

**STATISTICAL METHODS IN  
PSYCHIATRIC RESEARCH,  
WITH SPECIAL REFERENCE  
ON FACTOR ANALYSIS**

**JOUKO  
MIETTUNEN**

Department of Psychiatry,  
Department of Public Health Science and  
General Practice,  
University of Oulu  
Tampere School of Public Health,  
University of Tampere

OULU 2004

Abstract in Finnish





*JOUKO MIETTUNEN*

**STATISTICAL METHODS IN  
PSYCHIATRIC RESEARCH,  
WITH SPECIAL REFERENCE ON  
FACTOR ANALYSIS**

Academic Dissertation to be presented with the assent of the Faculty of Medicine, University of Oulu, for public discussion in the Väinö Pääkkönen Hall of the Department of Psychiatry, on June 4th 2004, at 12 noon.

OULUN YLIOPISTO, OULU 2004

Copyright © 2004  
University of Oulu, 2004

Reviewed by  
Docent Jaana Suvisaari  
Docent Erkki Alanen

ISBN 951-42-7366-4 (nid.)  
ISBN 951-42-7367-2 (PDF) <http://herkules.oulu.fi/isbn9514273672/>  
ISSN 0355-3221 <http://herkules.oulu.fi/issn03553221/>

OULU UNIVERSITY PRESS  
OULU 2004

## **Miettunen, Jouko, Statistical methods in psychiatric research, with special reference on factor analysis**

Department of Psychiatry, Department of Public Health Science and General Practice, University of Oulu, P.O.Box 5000, FIN-90014 University of Oulu, Finland; Tampere School of Public Health, University of Tampere, FIN-33014, Finland

2004

Oulu, Finland

### ***Abstract***

This interdisciplinary study describes in the first part the frequency with which various statistical research designs and methods are reported in psychiatric journals, and investigates how the use of these methods affect the visibility of the article in the form of received citations. In the second part focus is specifically on factor analysis, and the study presents two applications of this method.

Original research articles (N = 448) from four general psychiatric journals in 1996 were reviewed. The journals were the American Journal of Psychiatry, the Archives of General Psychiatry, the British Journal of Psychiatry and the Nordic Journal of Psychiatry. There were differences in the utilisation of statistical procedures among the journals. The use of statistical methods was not strongly associated with the further utilisation of an article. However, extended description of statistical procedures had a positive effect on the received citations.

Factor analysis is a statistical method based on correlations of the variables, which is often used when validity and structure of psychiatric instruments are studied. Exploratory factor analysis is designed to explore underlying latent factors, and in confirmatory factor analysis the aim is to verify the factor structure based on earlier findings in other data sets. Using data from the 31-year follow-up of the Northern Finland 1966 Birth Cohort Study this study aimed to demonstrate the validity and factor structure of scales measuring temperament (Tridimensional Personality Questionnaire, TPQ, and Temperament and Character Inventory, TCI) and alexithymia (20-item Toronto Alexithymia Scale, TAS-20).

The results of exploratory factor analysis indicated good performance of the TCI and TPQ, though the results suggested that some developmental work is still needed. Of the two scales, the TCI worked psychometrically better than the TPQ. A confirmatory factor analysis showed that the three-factor model of TAS-20 was in agreement with the Finnish version of the scale.

To conclude, future authors of psychiatric journals might apply these results in designing their research to present intelligible and compact analysis combined with a high quality presentation technique. Results of the factor analyses showed that the TPQ, TCI and TAS-20 can be used also in their Finnish versions.

*Keywords:* alexithymia, bibliometrics, personality, validity



## **Miettunen, Jouko, Tilastotieteelliset menetelmät, erityisesti faktorianalyysi, psykiatrisessa tutkimuksessa**

Psykiatrian klinikka, Kansanterveystieteen ja yleislääketieteen laitos, Oulun yliopisto, PL 5000, 90014 Oulun yliopisto; Terveystieteen laitos, Tampereen yliopisto, 33014 Tampereen yliopisto  
2004

Oulu, Finland

### ***Tiivistelmä***

Tämä poikkitieteellinen tutkimus kuvaa erilaisten tilastotieteellisten menetelmien yleisyyttä ja merkitystä psykiatriassa. Tutkimuksen ensimmäisessä osassa tutkitaan erilaisten tilastomenetelmien ja tutkimusasetelmien osuutta psykiatrisissa artikkeleissa ja lisäksi käytettyjen menetelmien vaikutusta artikkelien saamien viittausten lukumäärään. Tutkimuksen toisessa osassa keskitytään faktorianalyysiin ja esitetään kaksi siihen liittyvää sovellusta.

Aineiston muodostavat alkuperäistuloksia esittelevät artikkelit (N = 448) neljästä eri psykiatrian tieteellisestä yleislehdessä vuodelta 1996. Kyseiset lehdet ovat American Journal of Psychiatry, Archives of General Psychiatry, British Journal of Psychiatry ja Nordic Journal of Psychiatry. Lehdet erosivat toisistaan tilastotieteellisten menetelmien käytössä ja tulosten esittämisessä. Tilastotieteellisten menetelmien käytöllä ei ollut suurta vaikutusta artikkelien saamien viittausten lukumäärään, mutta laajalla menetelmien kuvauksella oli positiivinen vaikutus viittausten lukumäärään.

Faktorianalyysi on tilastotieteellinen tutkimusmenetelmä, jota käytetään tutkittaessa millaisista osatekijöistä erilaiset monimutkaiset ilmiöt koostuvat. Erityisesti tutkittaessa psykiatristen mittareiden validiteettia ja rakennetta faktorianalyysi on osoittautunut hyödylliseksi. Eksploratiivisessa faktorianalyysissä tarkoituksena on etsiä taustalla olevia piileviä muuttujia ja konfirmatorisessa faktorianalyysissä tarkoitetaan vahvistaa aiemmissa tutkimuksissa todettu mittarin faktorirakenne. Tässä tutkimuksessa hyödynnetään aineistoa Pohjois-Suomen vuoden 1966 syntymäkohortin 31 vuoden seurannasta. Aineiston avulla tutkitaan temperamenttia (Tridimensional Personality Questionnaire, TPQ, ja Temperament and Character Inventory, TCI) ja aleksitymiaa (20-item Toronto Alexithymia Scale, TAS-20) tutkivien mittareiden suomenkielisten käännösten validiteettia ja faktorirakennetta.

Eksploratiivisen faktorianalyysin tulokset kertoivat, että TPQ ja TCI toimivat hyvin myös suomenkielellä. Kuitenkin mittareissa on vielä kehittämisen varaa. TCI:n psykometriset ominaisuudet olivat paremmat kuin TPQ:n. Aleksitymiamittarin TAS-20 konfirmatorinen faktorianalyysi osoitti että aiemmin julkaistu kolmen faktorin malli toimi hyvin myös suomalaisella versiolla.

Psykiatristen artikkelien kirjoittajat voivat hyödyntää tämän tutkimuksen tuloksia suunnitellessaan psykiatrista tutkimusta suuntaan, jossa selkeä ja tiivis tulosten analysointitapa ja korkealaatuinen tulosten esitystapa korostuu. Faktorianalyysi soveltuu hyvin mittarin validiteetin tutkimiseen. Tutkimus osoitti TPQ-, TCI- ja TAS-20-mittareiden suomenkielisten versioiden validiteetin.

*Asiasanat:* aleksitymia, bibliometriikka, temperamentti, validiteetti





## Acknowledgements

This work was carried out at the Department of Psychiatry, at the Department of Public Health Science and General Practice, University of Oulu, and at the Tampere School of Public Health, University of Tampere. I wish to express my sincere gratitude to the following co-workers, without whom this study would not have been completed.

I am most grateful to my supervisors Professor Matti Isohanni, M.D., Ph.D., Department of Psychiatry, University of Oulu and Oulu University Hospital; Professor Matti Joukamaa, M.D., Ph.D., Department of Social Psychiatry, Tampere School of Public Health, University of Tampere and Department of Psychiatry, Tampere University Hospital; and Docent Pentti Nieminen, Ph.D., Medical Informatics Group, University of Oulu. The supervisors have given me great guidance during the work with the original articles and with the summary of the thesis.

I wish to express my warmest gratitude to Professor (emerita) Paula Rantakallio, M.D. Ph.D., and to Professor Marjo-Riitta Järvelin, M.D., Ph.D., Department of Public Health Science and General Practice, University of Oulu, for making it possible to use the valuable data, the Northern Finland 1966 Birth Cohort, as a study material.

I am also grateful to Professor (emeritus) Pekka Tienari, M.D., Ph.D., and to Professor Juha Veijola, M.D., Ph.D., for their help during this work and in various other projects. I also wish express my gratitude to Docent Juha Moring, M.D. Ph.D., the head of the Department of Psychiatry, Oulu University Hospital, for facilities and financial support for my study.

I owe my warmest thanks also to other co-authors in the original publications: Jesper Ekelund, M.D., Ph.D., Professor Nelson Freimer, M.D., Ph.D., Jari Jokelainen, M.Sc., Liisa Kantojärvi, M.D., Juha T Karvonen, M.D., Professor Jussi Kauhanen, M.D., Ph.D., Pirkko Kokkonen, M.D., Minna Koskinen, M.D., Professor Juhani Julkunen, Ph.D., Dirk Lichtermann, M.D., Ph.D., Kristian Läksy, M.D., Ph.D., and Professor Leena Peltonen, M.D., Ph.D. I wish to thank also docent Ritva Tikkanen, M.D., Ph.D., for the Finnish translation of the temperament questionnaires utilised in the original study III. I thank also Outi Väkevää, M.Sc., and Anna Vuolteenaho, M.A., for their corrections in the English language and Outi Mikkola, M.A., for her corrections in the language of the Finnish abstract. I thank Mr. Ville Varjonen for his help in typographic editing of the thesis.

I acknowledge also deep indebtedness to the official referees of the dissertation. I thank docent Jaana Suvisaari, M.D., Ph.D., National Public Health Institute, Helsinki, and docent Erkki Alanen, Ph.D., National Public Health Institute, Turku, for their valuable advice and constructive criticism.

Many persons at the Department of Psychiatry, University of Oulu and Oulu University Hospital have supported this work. I thank especially Professor Mauri Marttunen, M.D., Ph.D., Professor Pirkko Räsänen, M.D., Ph.D., Professor Hannu Koponen, M.D., Ph.D., Erika Lauronen, B.Med., Helinä Hakko, Ph.D., Pekka Laine, M.D., Ph.D., Kaisa Viilo, B.Sc., Ms. Raija Anttonen, Ms. Pirkko Kaan, Ms. Minna Lakkapää, Mr. Alpo Peltoniemi and Mr. Otto Moilanen for their help during the work.

The following foundations are acknowledged for having provided financial support for the study: Emil Aaltonen Foundation, Finnish Cultural Foundation, Jalmari and Rauha Ahokas Foundation, Finnish Academy and the European Commission, Quality of Life and Management of Living Resources Programme. The permission of the publishers to reprint the original articles is acknowledged.

Finally, my deepest thanks go to my friends and to my family, without their support this thesis would not have been completed. This work is dedicated to the memory of my mother and father.

Oulu, May 2004

Jouko Miettunen

## Main concepts

Alexithymia	personality style, difficulty in recognising and describing one's own emotions and externally oriented style of thinking
Bibliometrics	study of written documents and their citation influence, a subfield of information science and of statistics
Citation	written reference given to previous work
Confirmatory factor analysis	factor analysis conducted to test hypotheses (or confirm theories) about the factors one expects to find
Exploratory factor analysis	factor analysis conducted to discover what latent variables are behind a set of variables or measures
Factor analysis	statistical technique where a set of variables is presented in terms of a smaller number of hypothetical variables
Harm avoidance	tendency to respond intensively to signals of aversive stimuli
Normative data	data from a representative sample of normal subjects to allow comparisons between subjects and samples
Novelty seeking	tendency to respond with intense excitement to novel stimuli
Persistence	tendency to persevere in behaviours associated with reward
Personality	complex pattern of thoughts, emotions, and behaviours that is stable across time and situations
Psychometrics	science of measuring psychological aspects of a person, a subfield of psychology and of statistics
Reward dependence	tendency to respond intensively to signals of reward
Statistics	science of collecting, analysing, presenting, and interpreting data

Temperament

biologically rooted individual differences in  
behaviour tendencies

Validity

the degree to which the instrument measures what it  
should measure.

## Abbreviations

AGFI	adjusted goodness-of-fit index
AGP	Archives of General Psychiatry
AJP	American Journal of Psychiatry
BJP	British Journal of Psychiatry
CAIC	Consistent Akaike's Information Criteria
GFI	goodness-of-fit index
HA	harm avoidance (dimension of TPQ and of TCI)
MMPI	Minnesota Multiphasic Personality Inventory
NFBC 1966	Northern Finland 1966 Birth Cohort project
NJP	Nordic Journal of Psychiatry
NS	novelty seeking (dimension of TPQ and of TCI)
P	persistence (dimension of TCI)
RD	reward dependence (dimension of TPQ and of TCI)
RMS	root-mean-square residuals
SD	standard deviation
TAS-20	Toronto Alexithymia Scale (20-item version)
TCI	Temperament and Character Inventory
TPQ	Tridimensional Personality Questionnaire



## **List of original publications**

This thesis is based on the following original publications, which are referred to in the text by the Roman numerals I–IV.

- I Miettunen J, Nieminen P & Isohanni M (2002) Statistical methodology in general psychiatric journals. *Nord J Psychiatry* 56: 223–228.
- II Miettunen J & Nieminen P (2003) The effect of statistical methods and study reporting characteristics on the number of citations: a study with four general psychiatric journals. *Scientometrics* 57: 379–390.
- III Miettunen J, Kantojärvi L, Ekelund J, Veijola J, Karvonen JT, Peltonen L, Järvelin M-R, Freimer N, Lichtermann D & Joukamaa M (2004) A large population cohort provides normative data for investigation of temperament. *Acta Psychiatr Scand* (in press).
- IV Joukamaa M, Miettunen J, Kokkonen P, Koskinen M, Julkunen J, Kauhanen J, Jokelainen J, Veijola J, Läksy K & Järvelin M-R (2001) Psychometric properties of the Finnish 20-item Toronto Alexithymia Scale. *Nord J Psychiatry* 55: 123–127.

In addition, some unpublished data have been included in this thesis. The published original papers are reprinted by permission of Taylor & Francis (I, IV), Kluwer Publications (II) and Blackwell Publishing (III).





