences data. In this contribution, item response data are further classified using the following facets: Intended Behavior Type, Task Type, Intermediating Variable Type, Construct Level Type, Construct Dimension Type, Recording Type, Scaling Type, Person × Stimulus Interaction Type, Construct Relation Type, and Step Task Type. These facets cannot be completely crossed because some of the facets are nested within certain (combinations of) cells of crossed facets. The facets constitute a theory of item responses (TIR). The TIR and item response theory (IRT) complete each other: the TIR structures the item response data, while IRT explains these data. The TIR can be used to facilitate the choice of an appropriate IRT model.

[1]Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands; e-mail: *ml_Mellenbergh@macmail.psy.uva.nl*