# Contents

Abstract	
Tiivistelmä	
Acknowledgements	
Main concepts	
Abbreviations	
List of original publications	
Contents	
1 Introduction	17
2 Review of literature	19
2.1 Statistical methods in psychiatry	19
2.1.1 Use of statistical methods	19
2.1.2 Effect of methods on number of citations	20
2.2 Factor analysis	20
2.2.1 Concepts of factor analysis	20
2.2.2 Factor analysis in validity studies	22
2.3 Cloninger's model of temperament dimensions	23
2.3.1 Models of personality	23
2.3.2 Cloninger's TPQ and TCI	24
2.3.3 Validity studies of TPQ and TCI	25
2.3.4 Normative studies of TPQ and TCI	25
2.4 Alexithymia	26
2.4.1 Alexithymia construct	26
2.4.2 Instruments to measure alexithymia	26
2.4.3 Validity studies of TAS-20	27
2.5 Summary of the literature	27
3 Aims of the study	29
4 Material and methods	30
4.1 Study samples	30
4.1.1 Bibliometric sample of psychiatric articles (I, II)	30
4.1.2 Northern Finland 1966 Birth Cohort Study (III, IV)	30
4.1.3 Student sample (IV)	31

4.2 Assessments.....	33
4.2.1 Bibliometric studies (I, II).....	33
4.2.2 Factor analytic studies (III, IV).....	33
4.3 Variables.....	34
4.3.1 Bibliometric studies (I, II).....	34
4.3.2 Validity study of TPQ and TCI (III).....	35
4.3.3 Validity study of TAS-20 (IV).....	36
4.4 Statistical methods.....	37
4.4.1 Bibliometric studies (I, II).....	37
4.4.2 Validity study of TPQ and TCI (III).....	37
4.4.3 Validity study of TAS-20 (IV).....	38
4.5 Ethical considerations and personal involvement.....	38
5 Results.....	40
5.1 Statistical methodology in psychiatric journals (I).....	40
5.2 Effect of statistical methods on citations (II).....	46
5.3 Validity of TPQ and TCI (III).....	50
5.4 Validity of TAS-20 (IV).....	52
6 Discussion.....	54
6.1 Main findings.....	54
6.2 Discussion of results.....	54
6.2.1 Statistical methodology in psychiatric journals (I).....	54
6.2.2 Effect of statistical methods on citations (II).....	56
6.2.3 Validity of TPQ and TCI (III).....	57
6.2.4 Validity of TAS-20 (IV).....	58
6.3 Strengths and limitations of the study.....	59
6.3.1 Bibliometric studies (I, II).....	59
6.3.1.1 Strengths of the studies.....	59
6.3.1.2 Limitations of the studies.....	59
6.3.2 Factor analytic studies (III, IV).....	60
6.3.2.1 Strengths of the studies.....	60
6.3.2.2 Limitations of the studies.....	60
7 Conclusions.....	62
7.1 Main conclusions.....	62
7.2 Implications of the study.....	63
7.3 Future research.....	63
References	

# 1 Introduction

Statistics is a science of collecting, analysing, presenting and interpreting data. Statistical methods play an important role in empirical and quantitative psychiatric research. This is reflected in the high proportion of articles that are essentially statistical in their presentation (Pincus *et al.* 1993). Regardless of the quality of data and the variables chosen to express the results, the overt evidence of psychiatric research is produced as lists of numbers, tables, line graphs, bar and pie graphs, and other displays demonstrating the message called descriptive statistics.

The question of how research information is best communicated to other researchers is an ever-present problem. In psychiatric fields of medicine one way of disseminating research messages is traditional quantitative medical research, which is based on systematic data collection and usually also utilises statistical analysis methods. Many researchers have responded to advances in statistical software by adopting new statistical methods, e.g. latent modelling, and computer technology.

Bibliometrics is a science of studying written communication by systematically measuring and analysing research publications. Bibliometric methods can be used to evaluate methodological aspects of reported studies. Citation is a written reference given to previous work. This study applies citation counts to explore the effect of statistical methods and characteristics of study on a medical paper's visibility and later utilisation.

Garfield's impact factor measures the frequency with which an average article in a journal is cited (Garfield 1986). The use of impact factors in assessing scientific work has been criticised in many contexts (Adam 2002, West & McIlwaine 2002, Berghmans *et al.* 2003, Kurmis 2003). Nevertheless, in Finland government funding for medical research is partly based on publication points, with a sliding scale corresponding to the impact factor of the journals in which researchers publish their work (Adam 2002). This makes it especially interesting to perform bibliometric studies on medical research in Finland.

Statistics and bibliometrics are both independent sciences, but are frequently utilised in other fields of science. The current scientific research is team work (Isohanni *et al.* 2002) and also the study groups in psychiatry include more and more experts from other disciplines. The role of statisticians in this team work has become more important as the portion of quantitative studies has increased (Pincus *et al.* 1993).

In psychiatric research and clinical practice diagnosis is often based on interviews and various other instruments (questionnaires and rating scales). Often these instruments are developed in different countries and cultures, and the original instrument needs to be translated. A new translation should be validated, i.e. it should measure what it is meant to measure, similarly to the original scale.

Psychometrics is a science that combines psychology and statistics. On the other hand it is thought to be science of measuring psychological aspects of a person and on the other hand a subfield of statistics applied especially in psychology. Factor analysis is a statistical method that is often used when psychiatric or psychological instruments are validated. It is used for studying the dimensional structure of instruments. In exploratory factor analysis the aim is to discover what latent variables are behind a set of variables or measures, and in confirmatory factor analysis the aim is to test hypotheses (or confirm theories) about the factors one expects to find (Vogt 1993).

In this study factor analysis is utilised when studying the validity of measurements for alexithymia and personality in the Northern Finland 1966 Birth Cohort Project (NFBC 1966). NFBC 1966 consists of an unselected, general population based birth cohort of 12058 live births in the two northernmost provinces of Finland (Rantakallio 1988). A large part of the subjects participated in a follow-up study at the age of 31 years; this study included e.g. questionnaires to measure temperament and alexithymia.

Several models have been proposed for classifying temperament and personality. A widely adopted model has been that of Cloninger, who proposed that temperament can be divided into the following dimensions: novelty seeking (NS), harm avoidance (HA), reward dependence (RD) and persistence (P). Two questionnaires have been developed to measure these dimensions: Tridimensional Personality Questionnaire (TPQ) and Temperament and Character Inventory (TCI). (Cloninger 1986, 1987).

Alexithymia is a condition where a person has difficulties in recognising and describing one's own emotions and has an externally oriented style of thinking. Many different methods have been proposed for the assessment of alexithymia. In 1985 Taylor and his group developed a self-report questionnaire for the assessment of alexithymia, the Toronto Alexithymia Scale. (Taylor *et al.* 1985, 1997). There are no previous publications on validity of the Finnish translations of instruments for assessing Cloninger's temperament dimensions and alexithymia.

This interdisciplinary work combines fields of statistics, bibliometrics, psychiatry and psychology. The aim of this thesis is to study the use of statistical methods in psychiatric research and especially to utilise factor analysis in assessing structural validity of psychological scales measuring temperament (TPQ and TCI) and alexithymia (TAS-20).

## **2 Review of literature**

### **2.1 Statistical methods in psychiatry**

#### ***2.1.1 Use of statistical methods***

Several authors have carried out comprehensive studies of medical journals to see which statistical methods are most frequently used. Most of these statistical “content analyses” have examined a collection of subfield-oriented journals (Goldin *et al.* 1996, Wang & Zhang 1998) or other samples of research articles (Altman 1991b, Nieminen 1995, Walter 1995, Mariani & Marubini 2000, Hakko *et al.* 2002).

Thus far, there have not been many systematic assessments of methodology in scientific psychiatric literature. Previous evaluations include bibliometric studies of general psychiatric journals (Lloyd & Fletcher 1989, Pincus *et al.* 1993, McGuigan 1995) and some reviews of special fields (Cox *et al.* 1994, Nieminen 1995, Davis & Lipson 1996, Nieminen & Isohanni 1997, Morlino *et al.* 1997, Hakko *et al.* 2002). Hokanson *et al.* inventoried the statistical techniques described in 15 major psychiatric journals during 1983 and 1984. A dozen procedures typically encountered in intermediate-level statistics courses accounted for approximately 95% of all statistical methods reported. (Hokanson *et al.* 1986). Another study has looked at the statistical content of publications concerning therapeutic community and psychiatric ward. This survey concluded that standard statistical methods were characteristic tools in the reviewed articles, but an increasing number of research reports also employed multivariate techniques. (Nieminen 1995).

In psychiatric research early publications were often qualitative studies and included only descriptive data. The proportion of traditional research articles, which present and analyse with statistical methods original, firsthand data, collected in a systematic fashion, has risen significantly in major journals (Pincus *et al.* 1993). Another tendency seen in leading psychiatric journals is an increase of the diversity of statistical methods used.

The availability of statistical software packages for microcomputers has greatly facilitated extensive data analysis, thus increasing the quantity and complexity of usage.

The last 20 years have seen a further introduction of more advanced statistical techniques, such as regression modelling methods, structural equation models and time-to-event analysis (Walter 1995, Altman 1998, Everitt & Landau 1998, Altman 2000). The effect of this trend on the frequency and quality of the statistical methods used is still unclear.

### ***2.1.2 Effect of methods on number of citations***

The question of exactly which statistical concepts and techniques influence the scientific communication and to what extent this occurs is still largely unanswered.

Peritz analysed the association between study designs of clinical trials and citation counts. Sample size was associated with the frequency of citations: researchers preferred to cite large studies rather than small studies. (Peritz 1994). Nieminen (1996) has analysed the role of empirical research report type as an explanatory factor in citation performance of psychiatric research: quantitative reports were more frequently cited than qualitative ones.

Some non-methodological issues have affected the number of citations in the earlier studies. The nationality of the authors has had a clear impact on the number of citations: North American authors are cited more frequently than others (Lloyd & Fletcher 1989, Luwel 1999). Also the number of authors has had a positive impact on the number of citations in earlier studies (Herbertz 1995). In summary, there are not many studies related to the topic, and especially the association between the statistical procedures used and the citations received is presently unclear.

## **2.2 Factor analysis**

### ***2.2.1 Concepts of factor analysis***

Developed 100 years ago by Charles Spearman (1904), factor analysis is one of the oldest multivariate methods still in frequent use. Factor analysis refers to a variety of statistical techniques whose common objective is to represent a set of variables in terms of a smaller number of hypothetical variables. Factor analysis can be utilised especially when studying complex phenomena e.g. in social science or psychology. It has been a powerful tool in studies on intelligence and personality in particular. Another important use of factor analysis is in the construction of psychological scales, where factor analysis can be used to guide the selection of variables to the scale. (Kline 1994).

The first step of the analysis is to examine the interrelationships of the variables. Inspection of the correlation (or covariance) matrix may show that there are positive relationships among variables, and that within some subsets of variables the relationships are higher than those between the subsets. In exploratory factor analysis the aim is to

explore underlying latent factors, and in confirmatory factor analysis the aim is to verify the factor structure based on earlier findings. (Kim & Mueller 1978a, Metsämuuronen 2003).

There are some important concepts and definitions in factor analysis. *Factor* is a unidimensional construct or dimension within a data set which is characterised by the variables which it is comprised. A major step in exploratory factor analysis is to find the number of factors that can adequately explain the observed correlations among the observed variables. The most common methods of obtaining the initial solution include principal component analysis, principal axis factoring, maximum likelihood method, least-squares method, alpha factoring and image factoring (Kim & Mueller 1978a, 1978b, Nummenmaa *et al.* 1997). Principal component analysis is mathematically and philosophically different from the other methods, but the steps of analyses are similar (Metsämuuronen 2003).

To obtain an initial solution the researcher must provide either the number of common factors to be extracted or the criterion by which such a number can be determined. *Eigenvalue* is a mathematical property of a (correlation) matrix, which is also used when estimating the variance explained by each factor. A *scree-test* is a graphical analysis of eigenvalues, which also can be used in concluding the number of factors to be extracted. A common rule is to rotate factors with eigenvalues greater than 1. Kline (1994) notes that in large matrices this may overestimate the number of factors and he prefers the use of a scree-test.

In the next step initial solution is rotated, i.e. described by a different system of coordinates, to give a simplified presentation of the solution. There are several rotation techniques. *Orthogonal rotations* (e.g. varimax) give factors that are uncorrelated with each other, whereas *oblique rotations* (e.g. oblimin and promax) allow factors to be correlated with each other. (Kim & Mueller 1978b). In many situations different rotation methods give quite similar results, but in some situations the character of the data (phenomenon) may require either orthogonal or oblique rotation to be selected. *Factor loadings* are equivalent to correlations between factors and variables. A factor loading greater than 0.40 is often interpreted as significant, although other criteria are also used. The *communality* of an observed variable is the sum of squares of the factor loadings for that variable. (Kim & Mueller 1978a). The variables in factor analysis, especially when analysing instruments, are often ordinal, not continuous variables. Basic statistical programs performing exploratory factor analysis calculate Pearson correlation coefficients which are meant for normally distributed continuous variables; this may lead to lack of efficiency in the analysis. There are also factor analysis methods (Bartholomew 1987) and software (Muthén & Muthén 1998) developed especially to ordinal and binary data.

In confirmatory factor analysis one hypothesises before hand the number of factors and which variables are included in each factor. (Kim & Mueller 1978b). For example, when the researcher does not hypothesise the variable to load on a factor the loading is set to zero. Also the correlations between the hypothesised factors can be constrained in the model. Factor loadings of original variables and correlations between the factors are estimated most commonly using maximum likelihood method. The goodness-of-fit of the model is evaluated using various statistical indices, e.g. chi-square goodness-of-fit, the

goodness-of-fit index (GFI), and the root-mean-square residuals (RMS). (Finch & West 1997, Nummenmaa *et al.* 1997).

These presented and other methodological issues in factor analysis are discussed in books and articles published on factor analysis (Kim & Mueller 1978a, Kim & Mueller 1978b, Jöreskog & Sörbom 1979, Bartholomew 1987, Loehlin 1987, Child 1990, Comrey & Lee 1992, Joliffe & Morgan 1992, Kline 1994). There are also comprehensive books on factor analysis in Finnish (Nummenmaa *et al.* 1997, Metsämuuronen 2003). *Psychometrika* is a journal, which presents methods and statistical techniques for the evaluation of psychological data. The journal has published several methodological articles also on factor analysis. Articles up to the mid-1980s have been reviewed by Mulaik (1986).

Basic statistical packages, such as SAS (SAS Institute 1999), SPSS (SPSS Inc. 2001) and STATA (Stata Corporation 2001) can perform exploratory factor analysis. AMOS in SPSS (Arbuckle 1997), PROC CALIS in SAS (SAS Institute 1999), EQS (Bentler 1995), LISREL (Jöreskog & Sörbom 1996) and MPLUS (Muthén & Muthén 1998) are examples of statistical programs developed especially for confirmatory factor analysis and related statistical techniques. Related methods include e.g. other forms of structural equation modelling and path analysis (Loehlin 1987, Bollen 1989, Bentler & Stein 1992, Finch & West 1997, MacCallum & Austin 2000).

### ***2.2.2 Factor analysis in validity studies***

Factor analysis and related techniques have been used especially in social and behavioural science, psychology and psychiatry. Most of the applications have been on validity of questionnaires and other instruments.

*Validity* is the degree to which the instrument measures what it should measure. *Construct validity* refers to the extent to which a theoretical concept is adequately or successfully measured. The two subcategories of construct validity are: convergent validity and discriminant validity. *Convergent validity* refers to the consistency of the results yielded from measuring the same concept with different measures. *Discriminant validity* refers to ability of a measure to differentiate a concept from other concepts. *Content validity* refers to the extent in which the concept to be measured is adequately covered by items or questions. *Face validity* refers to the extent in which a measure appears to measure what it aims at measuring. *Criterion validity* refers to the extent in which a measure will match up with the scores obtained from using the external or independent criteria. There are two types of criterion validity. The first type is *concurrent validity*. This refers to the ability of a measure to predict the current status of an individual. The second type is *predictive validity*, it refers to the ability of a measure to predict future performance from the current one. There are also other subtypes of validity and also other definitions for the presented subtypes. (Blacker & Endicott 2000, Morgan *et al.* 2001).

Several articles have been published where authors have aimed to validate a Finnish version of an original scale or where the authors have used Finnish data to construct a



new scale. Factor analysis was already used for this purpose in the 1950s (Takala & Hakkarainen 1953). Finnish researchers on psychology and psychiatry have utilised factor analysis frequently since then. The latest applications include e.g. validity studies of following scales: a self-image questionnaire (Laukkanen *et al.* 2000), an assessment of magical beliefs about food and health (Lindeman *et al.* 2000), a test of theory of intelligence (Sternberg *et al.* 2001), assessment of strategies in families (Åstedt-Kurki *et al.* 2001), a strengths and difficulties questionnaire (Koskelainen *et al.* 2001), a care dependency scale (Dijkstra *et al.* 2003) and a food neophobia scale (Ritchey *et al.* 2003). Other recent applications include articles where factor analysis was used to study whether alexithymia and depression are distinct or overlapping constructs (Hintikka *et al.* 2001) and where factor structure of psychotic symptoms in schizophrenia was investigated (Salokangas *et al.* 2002, Arajärvi *et al.* 2004).

## 2.3 Cloninger's model of temperament dimensions

### 2.3.1 Models of personality

Personality is thought to be a complex pattern of thoughts, emotions, and behaviours that is stable across time and many situations. Personality has frequently been studied in context of various somatic and psychiatric disorders. Several models have been proposed for classifying personality and temperament. A widely adopted model has been that of Cloninger, who originally proposed that there are three genetically homogeneous and independent dimensions of temperament. *Novelty seeking* (NS) is a tendency to respond with intense excitement to novel stimuli, or cues for potential rewards or potential relief of punishment and thereby activating/initiating behaviour. *Harm avoidance* (HA) is defined as a tendency to respond intensively to signals of aversive stimuli, thereby inhibiting/stopping behaviour. *Reward dependence* (RD) is a tendency to respond intensively to signals of reward, especially social rewards, thereby maintaining and continuing particular behaviours. NS is thought to be linked to the dopamine, HA to the serotonin and RD to the norepinephrine system. (Cloninger 1986, 1987). These systems play an important role in the pathophysiology and treatment of major mental disorders.

Later Cloninger also included character dimensions in his personality model (Cloninger *et al.* 1994). Other commonly used personality models include Eysenck's three-factor model of personality measured with the Eysenck Personality Questionnaire, EPQ, (Eysenck & Eysenck 1975) and the five-factor model of personality measured with the NEO Personality Inventory, NEO-PI (Costa & McCrae 1985) and with the NEO Five Factor Inventory, NEO-FFI (McCrae & Costa 2004). Minnesota Multiphasic Personality Inventory (MMPI) includes also scales related to personality and psychopathology (Dahlstrom *et al.* 1982). There have been validation studies using partly also the Finnish translations of Eysenck Personality Questionnaire (Barrett *et al.* 1998) and of NEO Personality Inventory (Pulver *et al.* 1995).

### 2.3.2 Cloninger's TPQ and TCI

Originally Cloninger developed a structured clinical interview, the Tridimensional Inventory of Personality Style (TIPS), to measure these three dimensions. This interview served as a starting point for the Tridimensional Personality Questionnaire (TPQ). Cloninger subsequently elaborated his initial model into a seven-factor model of personality, and modified the TPQ to incorporate this extended model, developing a new questionnaire called the Temperament and Character Inventory (TCI). (Cloninger *et al.* 1993, 1994).

The TCI assesses four temperament dimensions (HA, NS, RD and persistence) and three character dimensions (self-directedness, co-operativeness and self-transcendence). Persistence (P) had been in the TPQ as part of the RD-factor. P includes a tendency to persevere in behaviours that have been associated with reward or relief from punishment. (Cloninger *et al.* 1993, 1994). NS, HA and RD dimensions represent higher-order personality dimensions, composed of similarly motivated but differently expressed behaviours; in accord with this they are divided into subscales (Cloninger *et al.* 1993).

Examples of NS items (all reverse scored in analyses) from the TCI include "I'm slower than most people to get excited about new ideas and activities", "I usually think about all the facts in detail before I make a decision" and "I'm better at saving money than most people". Examples of HA items include "I usually am confident that everything will go well, even in situations that worry most people" (reversed), "I often feel tense and worried in unfamiliar situations, even when others feel there is no danger at all" and "I often avoid meeting strangers because I lack confidence with people I do not know". The RD scale consists of items such as "I'm strongly moved by sentimental appeals", "I don't open up much even with friends" (reversed) and "Other people often think that I am too independent because I won't do what they want" (reversed). The P scale consists of items such as "I often push myself to the point of exhaustion or try to do more than I really can", "I am usually so determined that I continue to work long after other people have given up" and "I am satisfied with my accomplishments, and have little desire to do better" (reversed). (Cloninger *et al.* 1994). A description of the development of the temperament scales and of the differences between the TPQ and TCI can be found in the TCI user guide (Cloninger *et al.* 1994).

The relationships between Cloninger's TCI and the five-factor model of personality (Costa & McCrae 1985, McCrae & Costa 2004) have been studied by De Fruyt *et al.* (2000). The five-factor model consists of factors named neuroticism, extraversion, openness to experience, agreeableness and conscientiousness. De Fruyt *et al.* (2000) concluded that there is considerable overlap between the models. They found out that HA correlates positively with neuroticism and negatively with conscientiousness and openness to experience, NS correlates negatively with conscientiousness, RD positively with extraversion, and P positively with conscientiousness. In contrast to the five-factor model, Cloninger's model provides biologically based structure of personality (Cloninger 1986, 1987, Cloninger *et al.* 1993).

### 2.3.3 *Validity studies of TPQ and TCI*

After the presentation of the factorial structure of the original TPQ (Cloninger 1986) studies using various versions of the scale have been conducted to examine its psychometric properties in a variety of study populations and in various languages. The main findings of these studies are that both the HA and NS scales are psychometrically sound and conform to Cloninger's theory, whereas the RD scale is psychometrically weaker and may be composed of two factors. (Kozeny *et al.* 1989, Svrakic *et al.* 1991, Takeuchi *et al.* 1993, Aschauer *et al.* 1994, Weyers *et al.* 1995, Zohar *et al.* 2001, Chen *et al.* 2002, Stewart *et al.* 2004).

The TCI has shown good psychometric properties in the English version (Cloninger *et al.* 1993, Svrakic *et al.* 1993). There have also been some other factor analytic studies of the TCI in different languages. These studies have mostly confirmed that the temperament scale consists of four factors (NS, HA, RD and P), although the RD dimension is still psychometrically weaker than the other dimensions. (Brändström *et al.* 1998, de la Rie *et al.* 1998, Pélissolo & Lépine 2000, Kijima *et al.* 2000, Gutiérrez *et al.* 2001, Sung *et al.* 2002, Parker *et al.* 2003b). The study by Parker *et al.* (2003a) indicates that TCI is likely to have applicability to Chinese subjects and argues against properties being constrained by the English language or by western culture. There are no publications on Finnish translations, but in his unpublished thesis Puttonen has studied the validity of a Finnish translation of TCI using a likert-type scale (1 = totally disagree, ..., 5 = totally agree). The study included 2109 subjects at the age of 20–34 years, using a different translation than in the current study. (Puttonen 1998).

### 2.3.4 *Normative studies of TPQ and TCI*

Normative data are collected from a representative sample of normal subjects to allow comparisons between subjects and samples. The original normative data for the TPQ were collected in the USA (Cloninger *et al.* 1991). Subsequent normative data have been presented from different populations, e.g. from Italy (Manfredonia *et al.* 1991), Yugoslavia (Svrakic *et al.* 1991), Austria (Aschauer *et al.* 1994), England (Otter *et al.* 1995), Taiwan (Chen *et al.* 2002) and Israel (Zohar *et al.* 2001). Studies of the TCI include normative data from the USA (Cloninger *et al.* 1993), Sweden (Brändström *et al.* 1998), the Netherlands (Duijsens *et al.* 2000), France (Pélissolo & Lépine 2000), Japan (Kijima *et al.* 2000) and Belgium (Hansenne *et al.* 2001). These studies have mostly reported higher scores for females in all temperament dimensions. Distributions of the TCI dimensions according to age and sex have been studied also from a cross-cultural perspective comparing data from Sweden, Germany and the USA. The authors concluded that, because of significant differences between genders, sex specific norms for the dimensions of the TCI must be given. (Brändström *et al.* 2001).

To date, no normative data for the TPQ or TCI in the Finnish population have been available. Such normative data are necessary to interpret the scores obtained with the TPQ and TCI in Finnish-speaking subjects, but are also of interest when addressing

cross-cultural issues about the assessment of personality in various countries (Svrakic *et al.* 1991, Péliissolo & Lépine 2000, Brändström *et al.* 2001).

## 2.4 Alexithymia

### 2.4.1 *Alexithymia construct*

The concept of alexithymia was derived from repeated clinical observations of a cluster of specific cognitive characteristics among certain patients, especially those suffering from psychosomatic diseases (Sifneos 1973). Nowadays it seems, that alexithymia is associated with many somatic disease and is a useful construct for understanding the role of personality in the pathogenesis of somatic diseases (Taylor 2000). Associations between alexithymia and somatic or psychiatric disorders have been studied frequently, in more than 1000 publications (Taylor 2000).

The alexithymia construct and its historical background have been reviewed in detail e.g. by Taylor *et al.* (1997). Theoretically the alexithymia construct is thought to be composed of the following features: difficulty in identifying feelings and distinguishing them from the bodily sensations of emotion, difficulty in describing feelings to others, constricted imaginal processes, as evidenced by a paucity of fantasies, and an externally oriented style of thinking (Sifneos 1973, Taylor *et al.* 1997).

Alexithymia is considered to be a stable personality trait (Honkalampi *et al.* 2001). Taylor and colleagues (Taylor *et al.* 1997) also describe relations between alexithymia and various personality models; comparisons have been made e.g. between the Toronto Alexithymia Scale and Eysenck Personality Questionnaire (EPQ). One study has analysed the relationship of alexithymic features with Cloninger's TCI. The authors conclude that alexithymia is captured only partly by Cloninger's model. Of the temperament dimensions a low RD was the best predictor for alexithymia. (Grabe *et al.* 2001).

### 2.4.2 *Instruments to measure alexithymia*

Since Sifneos (1973) coined the term alexithymia 30 years ago, many different methods have been proposed for its assessment. The developer of the term himself introduced an interviewer-rated measure, the Beth Israel Hospital Questionnaire (BIQ) and a self-report scale, the Shalling-Sifneos Personality Scale (Apfel & Sifneos 1979, Sifneos 1986). In addition, many other questionnaires have been developed to measure the alexithymia construct: the MMPI Alexithymia Scale (Kleiger & Kinsman 1980), the Alexithymia Provoked Response Questionnaire (Krystal *et al.* 1986), and the Analogue Alexithymia Scale (Faryna *et al.* 1986). Projective tests have also been used for measuring

alexithymia, such as the Archetypal9 Test (Cohen *et al.* 1985) and the Rorschach ink plot test (Acklin & Alexander 1988). It does seem, however, that most of the above-mentioned methods have some limitations, and some of them may measure different aspects of alexithymia (Bagby *et al.* 1986a, Bagby *et al.* 1988, Norton 1989, Bagby *et al.* 1991, Parker *et al.* 1991, Bagby & Taylor 1997). Newer alexithymia scales include the California Q-set Alexithymia Prototype (Haviland 1998), the Amsterdam Alexithymia Scale (Bermond *et al.* 1999) and the Bermond-Vorst Alexithymic Questionnaire (Vorst & Bermond 2001).

In 1985 Taylor and his group developed a self-report questionnaire for the assessment of alexithymia, the Toronto Alexithymia Scale (Taylor *et al.* 1985). Originally it consisted of 26 items. More recently, an improved, revised version of the scale, the 20-item Toronto Alexithymia Scale (TAS-20) has been introduced (Bagby *et al.* 1994a).

### **2.4.3 Validity studies of TAS-20**

The original 26-item Toronto Alexithymia Scale (Taylor *et al.* 1985) has a factor structure congruent with the alexithymia construct (Bagby *et al.* 1986b, Haviland *et al.* 1988, Bagby *et al.* 1990). However, the TAS-26 items directly assessing daydreaming and other imaginal activity were eliminated because of low corrected item-total correlations and/or high correlations with a measure of social desirability (Taylor *et al.* 1997). Internal consistency, test-retest reliability, and convergent, discriminant, and concurrent validity of TAS-20 have been shown to be good (Bagby *et al.* 1994a, Bagby *et al.* 1994b, Parker *et al.* 2003c, Taylor *et al.* 2003). Factor analyses of the measure showed that the scale consists of three factors that reflect the facets of alexithymia: difficulty in identifying feelings and distinguishing them from the bodily sensations of emotion (factor 1), difficulty in describing feelings to others (factor 2), and an externally oriented style of thinking (factor 3) (Parker *et al.* 1993). Consistency of the factor structure of TAS-20 has been shown with different English-speaking samples, with versions translated into almost twenty languages (Taylor *et al.* 2003). Recently some criticism to this factor solution has been presented and also other factorial structures have been studied (Loas *et al.* 1996, Kooiman *et al.* 2002, Müller *et al.* 2003). Content validity of the TAS-20 has been criticised by Sifneos. He preferred the TAS-26 as it also included items assessing daydreaming and other imaginal activity. (Sifneos 1996).

## **2.5 Summary of the literature**

The previous literature on the topic of this thesis can be summarised briefly as follows. The use of multivariate methods has increased in the last twenty years, but there are only few studies on the frequency of various statistical methods in psychiatric research. The influence of statistical methods on the number of citations is largely unknown. Factor analysis is a specific statistical method, which has been utilised in validity studies of

psychological scales. TAS-20 and Cloninger's temperament questionnaires (TPQ and TCI) have been validated in English and in various other languages, both not in Finnish. Most of the previous studies have confirmed the three-factor structure of TAS-20 and the four-factor structure of Cloninger's temperament dimensions. Although, in Cloninger's temperament model the reward dependence dimension has been psychometrically weaker than other dimensions in the earlier studies.

### **3 Aims of the study**

The bibliometric part of this study (I, II) utilises a sample of psychiatric articles and aims to:

1. Study how often tables, figures and various statistical methods are used to summarise the findings of research.
2. Find out whether there are differences between general psychiatric journals in the statistical content of published papers.
3. Study whether the use of different statistical methods and characteristics of study report affect the number of received citations.

The factor analytic part of this study (III, IV) utilises the Northern Finland 1966 Birth Cohort data and aims to

4. Present the validity, the Finnish normative data, and compare psychometric properties of the Tridimensional Personality Questionnaire (TPQ) and of the Temperament and Character Inventory (TCI).
5. Examine the factor structure and the validity of the Finnish version of the 20-item Toronto Alexithymia Scale (TAS-20).

## **4 Material and methods**

### **4.1 Study samples**

#### ***4.1.1 Bibliometric sample of psychiatric articles (I, II)***

The evaluation was limited to articles published in four general psychiatric journals: The American Journal of Psychiatry (AJP), The Archives of General Psychiatry (AGP), The British Journal of Psychiatry (BJP) and The Nordic Journal of Psychiatry (NJP). These journals were selected because AGP, AJP and BJP have the highest impact factors (IF) of all general psychiatric journals, and NJP is a Nordic journal representing the large group of journals with markedly lower IF. In 2002 the impact factors for the four journals were: AGP 11.622, AJP 6.458, BJP 4.374 and NJP 0.738.

For each of the journals, original quantitative research articles published in 1996 were scrutinised, using the same definition as Pincus *et al.* (1993). The reviewed articles reported original research findings based on the systematic collection of data and on the use of various statistical data analysis tools. The required sample size was at least 10 patients or study units or serial measurements in studies with fewer subjects. Articles that did not fulfil these requirements were excluded. Brief reports of AJP were not included due to their having a restricted number of pages and tables. The year 1996 was selected as it ensured a long enough period for citations to accumulate, and the sample still represented well the current use of statistical methods in psychiatric research. The total number of reviewed articles was 448.

#### ***4.1.2 Northern Finland 1966 Birth Cohort Study (III, IV)***

This study is part and one extension of the Northern Finland 1966 Birth Cohort project (NFBC 1966). The original sample was collected from a geographically defined area of

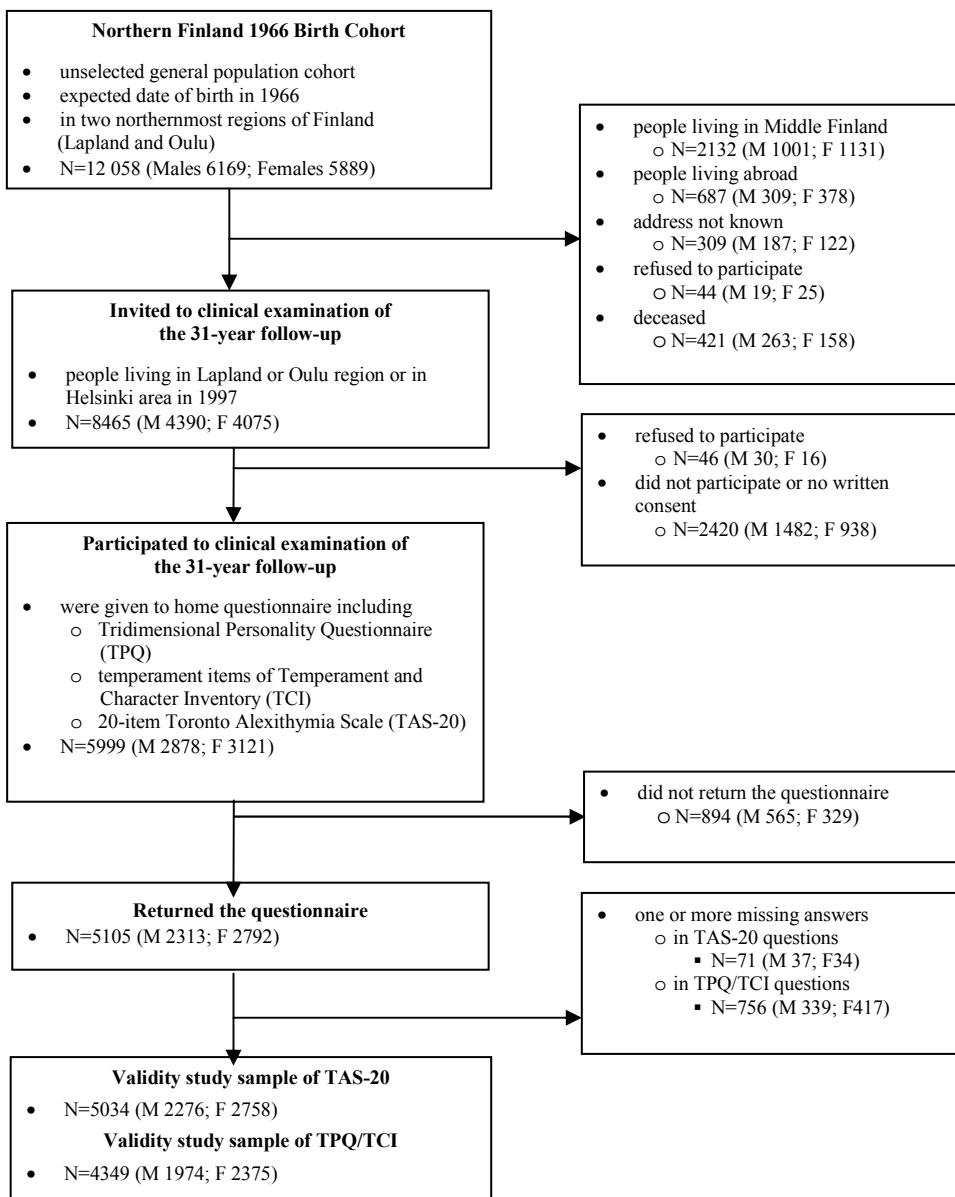


the two northernmost provinces of Finland. It consisted of an unselected, general population-based birth cohort of 12058 live births in Northern Finland whose expected delivery date fell in 1966, representing 96% of all such births (Rantakallio 1988). In 1997, 8411 cohort members were still living in the northernmost provinces of Finland and in the Helsinki area. As a part of the 31-year follow-up study they were invited to a clinical examination. Figure 1 describes the composition of the present study population. Of the invited subjects, 5915 participated in the field study and gave a written informed consent.

The questionnaires including TAS-20 and the temperament questions were given to the cohort members who participated in a clinical examination; 5105 responded (Figure 1). The TAS-20 questionnaire was filled in without any missing answers by 85.1% (5034) of the subjects. The temperament items were filled in without any missing answers by 72.5% (4349) of the subjects. Demographic data and comparisons between participants and non-participants are presented in the original articles III (TPQ/TCI) and IV (TAS-20).

#### ***4.1.3 Student sample (IV)***

The validity of TAS-20 was examined further by using a sample consisting of 243 university students from the University of Turku and 273 students from the Turku Nursing College. The sample was collected by child psychiatrist Minna Koskinen, MD, and psychiatrist Matti Joukamaa, MD, PhD. The students were informed that the aim of the research was to evaluate a personality scale. Informed consent was obtained from all participants. Of the participants 108 (21%) were male and 405 (78%) female (the sex of three subjects was unknown). The mean age for the total sample was 24.8 years (standard deviation, SD = 5.44); for men alone, 24.9 years (SD = 3.87); and for women, 24.8 years (SD = 5.80).



**Fig. 1. The flow chart of the validity study population of the Finnish translations of the Toronto Alexithymia Scale – 20 and of Cloninger's temperament dimensions. Excluded cases with the reason for exclusion presented on the right.**

## 4.2 Assessments

### 4.2.1 *Bibliometric studies (I, II)*

An experienced biostatistician (docent Pentti Nieminen) has earlier read and evaluated the statistical content of 223 psychiatric research articles using the same classification as in this study (Nieminen 1995, Nieminen & Isohanni 1997). The author was trained by him. After that the author classified the material of this study, for statistical methods using the categorisation described in Table 1.

Citations were collected from ISI (Institute for Scientific Information) Web of Science databases in November 2002. The databases were *Science Citation Index*, *Social Sciences Citation Index* and *Arts & Humanities Citation Index*. So the time-span to receive citations was approximately six years. A literature search in the database of the American Psychological Association, PsycINFO, was performed to estimate the number of factor analytic studies in psychiatry.

### 4.2.2 *Factor analytic studies (III, IV)*

The original English versions of the TPQ and TCI were translated into Finnish by one investigator (docent Ritva Tikkanen) and back-translated blindly into the original English scale by a professional English translator. The original version and the back-translation were compared and corrections were made accordingly. The translation was tested in a sample of 50 laboratory workers, the results indicated that no questions needed revision. The character items of the TCI were not used in this study because the main goal in using this questionnaire in the NFBC 1966 is to identify genetic factors underlying personality (Ekelund *et al.* 1999, Lichtermann *et al.* 2001), and character traits have been thought to be less heritable than temperament traits (Cloninger *et al.* 1993). Some researchers now suggest that character traits are also highly heritable (Ando *et al.* 2002).

The original English version of the TAS-20 was translated into Finnish by one investigator (docent Jussi Kauhanen) and back-translated blindly into the original English scale by another (see original article IV for more details).

The temperament and alexithymia items were asked as a part of a questionnaire also including various other mental health related true/false questions; a total of 351 questions was presented to subjects.

*Table 1. Categories of statistical procedures used to assess the statistical content of the articles.*

Category	Brief description
Basic methods	
Comparison of means	Test for differences in mean values in text, table or figure
t-tests	One-sample, matched-pair and two-sample t-tests
analysis of variance	Analysis of variance, analysis of variance with repeated measures, analysis of covariance (F-tests)
Cross-tabulation	Test for contingency table (Chi-square, Fisher's exact and McNemar's tests)
Non-parametric tests	Mann-Whitney, sign, Wilcoxon signed-rank, Kruskal-Wallis and Friedman tests
Correlation	Product-moment correlation, simple linear regression, non-parametric Spearman's and Kendall's correlations
Multiple comparisons	Procedures for handling multiple inference on same data sets including Bonferroni techniques, methods of Scheffe, Tukey etc.
Multivariate methods	
Regression models	Linear, non-linear, logistic and Cox's regression models
Manova	Multivariate analysis of variance
Factor analysis	Exploratory and confirmatory factor analysis, structural equation models
Other multivariate methods	Discrimination analysis, cluster analysis, log-linear modelling etc.
Specific methods	
Epidemiological statistics	Relative risk, odds ratio, log odds, incidence
Reliability	Percentage of agreement, sensitivity, specificity, kappa coefficient, intraclass correlation coefficients, intra- and inter-assay coefficient of variation
Other specific methods	Survival analysis, effect size, phi, adjustments for t-test etc.

## 4.3 Variables

### 4.3.1 Bibliometric studies (I, II)

Each of the selected articles was classified into one of three study design groups as defined by Hulley & Cummings (1988): *experimental*, *cross-sectional* and *longitudinal*. An experimental study was defined as one in which one or more variables were controlled by the investigator in order to monitor the effect on process or outcome. Cross-sectional and longitudinal studies are observational studies in which the investigator stands apart from the study arrangements. In a cross-sectional study, each subject is

examined on only one occasion, whereas in a longitudinal study, each subject is followed over a period of time and some measurements are taken in two or more points in time.

Studies were categorised as using p-values if the results of formal statistical significance testing were reported and as using confidence intervals if these were reported. The usage of tables and figures was also assessed. The presence or absence of specific statistical procedures and techniques in each article was reviewed. Table 1 lists the categories of statistical methods used, and gives a brief description of their content.

Psychiatric subfield of the studies was also determined by the author of this thesis. Each study was only classified into one category, best describing the content of the article in the question. The article of Morlino *et al.* (1997) and Medical Subject Headings from Medline database were used when designing categorisation. The following classification was used: epidemiology, clinical topics, psychopharmacology, biological topics and others. Epidemiological articles also included etiological (non-biological) articles, clinical topics included e.g. rehabilitation and treatment articles, biological articles included e.g. genetic and imaging studies, and other articles were e.g. validity studies of psychometrical scales and studies of prevalence.

The following information was also obtained: whether data analysis procedures were described in the methods part of the research report, whether there was any reference to the statistical methods used, and whether the statistical software used in the research was named. The sample size of each study was also studied, for the analyses the sample size was dichotomised with 100 subjects as a cut-point (median = 116.5).

Also other study characteristics variables were investigated. The country of the correspondence address and number of the authors were also studied to receive a reference difference between citation counts. The country of the corresponding author was classified: North Americans and others. The impact of the number of authors on citations was also investigated. The number of authors was dichotomised, with median as a cut-point (five or less vs. six or more authors).

The number of citations as such was the main outcome variable when analysing the influence of different variables on the number of citations. The effect of the journal was also taken into account, and to do this the median number of citations in each journal was used as the expected number of citations.

### ***4.3.2 Validity study of TPQ and TCI (III)***

For the validity study of the TPQ and TCI information about the level of education of the subjects was collected in 1997 from the National Educational Registry of Statistics, Finland. The categories used were basic ( $\leq 9$  years), upper secondary (10–12 years) and tertiary ( $\geq 13$  years). The place of residence in 1997 was also recorded: Northern Finland (Oulu and Lapland region) or the capital area around Helsinki. The results were also presented by gender. The TPQ/TCI items were analysed using the subscales developed by Cloninger and colleagues (Cloninger *et al.* 1993, 1994). Names of the subscales and number of variables in the subscales can be seen in Table 9 in the results section.

### 4.3.3 Validity study of TAS-20 (IV)

Marital status and vocational education were queried in the same questionnaire as the TAS-20. For the analyses marital status was dichotomised as never married/ever married + cohabiting. Vocational education was divided into four classes: no professional training, vocational school, institutional level, and college or higher education. The factor model was tested in two categories of vocational education (no professional school versus professional school/higher). Level of education used in the original study III was not used as it was not available when original article IV was done. Results were presented also by gender. In the student sample distributions of age and sex were reported.

Statistical analyses were performed for single TAS-20 items. The questions of the items are presented in Table 2. Questions 4, 5, 10, 18, and 19 were reversed for the analyses.

*Table 2. Original questions of Toronto Alexithymia Scale – 20 item version (TAS-20)*

Item	Question
1	I am often confused about what emotion I am feeling
2	It is difficult for me to find the right words for my feelings
3	I have physical sensations that even doctors don't understand
4	I am able to describe my feelings easily
5	I prefer to analyze problems rather than just describe them
6	When I am upset, I don't know if I am sad, frightened, or angry
7	I am often puzzled by sensations in my body
8	I prefer to just let things happen rather than to understand why they turn out that way
9	I have feelings that I can't quite identify
10	Being in touch with emotions is essential
11	I find it hard to describe my feelings more
12	People tell me to describe my feelings more
13	I don't know what's going on inside me
14	I often don't know why I am angry
15	I prefer talking to people about their daily activities rather than their feelings
16	I prefer to watch "light" entertainment shows rather than psychological dramas
17	It is difficult for me to reveal my innermost feelings, even to close friends
18	I can feel close to someone, even in moments of silence
19	I find examination of my feelings useful in solving personal problems
20	Looking for hidden meanings in movies or plays distracts from their enjoyment

## 4.4 Statistical methods

### 4.4.1 *Bibliometric studies (I, II)*

When comparing the use of different methods percentages were presented for the four journals. The distribution of received citation was extremely skewed to the right, so median was an appropriate statistic to summarise the citations. The median number of citations and the median of ratios (number of citations/journal median citation) by the use of statistical procedures and by the characteristics of the reported study in each psychiatric subfield and in the total sample were reported. The ratio was used to adjust for the effect of journal on number of citations. To test the null hypothesis of no differences in the citations between different study characteristics, the statistical significances were evaluated by the Mann-Whitney test (two groups) or by the Kruskal-Wallis test (more than two groups) (Altman 1991a). The results on received citations (original study II) are presented also separately for different psychiatric subfields.

### 4.4.2 *Validity study of TPQ and TCI (III)*

The skewness and kurtosis for all temperament scales and subscales was assessed to check the appropriateness of the data for factor analysis. Skewness is a measure of asymmetry and kurtosis is a measure of flatness or peakedness of the distribution of variables (Altman 1991a). Criteria for skewness and kurtosis for the subscales were the range (-1, 1). These were the criteria also in the only study (Chen *et al.* 2002) on Cloninger's scales that reported results on skewness and kurtosis of the subscales. An exploratory factor analysis with principal component analysis (promax rotation and with an unlimited number of factors, eigenvalues over 1 as a criterion) was made to test the replicability of the factor models for the TPQ/TCI subscales. Results were reported using the term factor, instead of the statistically more appropriate term principal component. This has been the case also in all the previous studies.

As most of the previous studies have found out that RD is psychometrically weaker and may consist of two factors, exploratory factor analysis was used instead of confirmatory factor analysis. Since the individual items are 0/1 variables and skewed, the presented factor analyses were done for subscales, as has also been done in most previous studies. The selection of different options and presentation of test statistics in factor analysis was based on the validity studies of the original instruments. This was done to enable comparisons of current results to the results of the previous studies.

Cronbach's alpha coefficients were used to test the internal consistency of the scales (Cronbach 1951). Cronbach's alpha measures how well a set of items measures a single unidimensional latent construct. Landis & Koch (1977) gave some benchmarks for interpretation of Cronbach's alphas: below 0.50 poor, 0.51–0.60 slight, 0.61–0.70 fair, 0.71–0.80 moderate, 0.81–0.90 substantial and 0.91–1.00 almost perfect.

Normative data were presented in forms of mean values and standard deviations. Analysis of covariance (Altman 1991a, Norušis 1993) was used to test differences between genders with control for education. TPQ and TCI dimensions were included as dependent and gender and education as independent variables in the analyses. Population distributions for the absolute scores of the higher-order temperament scales were also presented. Relationships between the higher-order temperament dimensions were tested with Pearson product moment correlations (Altman 1991a, Norušis 1993). All tests were two-tailed. Results of factor analyses were presented only for the total sample, but the other results were presented also for males and females separately. All the results were presented separately for the TPQ scale and for the temperament dimensions of the TCI.

#### **4.4.3 Validity study of TAS-20 (IV)**

According to the model developed by Bagby *et al.* (1994a), each of the TAS-20 items was considered to be a measure of only a single factor in the three-factor model. An exploratory factor analysis (principal component analysis, with varimax rotation and with an unlimited number of factors, eigenvalues greater than 1 as a criterion) and a confirmatory factor analysis (maximum likelihood estimation, with an oblique method) were done to test the replicability of the three-factor model for the TAS-20. As done by Parker *et al.* (1993) and following the recommendation of Cole (Cole 1987), the goodness-of-fit was evaluated by using four criteria: chi-square goodness-of-fit (a non-significant chi score), the goodness-of-fit index (GFI, score  $\geq 0.85$ ), the adjusted goodness-of-fit index (AGFI, score  $\geq 0.80$ ), and the root-mean-square residuals (RMS, score  $\leq 0.10$ ). The Cronbach's alpha was also used for testing the internal consistency of the whole TAS-20 scale and the three subscales (Cronbach 1951). All analyses were made for the total sample and also for men and women separately. The three-factor solution with a two-factor (Loas *et al.* 1996) and a one-factor model (Bagby *et al.* 1994a) were compared by using Consisted Akaike's Information Criteria (CAIC). A low CAIC value indicates the better model (Arbuckle 1997). The statistical analyses were made with SPSS (SPSS Inc. 2001) and Amos software (Arbuckle 1997).

### **4.5 Ethical considerations and personal involvement**

The 31-year follow-up survey of NFBC 1966 was approved by the Ethical Committee of Faculty of Medicine, Oulu University on 17 June 1996. After complete description of the study to the subjects, written informed consent was obtained.

The author of this thesis has evaluated all the articles (N = 448) used in the bibliometric part of this thesis and entered the data into computer file. The author has participated in designing and reporting of both the bibliometric articles (I, II).



The author has participated in the Northern Finland 1966 Birth Cohort study as a researcher since 1998. The author has participated in designing and reporting of both the factor analytic studies (III, IV).

All the statistical analyses, presentations of methods, and presentations of results in the original articles and this summary have been done by the author.

## 5 Results

### 5.1 Statistical methodology in psychiatric journals (I)

The general characteristics of the analysed articles are presented in Table 3. The journals differed from each other: the journals of national psychiatric associations (AJP, BJP and NJP) published more articles reporting studies based on cross-sectional or longitudinal design. When compared to the other journals, a large proportion (25%) of the articles published in AGP had an experimental design.

A clear majority of all articles applied some multivariate or specific methods. Statistical significance tests (p-values) were reported in 92% of all articles. Confidence intervals were reported in 29% of all the articles. Confidence intervals were less frequently used in AJP and NJP, while they played a more visible role in the other journals. Tables were included in 92% and figures in 43% of all the articles. There was some variation between the journals, especially regarding frequency of figures.

The evaluated journals had a different emphasis in the documentation of statistical methods and software used. A description of data analysis procedures in the methods part of the report was more frequently presented in AGP (92%) and AJP (79%) than in BJP (54%) or NJP (46%). Also the proportion of articles that included reference to statistical literature was lowest in BJP and NJP. Of all reviewed articles, 29% reported the selection and use of software.

The frequency and the percentage of different statistical procedures used in the four journals are presented in Table 4. The most commonly reported procedure was comparison of means in its various forms, with 60% of the articles comparing mean values. Non-parametric techniques were seldom mentioned. There were some differences between the journals in the application of methods. Standard methods of continuous variables (e.g. comparison of means and multiple comparisons) were more characteristic tools in AJP and AGP articles. Non-parametric tests were especially popular in NJP, where 25% of the articles used it as an analysing method. Computer dependent multivariate methods were frequently used, although these techniques had not replaced the traditional significance tests in data analyses. In particular, the application of statistical regression techniques was reported in more than one third of all articles.

Specific epidemiological statistics and methods were used in 50% of the papers. For instance, odds ratio and survival analysis were frequently applied.

*Table 3. Study design and general characteristics of statistical procedures in articles of four general psychiatric journals in 1996.*

Study design	American Journal of Psychiatry (n = 135)		Archives of General Psychiatry (n = 89)		British Journal of Psychiatry (n = 185)		Nordic Journal of Psychiatry (n = 39)		Total (n = 448)	
	n	%	n	%	n	%	n	%	n	%
Experimental	17	12.6	23	25.8	19	10.3	5	12.8	64	14.3
Cross-sectional	84	62.2	48	53.9	122	65.9	27	69.2	281	62.7
Longitudinal	34	25.2	18	20.2	44	23.8	7	17.9	103	23.0
Basic methods only	33	24.4	25	28.1	58	31.4	18	46.2	133	29.7
Multivariate or specific methods	100	74.1	62	69.7	115	62.2	16	41.0	294	65.6
P-values	129	95.6	86	96.6	167	90.3	32	82.1	414	92.4
Confidence intervals	25	18.5	30	33.7	67	36.2	8	20.5	130	29.0
Statistical tables	123	91.1	84	94.4	168	90.8	36	92.3	411	91.7
Statistical figures	59	43.7	52	58.4	65	35.1	15	38.5	191	42.6
Extended description of procedures	107	79.3	82	92.1	99	53.5	18	46.2	306	68.3
Reference to statistical literature	56	41.5	41	46.1	39	21.1	11	28.2	147	32.8
Software reported	30	22.2	30	33.7	60	32.4	8	20.5	128	28.6

*Table 4. Statistical methods in articles of four general psychiatric journals in 1996.*

	American Journal of Psychiatry (n = 135)		Archives of General Psychiatry (n = 89)		British Journal of Psychiatry (n = 185)		Nordic Journal of Psychiatry (n = 39)		Total (n = 448)	
	n	%	n	%	n	%	n	%	n	%
Basic methods	125	92.6	84	94.4	157	84.9	32	82.1	398	88.8
Comparison of means	103	76.3	63	70.8	89	48.1	12	30.8	267	59.6
t-test	83	61.5	48	53.9	71	38.4	7	17.9	209	46.7
analysis of variance	62	45.9	39	43.8	44	23.8	7	17.9	152	33.9
Cross-tabulation	72	53.3	62	69.7	95	51.4	19	48.7	248	55.4
Non-parametric test	12	8.9	16	18.0	36	19.5	10	25.6	74	16.5
Correlation	46	34.1	34	38.2	65	35.1	12	30.8	157	35.0
Multiple comparisons	31	23.0	29	32.6	15	8.1	4	10.3	79	17.6
Multivariate methods	72	53.3	41	46.1	66	35.7	10	25.6	189	42.2
Regression analysis	53	39.3	32	36.0	43	23.2	8	20.5	136	30.4
Manova	19	14.1	5	5.6	5	2.7	2	5.1	31	6.9
Factor analysis	3	2.2	4	4.4	17	9.2	0	0.0	24	5.4
Other multivariate methods	5	3.7	5	5.6	6	3.2	1	2.6	41	9.2
Specific methods	68	50.4	49	55.1	94	50.8	12	30.8	223	49.8

The sample of articles included 24 (5.4%) studies using factor analysis. These are summarised in Table 5. Sixteen of the articles used only exploratory factor analysis, six studies used only confirmatory factor analysis, while two of the studies applied both of these. The 18 articles using exploratory factor analysis did not always state clearly what methods were used. Principal component analysis was used in 10 of 12 articles reporting the method for extraction of the initial solution. Twelve articles reported the rotation method, 9 articles used orthogonal (varimax) rotation and 3 used oblique rotation. Most of the articles were published in BJP, whereas NJP did not publish any articles using factor analysis in 1996. In almost all of the articles factor analysis was applied in questionnaires and other instruments. There were only two exceptions (Al-Mousawi *et al.* 1996, Sackeim *et al.* 1996). Four of the articles studied the structure of psychotic symptoms (Cardno *et al.* 1996, Lenzenweger & Dorkin 1996, Maziade *et al.* 1996, Vazquez-Barquero *et al.* 1996).

The number of studies using factor analysis is generally large, a PsycINFO database article search with the keyword “factor analysis” gave 805 references for the year 1996. When no time limit was used the search gave 20930 articles in January 2004.

Table 5. Studies using factor analysis in four general psychiatric journals in 1996.

Reference	Journal*	Methods**	Application***
Alexopoulos <i>et al.</i> (1996)	AJP	EFA	Structure of the <i>Hamilton Depression Scale (HAM-D)</i>
Al-Mousawi <i>et al.</i> (1996)	BJP	EFA	Factorisation of regional metabolic values in schizophrenia and mania
Baillie & Mattick (1996)	BJP	EFA	Structure of the <i>Benzodiazepine Dependence Questionnaire (BDEPQ)</i>
Barry & Crosby (1996)	BJP	EFA	Calculate factor scores for the <i>Brief Psychiatric Rating Scale (BPRS)</i>
Cardno <i>et al.</i> (1996)	BJP	EFA	Structure of psychotic symptoms ( <i>OPCRIT</i> )
Chen <i>et al.</i> (1996)	BJP	EFA	Structure of the <i>High Royds Evaluation of Negativity Scale (HEN)</i>
Cox <i>et al.</i> (1996)	BJP	CFA	Structure of the <i>Fear Questionnaire</i> with social phobia patients
Harvey <i>et al.</i> (1996)	BJP	EFA	Structure of the <i>MRC Social Behaviour Schedule (SBS)</i> and of the <i>Manchester Scale</i>
Horowitz <i>et al.</i> (1996)	AJP	EFA	Structure of the <i>Self-Regard Questionnaire</i>
Kendler (1996)	AJP	EFA/CFA	Structure of the <i>Parental Bonding Instrument</i>
Kendler <i>et al.</i> (1996)	AGP	CFA	Validation of <i>DSM-III-R</i> depressive syndromes
Laor <i>et al.</i> (1996)	AGP	EFA	Children's stress symptoms ( <i>CBCL</i> and <i>PCASS</i> ) after traumatic events
Lenzenweger & Dworkin (1996)	BJP	CFA	Dimensions of schizophrenia (manual of psychotic symptoms developed by the authors, <i>PASPSA</i> and <i>ZPSCS</i> )
Madden <i>et al.</i> (1996)	AGP	EFA/CFA	Seasonal changes in symptoms of mood and behaviour ( <i>SPAQ</i> )
Maziade <i>et al.</i> (1996)	BJP	EFA	Stability of symptom dimensions ( <i>DSM-III-R</i> ) in early-onset schizophrenia
Nakagawa <i>et al.</i> (1996)	BJP	CFA	Structure of two measures of obsessive-compulsive disorder ( <i>CC</i> and <i>YBOCS</i> )
Naughton <i>et al.</i> (1996)	BJP	CFA	Structure of the <i>Adolescent to Adult Personality Functioning Assessment (ADAPFA)</i>
Romans <i>et al.</i> (1996)	BJP	EFA	Structure of a <i>self-esteem scale</i> in a study of childhood sexual abuse
Sackeim <i>et al.</i> (1996)	AGP	EFA	EEG profile changes after electroconvulsive therapy
Stevenson <i>et al.</i> (1996)	BJP	CFA	Children's mental health variables ( <i>BCL</i> , <i>EAS</i> , <i>GHQ</i> and <i>W-W-P</i> )
Thomas <i>et al.</i> (1996)	BJP	EFA	Structure of the <i>Brief Syntactic Analysis (BSA)</i>
Vazquez-Barquero <i>et al.</i> (1996)	BJP	EFA	Symptom structure in first episode schizophrenia patients ( <i>SANS</i> and <i>SAPS</i> )
Wolff <i>et al.</i> (1996a)	BJP	EFA	Structure of the <i>Community Attitudes toward the Mentally Ill (CAMI)</i>
Wolff <i>et al.</i> (1996b)	BJP	EFA	Use of factors of the <i>Community Attitudes toward the Mentally Ill (CAMI)</i>

\* AGP = Archives of General Psychiatry, AJP = American Journal of Psychiatry, BJP = British Journal of Psychiatry. There were no factor analytic articles in Nordic Journal of Psychiatry in 1996.

\*\* CFA = Confirmatory Factor Analysis, EFA = Exploratory Factor Analysis

\*\*\* BCL = Behaviour Checklist, CBCL = Child Behaviour Checklist, CC = Compulsion Checklist, DSM-III-R = Diagnostic and Statistical Manual of Mental Disorders, third edition, revisited, EAS = Emotionality, Activity and Sociability temperament questionnaire, EEG = Electroencephalogram, GHQ = General Health Questionnaire, MRC = Medical Research Council, OPCRIT = Operational Criteria Checklist for Psychotic Illness, PASPSA = Phillips Abbreviated Scale of Premorbid Sexual Adjustment, PCASS = Preschool Children's Assessment of Stress Scale, SAPS = Scale for the Assessment of Positive Symptoms, SANS = Scale for the Assessment of Negative Symptoms, SPAQ = Seasonal Pattern Assessment Questionnaire, W-W-P = Weiss-Werry-Peters Activity Scale, YBOCS = Yale-Brown Obsessive-Compulsive Scale, ZPSCS = Zigler-Phillips Social Competence Scale.

## 5.2 Effect of statistical methods on citations (II)

The median of received citations was 23.5 in the total sample of 448 articles. In different subfields the respective values were as follows: epidemiology 21, clinical topics 22, psychopharmacology 32, biological topics 37 and others 15. This difference between various psychiatric subfields was statistically significant (Kruskal-Wallis test,  $p < 0.001$ ). Medians of citations up to November 2002 for different psychiatric subfields by the journals reviewed are presented in Table 6. There were statistically significant (Kruskal-Wallis test,  $p < 0.001$ ) differences between journals, articles published in AGP receiving the most citations (median 54).

The medians of citations and ratios (number of citations/journal median citation) by the use of statistical procedures in the total sample and by psychiatric subfields are presented in Table 7. In the total sample articles using basic methods received more citations (Mann-Whitney test,  $p = 0.003$ ) compared to empirical research reports, which did not apply any of the basic statistical analysis tools. The use of multivariate and specific methods did not increase the number of citations. Also the use of statistical figures (Mann-Whitney test,  $p < 0.001$ ), reference to statistical literature (Mann-Whitney test,  $p = 0.007$ ) and extended description of statistical procedures (Mann-Whitney test,  $p < 0.001$ ) had a statistically significant positive effect on the number of citations. The trends were similar in all psychiatric subfields, except for the fact that reference to statistical literature had a more positive impact on the number of citations in clinical, psychopharmacological and biological studies than in epidemiological studies.

When the effect of journal was taken into account (using the ratio observed/expected citations) the difference remained statistically significant in the extended description of procedures (Mann-Whitney test,  $p = 0.02$ ). The other differences in the total sample were no longer statistically significant.

Citations by psychiatric subfields are presented in Table 8 for study design, data size and author characteristics. The study design was statistically significantly associated to the number of citations. Especially in clinical studies, studies with experimental design were most often cited in further studies. The sample size did not affect the citations.

The country of correspondence and the number of authors were related to the citation frequency. Authors outside the USA and Canada had a decreased probability of receiving citations. North American corresponding authors (median 36) received more citations than others (median 14). The difference between more than five (median 33) and five or less (median 18) authors was also large. After taking the effect of the journal (median citations of the journal) into account, more than five authors still had a statistically significant positive effect on the citations. These differences were greater than the effect of reported statistical techniques on the number of received citations.



Table 6. The number of original research articles and median (range) of citations (up to November 2002) in different subfields of psychiatry in four general psychiatric journals in 1996.

	Epidemiology		Clinical topics		Psychopharmacolog		Biological topics		Others		Total	
	n	Md (range)	n	Md (range)	n	Md (range)	n	Md (range)	n	Md (range)	n	Md (range)
American Journal of Psychiatry	21	27 (12-111)	51	30 (7-140)	14	35.5 (3-114)	25	52 (11-131)	24	25 (1-99)	135	30 (1-140)
Archives of General Psychiatry	20	57.5 (8-178)	19	38 (9-161)	13	64 (12-136)	29	57 (6-225)	8	51 (15-128)	89	54 (6-225)
British Journal of Psychiatry	40	18 (2-92)	40	17.5 (3-77)	17	18 (3-103)	20	13 (4-65)	68	13.5 (1-48)	185	17 (1-103)
Nordic Journal of Psychiatry	8	2 (0-11)	15	1 (0-12)	5	2 (0-5)	2	1.5 (1-2)	9	2 (0-4)	39	2 (0-12)
Total	89	21 (0-178)	125	22 (0-161)	49	32 (0-136)	76	37 (1-225)	109	15 (0-128)	448	23.5 (0-225)

Table 7. Use of statistical methods and median of the citations (up to November 2002) of the original research articles of four general psychiatric journals<sup>a</sup> in 1996 by psychiatric subfields.

	Epidemiology (n = 89)			Psychopharmacology (n = 49)			Biological topics (n = 76)			Others (n = 109)			Total (n = 448)					
	n	Md	Ratio	n	Md	Ratio	n	Md	Ratio	n	Md	Ratio	n	Md	Ratio			
Basic methods																		
No	13	15	0.57	8	15	0.97	5	5	1.00	4	55	1.51	20	17	0.88	50	16	0.94
Yes	76	23	1.03	117	23	0.90	44	36	1.12	72	36	1.13	89	15	0.83	398	25	1.00
Multivariate methods																		
No	42	22.5	1.03	76	22	1.00	30	39	1.18	44	35.5	1.17	67	13	0.82	259	22	1.00
Yes	47	20	1.00	49	24	0.87	19	26	0.97	32	45.5	1.05	42	18	0.92	189	24	0.97
Specific methods																		
No	29	18	0.90	62	23.5	1.03	33	40	1.22	47	31	1.00	54	14.5	0.88	225	23	1.00
Yes	60	23.5	1.09	63	22	0.87	16	27	0.95	29	46	1.19	55	16	0.82	223	25	0.98
Confidence intervals																		
No	45	20	1.00	97	22	0.94	38	36.5	1.18	65	38	1.20	73	14	0.82	318	22	1.00
Yes	44	27.5	1.09	28	26	0.95	11	29	1.02	11	33	1.00	36	19	0.87	130	26	1.00
Statistical figures																		
No	63	20	1.00	68	18.5	1.00	17	18	1.06	38	46.5	1.17	71	15	0.82	257	20	1.00
Yes	26	23	1.06	57	32	0.90	32	39.5	1.34	38	34	1.00	38	15.5	0.97	191	28	1.00
Reference to statistical literature																		
No	61	21	1.00	80	20	1.00	32	30	1.06	56	34	1.04	72	13	0.82	301	21	1.00
Yes	28	20.5	0.88	45	30	0.87	17	39	1.30	20	56	1.60	37	18	0.97	147	29	1.00
Software reported																		
No	57	24	0.94	96	23.5	0.99	36	33	1.06	45	31	1.02	86	14.5	0.81	320	22.5	1.00
Yes	32	19.5	1.06	29	18	0.65	13	26	1.17	31	54	1.20	23	20	0.97	128	25	1.05
Extended description of procedures																		
No	29	15	0.94	37	17	0.82	13	12	1.06	14	28	1.13	49	12	0.76	142	15	0.88
Yes	60	27	1.03	88	27	0.99	36	39.5	1.18	62	45.5	1.12	60	18.5	0.92	306	29	1.00

Ratio = median of the ratios (number of citations/ journal citation median). Statistical differences are tested with Mann-Whitney U test,  $p < 0.05$  are in italics.  
<sup>a</sup>the American Journal of Psychiatry, the Archives of General Psychiatry, the British Journal of Psychiatry and the Nordic Journal of Psychiatry.

Table 8. Median of the citations (up to November 2002) of the original research articles of the American Journal of Psychiatry, the Archives of General Psychiatry, the British Journal of Psychiatry and the Nordic Journal of Psychiatry in 1996 by psychiatric sub-fields

	Epidemiology (n = 89)			Epidemiology (n = 89)			Psychopharmacology (n = 49)			Biological topics (n = 76)			Others (n = 109)			Total (n = 448)		
	n	Md	Ratio	n	Md	Ratio	n	Md	Ratio	n	Md	Ratio	n	Md	Ratio	n	Md	Ratio
Study design																		
experiment	2	6	0.38	18	42.5	1.57	34	43	1.18	4	60.5	1.12	6	23.5	0.78	64	40	1.15
cross-sectional	66	19.5	0.97	71	20	1.00	7	20	1.30	58	43.5	1.19	79	15	0.82	281	22	1.00
longitudinal	21	26	1.18	36	20	0.77	8	18	0.93	14	20.5	0.56	24	16	1.03	103	20	0.87
Data size																		
99 or less	20	15.5	1.00	69	20	0.87	30	28.5	1.11	48	37	1.18	32	11	0.79	199	22	1.00
100 or more	69	27	1.00	56	26	1.00	19	34	1.07	28	40.5	1.04	77	16	0.87	249	24	1.00
International co-operation																		
No	68	21	1.00	110	24.5	1.00	40	33	1.12	60	43.5	1.19	96	15	0.81	374	24	1.00
Yes	21	20	0.78	15	17	0.60	9	18	1.06	16	34	0.95	13	18	1.12	74	19	1.00
Country of the correspondence																		
USA or Canada	42	31.5	1.03	67	33	1.00	24	53	1.20	47	59	1.37	30	25	0.74	210	36	1.05
Other	47	18	1.00	58	12.5	0.85	25	18	1.06	29	21	0.97	79	13	0.88	238	14	0.95
Number of authors																		
1-5	59	20	0.88	79	18	0.87	25	18	1.06	31	29	0.97	85	15	0.83	279	18	0.90
6 or more	30	27	1.15	46	32	1.07	24	48	1.26	45	52	1.37	24	18.5	0.87	169	33	1.11

Ratio = median of the ratios (number of citations/ journal citation median). Statistical differences are tested with Mann-Whitney U test, except in study design with Kruskal-Wallis test,  $p < 0.05$  are in italics.

### 5.3 Validity of TPQ and TCI (III)

The skewness and kurtosis for all high-order temperament scales of both the TPQ and TCI fell in the range (-1, 1) in the total sample and also in males and females. The subscales were in the same range, the only exception being the attachment (RD3) subscale of the TCI, which was negatively skewed (-1.007) among females.

In the case of TPQ the exploratory factor analysis gave a four-factor model, which accounted for 61.8% of the total variance. Of the three higher-order dimensions RD was the only one performing poorly, having high loadings in two factors. Results in the TCI were quite similar; the four-factor solution accounted for 61.4% of the total variance. Results of the exploratory factor analyses, with factor loadings, communalities, eigenvalues and percents of the total variance explained are presented in Table 9. Factor analyses were done also separately for males and females, the resulted factor loadings were comparable with those of the whole sample (results not shown).

The distributions of the absolute scores of the higher-order temperament dimensions of the TPQ and TCI are presented in the original study III (Figure 2). The scores of the persistence dimension of the TCI are not presented, as this dimension has a markedly lower number of items (8) compared to the other higher-order dimensions. Females had higher scores than males in all the higher-order temperament dimensions of the TPQ and TCI, except for the persistence dimension of the TCI. Mean values for the TCI higher-order dimensions were for males 19.7 for NS, 13.0 for HA, 13.2 for RD and 4.5 for P, and for females 20.8 for NS, 14.9 for HA, 16.0 for RD and 4.1 for P. The difference in means was also statistically significant in most of the subscales. The Finnish normative data for the TPQ and TCI and statistical comparisons between genders are presented in detail in Tables 2 and 3, respectively, of the original study III.

The internal consistency of each subscale was assessed. Cronbach's alphas for the three higher-order dimensions of the TPQ were 0.74 for NS, 0.85 for HA and 0.61 for RD. The highest alpha value in the TPQ subscales was 0.69 for the shyness subscale (HA3), whereas the lowest alpha (0.41) was in the disorderliness subscale (NS4). Alphas for the four higher-order temperament dimensions (NS, HA, RD, P) of the TCI were 0.78, 0.85, 0.69, and 0.55, respectively. The highest alpha value in the TCI subscales was 0.74 of the extravagance subscale (NS3) and the lowest alpha (0.42) was in the disorderliness subscale (NS4). The number of items and Cronbach's alphas for the TPQ and TCI scales are also presented in tables 2 and 3 of the original study III.

Pearson product moment correlation coefficients for the relationships among the temperament dimensions in the TPQ and TCI are presented for the total sample and for males and females separately in the original study III (Table 4). The correlation between NS and RD was slightly positive, and between HA and both NS and RD slightly negative.

Table 9. Parameter estimates of the exploratory factor analysis <sup>a</sup> of the Tridimensional Personality Questionnaire (TPQ) and of the temperament dimensions of the Temperament and Character Inventory (TCI) in the 31-year follow-up study of the Northern Finland 1966 Birth Cohort (NFBFC 1966; N = 4349)

	TPQ				Temperament dimensions of TCI					
	F1	F2	F3	F4	Communalities	F1	F2	F3	F4	Communalities
% of variance explained <sup>b</sup>	24.4	16.2	12.0	9.2		25.1	15.0	12.1	9.2	
Eigenvalues	2.926	1.944	1.446	1.101		3.014	1.800	1.448	1.103	
Promax rotated solution										
Novelty seeking (NS)					0.49	-0.48	0.25	0.36	0.16	0.58
Explorative excitability (NS1)	-0.53	0.27	0.13	0.14	0.49	-0.48	0.25	0.36	0.16	0.58
Impulsiveness (NS2)	0.10	0.80	-0.03	0.01	0.63	-0.03	0.82	-0.18	-0.01	0.61
Extravagance (NS3)	-0.16	0.48	0.38	0.06	0.52	0.00	0.60	0.28	0.03	0.53
Disorderliness (NS4)	0.04	0.83	-0.23	0.01	0.66	-0.05	0.77	-0.27	0.02	0.57
Harm avoidance (HA)										
Worry/pessimism (HA1)	0.77	0.10	0.10	0.15	0.57	0.79	0.05	-0.06	0.24	0.59
Fear of uncertainty (HA2)	0.79	-0.11	0.13	0.21	0.64	0.78	-0.15	0.11	0.17	0.62
Shyness (HA3)	0.79	0.05	-0.12	0.00	0.63	0.76	-0.03	-0.19	0.04	0.62
Fatigability (HA4)	0.71	0.16	0.22	-0.06	0.59	0.75	0.13	0.07	-0.10	0.62
Reward dependence (RD)										
Sentimentality (RD1)	0.27	0.07	-0.16	0.91	0.76	0.37	0.11	0.16	0.84	0.75
Persistence (RD2/P)	-0.24	-0.04	-0.70	0.38	0.67	-0.35	-0.21	-0.32	0.59	0.68
Attachment (RD3)	-0.21	-0.05	0.33	0.60	0.63	-0.21	-0.05	0.72	0.22	0.64
Dependence (RD4)	0.07	-0.19	0.76	0.18	0.63	0.09	-0.25	0.76	-0.14	0.56

<sup>a</sup> Extraction Method: Principal Component Analysis. Rotation Method: Promax with Kaiser Normalization, with eigenvalues > 1 <sup>b</sup> In total 61.8 (TPQ) and 61.4 (TCI).

#### 5.4 Validity of TAS-20 (IV)

Exploratory factor analysis of TAS-20 in the cohort sample gave a four-factor model, which accounted for 49.8% of the total variance. Of the three originally suggested factors, items from factors 1 and 2 formed their own factors, whereas items from factor 3 formed two separate factors. When a three factor-solution was carried out with exploratory factor analysis, each item had a factor loading  $> 0.40$  on their suggested factors.

The parameter estimates of the confirmatory factor analysis for the items and for the relationships between factors are presented among all cohort subjects in Figure 2. The parameter estimates for the items of the TAS-20 and the model statistics from confirmatory factor analysis and Cronbach's alpha values are presented by gender in the original study IV (Table 1 and Table 2, respectively). For all subjects the chi-square goodness-of-fit (chi-square = 3281.30,  $p < 0.001$ ) was significant; and the GFI (0.935), the AGFI (0.918) and the RMS (0.061) were acceptable. The Cronbach's alpha for the total TAS-20 scale was 0.83, and for the three subscales (factors 1, 2, and 3), 0.81, 0.77, and 0.66, respectively. None of the alphas of the three factors improved when an item was deleted. The CAIC for the three-factor solution was 3690.8 (men 1917.5; women 2187.6); for the two-factor solution, 4978.8 (men 2344.8; women 2564.8); and for the one-factor solution, 8945.5 (men 3884.9; women 4279.9) in the total sample. The three-factor model also worked well (GFI, AGFI, and RMS were acceptable) when different subgroups on the basis of marital status (never married versus ever married/cohabitation) and educational level (no professional school versus professional school/higher) were analysed separately.

In the student sample the chi-square goodness-of-fit was significant (chi-square = 517.97,  $p < 0.001$ ). The GFI was 0.900, the AGFI 0.874, and RMS 0.064. All parameter estimates for the items of the TAS-20 were significant at the 0.05 level.

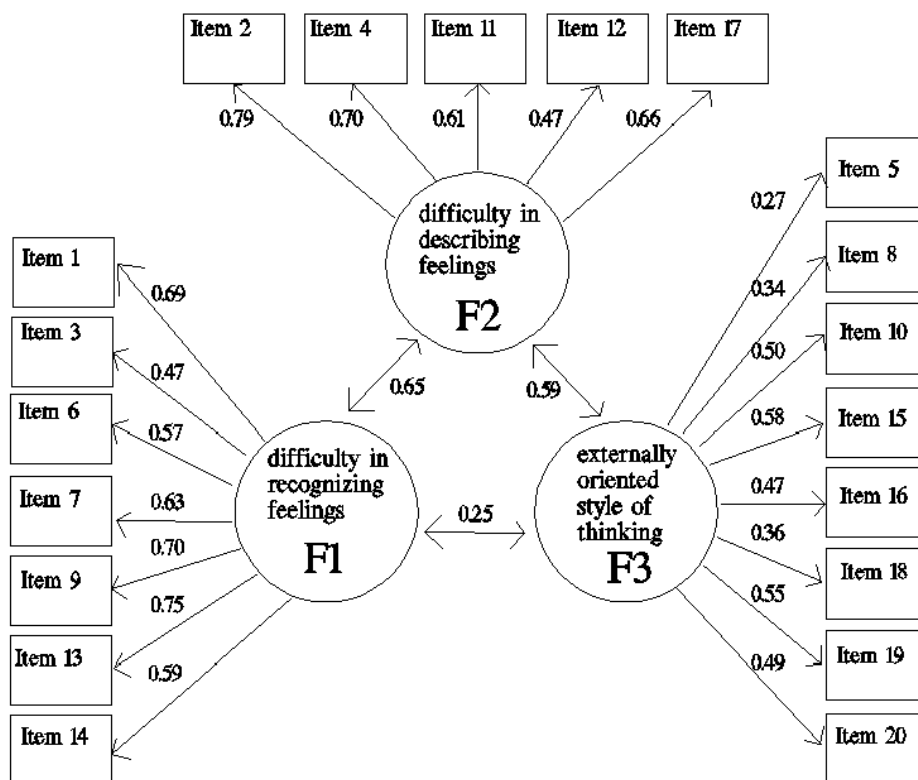


Fig. 2. The parameter estimates (item loadings to factors) for Toronto Alexithymia Scale – 20 (TAS-20) and correlations between factors in confirmatory factor analysis in the Northern Finland 1966 Birth Cohort Study (N = 5034). Item questions are presented in Table 1.

## **6 Discussion**

### **6.1 Main findings**

The main study findings corresponding to the presented aims are:

1. Tables, figures and various statistical methods are commonly used in psychiatric articles (I).
2. There are differences in the use of statistical methods between general psychiatric journals (I).
3. Use of different statistical methods had a low effect on the number of received citations, but extended description of methods had a positive effect on the number of received citations (II).
4. The Temperament and Character Inventory (TCI) has better psychometric properties than the Tridimensional Personality Questionnaire (TPQ). The scales can be used also in Finland, although some development could be done in the reward dependence dimension (III).
5. The Finnish version of the 20-item Toronto Alexithymia Scale (TAS-20) can be used and has a similar structure as the original version (IV).

### **6.2 Discussion of results**

#### ***6.2.1 Statistical methodology in psychiatric journals (I)***

The analysis of the bibliometric data shows that traditional statistical significance testing is still widely used in psychiatric data analysis, but there appears to be a shift towards the use of confidence intervals when compared to earlier findings (Hokanson *et al.* 1986, Everitt & Landau 1998). The results implicate also that the use of sophisticated statistical modelling and multivariate regression analysis has increased in major psychiatric



journals, psychiatric journals have different profiles in their statistical content, and statistical procedures are usually extensively described.

The most frequent statistical problems found in medical articles are misinterpretation of p-values or non-significant results and confusion between statistical and clinical significance (Jamart 1992, Walter 1995). This study shows that despite these well-known weaknesses, significance testing is still a widely used approach also in analysing psychiatric data. A possible explanation for this is that the inclusion of p-values convinces others (including some editors and referees of psychological and medical journals) of the report's "scientificity", validity and credibility.

Recently there has been a shift towards the use of confidence interval estimates (Altman 1991b, Walter 1995), especially in several prestigious medical journals (Bulpitt 1987, Sterne & Smith 2001). Some major biomedical journals now require authors to use confidence interval estimates instead of p-values when data warrant this approach (Walter 1995). In the reports of this sample, confidence interval estimates were still rarely seen compared to p-values in all four journals.

In some of the articles the authors did not clearly specify which research strategies or statistical methods were used in the data analysis process. Müllner *et al.* (2002) have noted a similar lack in reporting, when they studied how statistical methods to adjust for confounding are described in medical journals. An extended description of procedures was most often lacking from articles in BJP and NJP. Of all the articles in this sample 128 (29%) made reference to statistical software. It is easier for a reader to judge the methodological aspects of a report and to repeat the statistical analyses when the author indicates the software used to run statistical computations.

Some authors (Altman *et al.* 2000, Cummings & Rivara 2003) offer suggestions for the presentation of statistical information to the readers of general medical journals. The general principles include first a clear description of design and data collection of the study. In the methods section all the statistical procedures used should be identified. Common techniques do not need to be described, but methods with more than one form should be identified unambiguously. More complex methods need some explanation and reference. The name of the statistical software should be given, at least when applying more complex methods. In the results section, adequate description of the data should precede and complement formal statistical methods. Altman *et al.* (2000) suggest reporting the observed values of the test statistics and not just the p-values, whereas Cummings & Rivara (2003) do not consider this necessary. Authors of both of these articles encourage e.g. presenting confidence intervals when appropriate and graphical displays of results. Finally Altman *et al.* (2000) recommend researchers to seek the advice of a statistician when possible.

Tables (92% of the articles) and figures (43%) were commonly used in the four general psychiatric journals. The use of illustrative tables and figures helps readers to comprehend the main finding(s) of the article. Figures need not only be diagrams or curves, more innovative figures can also be used.

Logistic models and Cox regression have been used with increasing frequency in health research literature in situations where previously other methods – a stratified analysis of probabilities or odds (Hosmer & Lemeshow 1989) and the life-table analysis (Streiner 1995), respectively – used to be the analysis of choice. One of the advantages of

these modelling adjustment techniques is the capacity to control the influence of many confounders simultaneously.

The number of articles applying factor analysis in AGP, AJP and BJP had increased compared to the study by Hokanson *et al.* (1986). They reported 10 factor analytic studies in 1983 and 7 in 1984 in these three journals, whereas this study found that in 1996 the corresponding number of the studies was 24.

This study revealed that the articles of the selected journals differ in their use of sophisticated multivariate methods. The psychiatric subfield of an article affects the choice and frequency of advanced statistical techniques. This study replicated the earlier findings that the general psychiatric journals have diverging scientific fields of interest (Pincus *et al.* 1993, Morlino *et al.* 1997). AJP and NJP had more clinical articles than the other two journals, BJP had more other articles (e.g. prevalence and validity studies) and AGP had more biological articles compared to other evaluated journals. The subfields of psychiatry emphasise research methods differently. Instructions for authors differ between journals. For example, restrictions as to length of articles affect description and presentation of methods and results. The use of specific referees for statistical issues may also affect the quality and presentation of statistical methods. The readership of a journal may affect the content of the articles. AJP, BJP and NJP are journals of national psychiatric associations and their readerships may include more clinicians than the readership of AGP. These differences between the selected journals may explain the finding that the journals have somewhat different policies concerning statistical methodology.

### ***6.2.2 Effect of statistical methods on citations (II)***

As far as the author knows, this was the first time when it was investigated whether or not there was any association between the statistical methods applied in the research report and the citations these reports received from other researchers. The choice of statistical procedures does not always seem to be important for readers to comprehend and retain information from a publication. Reason for this could be the fact that some characteristics of the study (e.g. country of the origin) have a very strong effect and that effect of some characteristics (e.g. methods) are thereby difficult to sort out. However, articles with graphics and clear documentation of research methods did receive more citations than articles without these characteristics.

The effect of statistical methodology on citations was unfortunately low compared to the effect of the country of the corresponding author or the number of authors. The articles were cited and acknowledged more often if they were published in the visible North American journals (AGP and AJP) and if the topic was psychopharmacological or biological. This study indicates also that in clinical studies experimental design is valued when articles to be cited are selected. In psychopharmacological studies a large number of authors may be due to a study being a multicentre trial. This design is likely to have a positive effect on the number of citations.

The use of basic statistical methods is appreciated when results from empirical research articles are utilised. If an article did not employ any of the basic statistical analysis tools, its citation frequency remained low. The association with other techniques was slight. For example, the use of multivariate methods did not affect citation performance. This result implies that the importance of statistics in the communication system of medical literature is vague.

According to the results, the usage of statistical figures in both clinical and psychopharmacological research reports is associated with higher citation counts. Using tables and graphics instead of or in addition to summary statistics or words is sometimes useful. Investigators will be drawn to a presentation because tables and graphics appear friendlier than numbers in text. An effective use of graphics can convey more information than is possible by any other means.

Sufficient description of the methods used, reference to statistical literature and notification of software were essential factors for article to be frequently cited. In order to benefit the most from the articles and to assess the appropriateness of the methods used, journal readers should be informed of the methods used to summarise raw data. A distinct section of statistical methods clarifies the applied methods and techniques to the reader and also enables the reader to apply methods (and cite the study) in his/her own research. As the only one of the four journals, AGP advised authors to use a distinct section for the description of statistical methods. In this sample, the documentation of statistical methods used was not always sufficient. Reference was made to statistical literature in only 33% of the articles. When a published study does not report clearly its research methods and findings, a reader attempting to evaluate its scientific validity may rely largely on the authors' reputation and writing style, or he/she may simply rely on the journal's reputation.

### **6.2.3 Validity of TPQ and TCI (III)**

Cronbach's alpha coefficients for the higher-order dimensions and for the subscales of the TPQ and TCI were close to those reported earlier in different samples and in various languages (Cloninger *et al.* 1991, Cloninger *et al.* 1993, Brändström *et al.* 1998, de la Rie *et al.* 1998, Herbst *et al.* 2000, Pélişolo & Lépine 2000, Gutiérrez *et al.* 2001, Sung *et al.* 2002). The persistence subscale (P) in the TCI had the lowest alpha value of the higher-order scales; one reason for this is that P has only 8 items, and alpha tends to increase with more items (Norusis 1993). The shyness subscale (HA3) has also previously been among the subscales with the highest alpha value (Cloninger *et al.* 1991, Cloninger *et al.* 1993, Brändström *et al.* 1998, de la Rie *et al.* 1998, Pélişolo & Lépine 2000). The Cronbach's alpha values indicated that the internal consistency of the TCI temperament subscales was generally better than that of the TPQ subscales, the only exceptions being the fatigability (HA4) and the dependence (RD4) subscales. When compared with the presented (Landis & Koch 1977) benchmarks of Cronbach's alphas, the values of this study can be characterised mostly as fair, moderate or substantial.

Exploratory analyses to evaluate the structural validity of the subscales in the temperament scales were performed. The finding that sentimentality (RD1) differs from other RD dimensions in the TCI replicates the finding of Gutiérrez *et al.* (2001) in a Spanish sample. Sentimentality (RD1) and persistence (P) were in the same factor also in the unpublished Finnish TCI study (Puttonen 1998). The finding that explorative excitability (NS1) has a strong negative loading in the HA factor was also replicated; this has been shown in at least two prior studies (Pélissolo & Lépine 2000, Gutiérrez *et al.* 2001).

The correlations between the higher-order scales were close to those reported in the US samples (Cloninger *et al.* 1991, 1993). There were no large gender differences in the correlations. The sample size is very large, so correlations between scales were mostly statistically significant, even though most of these correlations were below 0.30. The negative correlation between HA and RD was smaller in the TCI (-0.02) than in the TPQ (-0.20), supporting the original psychobiological model in which these were considered independent factors (Cloninger 1986, 1987).

Previous studies have shown that education has an impact on temperament scores, for example HA decreases with more education (Chen *et al.* 2002). Study participants were more educated than the non-participants, and this fact might influence the results obtained. Consequently, education was adopted as one parameter in the statistical tests for gender differences. However, the female subjects in this sample were more educated than males, but females still had statistically significantly higher scores in the HA subscale. The evident differences between males and females in different temperament dimensions can be seen clearly in the presented distribution lines (Figure 2 of the original study III). Also here, the large sample size works further in favour of statistically significant p-values.

The presented data offer possibilities to make inter-population comparisons of temperament scores using also Finnish normative data. When compared with the original US normative samples of the TPQ/TCI (Cloninger *et al.* 1991, 1993) the 31-year old Finnish cohort had higher mean scores in NS and HA, whereas the US sample had higher scores in RD and P. These differences may be explained by cultural differences between the two countries, and/or by differences in the composition of the samples within each country.

#### **6.2.4 Validity of TAS-20 (IV)**

It was possible to replicate the originally established three-factor model for the TAS-20 (Bagby *et al.* 1994a, 1994b). Three of four criteria of goodness-of-fit met the standards for adequacy-of-fit. Only the chi-square goodness-of-fit score was not adequate. This was, however, not surprising, because the sample was very large (Cole 1987, Marsh *et al.* 1988, Arbuckle 1997), and a similar result has been reported in other studies concerning psychometric properties of the TAS-20 (Bagby *et al.* 1994a, Parker *et al.* 1993, Pandey *et al.* 1996).

Factor 1 and factor 2 exhibited quite a strong correlation, which was as expected. This is consistent with the results of previous studies dealing with the TAS-20 (Parker *et al.* 1993, Pandey *et al.* 1996), as was also the significant correlation between factors 2 and 3. The internal consistency of the total score was good and sufficient also in the three factors. There were no sex differences, showing that the scale functions for men as well as for women. The three-factor solution performed better than the two- and one-factor solutions.

The model in confirmatory factor analysis was evaluated using several coefficients and criteria of goodness-of-fit. This study applied similar criteria than in the earlier applications of TAS-20, but also slightly different and stricter criteria have been presented (Arbuckle 1997, Nummenmaa *et al.* 1997, Metsämuuronen 2003).

### **6.3 Strengths and limitations of the study**

#### ***6.3.1 Bibliometric studies (I, II)***

##### *6.3.1.1 Strengths of the studies*

This sample of psychiatric articles was more extensive than in most of the previous studies as more than one journal was selected. The selected journals included major general psychiatric journals, so the selection was not restricted to a subfield of psychiatric research, as has been the case in most of the previous studies. Furthermore, the selected articles included all original quantitative studies published in 1996. Therefore, the results are more generalisable than those from the previous studies. The relative long follow-up of six years gives good estimates for the later utilisation of the articles in the form of received citations. As far as the author knows, this was also the first study where the effect of statistical methods to the number of citations was explored.

##### *6.3.1.2 Limitations of the studies*

There are many problems with the use of citation counts in psychiatry that stem from differences between researchers in searching and selecting articles for references and conventions between fields of science (Nieminen & Isohanni 1999, Adam 2002, West & McIlwaine 2002, Berghmans *et al.* 2003, Kurmis 2003). In the current study, neither frequent use of statistical methods nor the number of citations was thought to be direct indicator of quality of the article. Citation counts were used as one method of providing a quantitative expression (measure) of the utilisation, acceptance and visibility of a particular paper in international scientific literature.

There are great variations between the various branches of psychiatry as to the period of time during which the citations are accumulated. The use of citation analysis for the evaluation of different research subfields is further complicated by the differences in the citation practices within these fields (Nieminen & Isohanni 1997). The starting point of the citation process is the author of the citing text. One assumption underlying citation analysis is that the citing author selects the references in a rational manner, but the act of citing is essentially subjective. References are issued for various reasons and they have different functions in texts (Egghe & Rousseau 1990).

Only one rater (the author of this thesis) evaluated the articles. This may have resulted in more incorrect ratings than would have been the case with several raters, but on the other hand this guaranteed that the articles were rated similarly. Also whenever the author faced problems in assessing a statistical procedure in a specific paper, its classification was subsequently discussed with an experienced biostatistician, docent Pentti Nieminen. The classification of psychiatric subfields was especially difficult, as many of the articles included characteristics from several subfields. It would have been interesting to study the effect of statistician as an author, but it was not possible to evaluate this from the articles.

### ***6.3.2 Factor analytic studies (III, IV)***

#### *6.3.2.1 Strengths of the studies*

The sample sizes in both of the studies (III, IV) were very large, larger than in any of the previous studies. Normative data for temperament dimensions from the TPQ and TCI are presented, as far as the present author knows, in the largest study sample analysed to date. This is also the first study describing the population distribution of Cloninger's temperament dimensions.

The tested population cohort represents well all 31-year old persons born in one large geographic area of the genetically isolated population of Finland (Rantakallio 1988). Most previous factor analytic studies of these tests have a substantially smaller sample size and/or the samples characteristically do not represent the general population, but rather selected subpopulations, such as students.

#### *6.3.2.2 Limitations of the studies*

One limitation of the 31-year follow-up of the NFBC 1966 was that the participants consisted of only one age cohort. On the other hand, the age distribution in other validation studies of the TAS-20 has been quite narrow as well. The studies by Bagby *et al.* (1994a, 1994b), Loas *et al.* (1996), and Parker *et al.* (1993) used students as subjects. Also previous studies of the TPQ and TCI have mainly sampled college students. This

fact may affect the possibility for making generalised inferences about temperament, as age affects certain scores; for example, both Brändström *et al.* (2001) and Chen *et al.* (2002) have found that NS correlates negatively and HA positively with age.

Construct validity is never actually achieved (Morgan *et al.* 2001). Regarding the validity of the TPQ and TCI, several additional studies using different designs could strengthen these conclusions; a test-retest study, comparisons to other well-described personality scales and also evaluation of the predictive validity of the TPQ and TCI should be conducted in this and in other large population samples. Also exploratory and confirmatory factor analyses using original items, not subscales, could give more information on factor structure of these instruments.

## 7 Conclusions

### 7.1 Main conclusions

The current study added new information regarding the status of statistical methods in psychiatry. The previous studies have been very few and the only study especially on general psychiatric journals has been done in mid-1980s (Hokanson *et al.* 1986). The last ten years have been marked by a rapid expansion in computing capability and in the use of statistical modelling and other new multivariate methods. The results in this study indicate that these more demanding methods are frequently adopted in visible psychiatric research.

This study revealed that articles with clear documentation of methods receive more citations. A comprehensible and compact analysis combined with a high-quality presentation technique in a research publication gives the reader an impression of emphasis on analysis and reporting. Furthermore, a distinct section on statistical methods could help readers to understand better how the statistical analysis is done and whether the methods are correct or not. The information from such research publications is more easily assimilated by other scientists and utilised in further studies. In this study the effect of statistical methodology on citations was unfortunately low, other characteristics of the articles had more influence.

The results of this study sustain Cloninger's psychobiological model underlying TPQ and TCI, and support the utility of these tests for assessing temperament dimensions. The NFBC 1966 data suggest that the TCI dimensions are psychometrically superior to those of the TPQ, but also indicate that the reward dependence dimension could still be improved.

The NFBC 1966 data show that the TAS-20 scale operates satisfactorily also as a Finnish version. The validity of TAS-20 was studied using two different samples. The results were satisfactory with both samples. Many of the previous validation studies can be criticised because they have used student samples, and also because sample sizes have been relatively small.



The studied cohort sample is in fact the largest sample of any alexithymia or temperament study to date. As a conclusion, TPQ, TCI and TAS-20 can be used also in their Finnish versions.

## 7.2 Implications of the study

Instructors could apply the results of this study in designing training for quantitative medical research skills and psychiatric journals could utilise these findings in their instructions to authors. For readers planning psychiatric research these findings offer suggestions for acquiring statistical skills that might be helpful in their own further studies. An understandable analysis with good presentation (tables and figures) in research publication helps other researchers to utilise the results in their own further studies. These findings are important for both the authors of scientific articles and for the editors of the journals.

The presented validity of Cloninger's temperament scales and of TAS-20 has several implications, as the results of factor structure and validity of the scales enable more credible reporting from the numerous Finnish studies using these scales. The future authors can now lean on to these presented results on validity of these Finnish versions of the TPQ, TCI and TAS-20. Some of the possible future research topics with these scales are discussed in the next chapter.

## 7.3 Future research

It will be interesting to follow-up the use of statistical methods by selecting a new sample e.g. using original articles of 2001 (five-year follow-up) from the same journals as in this study. Frequencies of different statistical methods in general psychiatric journals may change rapidly, as also journals' level of interest in different psychiatric subfields may change. Also more detailed analyses by psychiatric subfields could be a topic for future research. The methodological quality of articles could also be studied, in general and e.g. regarding studies using factor analysis. The effect of the quality of methods on received citations could also be an interesting topic for future research.

Factor analysis should be utilised when studying validity of other Finnish translations of psychometric scales. At the moment there are several scales in use that have not been validated. Validity includes several subtypes and they need to be checked in Cloninger's temperament scales and in TAS-20 using other designs. Also other factor models of TAS-20 could be tested in this sample, like that of Müller *et al.* (2003). This large cohort sample also makes it possible to study the validity of the scales by subgroups, e.g. by education.

The normative data of TPQ and TCI published in this study enable comparisons of Finnish data to other presented normative samples. The use of both TPQ and TCI allows us to calculate estimates that can be used to combine the previously published TPQ

studies with TCI studies. This has not been possible until now. First preliminary comparisons to other samples have been already presented (Miettunen *et al.* 2003).

The comprehensive datasets of NFBC 1966 make it possible to study various hypotheses with Cloninger's temperament dimensions and with TAS-20 alexithymia factors. The structural validity of the Finnish version of Cloninger's temperament dimensions sets the stage for genetic investigation of temperament in this cohort. The unbiased ascertainment strategy for the original cohort should enhance confidence in the inferences drawn from such investigation. Cloninger's temperament scores can also be utilised e.g. in studies on life-style factors, on somatic disorders, on criminality and on various psychiatric disorders (e.g. personality disorders). TAS-20 can be used when studying the association between alexithymia and e.g. depression or somatisation.

In summary, bibliometric articles provide basis for future research on psychiatric research articles e.g. in the form of a follow-up of the current study. The results of the factor analytic part of this study enable several future implications as the first results of validity of the Finnish translations of TPQ, TCI and TAS-20 are now presented.

## References

- Acklin MW & Alexander G (1988) Alexithymia and somatization. A Rorschach study of four psychosomatic groups. *J Nerv Ment Dis* 176: 343–350.
- Adam D (2002) The counting house. *Nature* 415: 726–729.
- Alexopoulos GS, Vrontou C, Kakuma T, Meyers BS, Young RC, Klausner E & Clarkin J (1996) Disability in geriatric depression. *Am J Psychiatry* 153: 877–885.
- Al-Mousawi AH, Evans N, Ebmeier KP, Roeda D, Chaloner F & Ashcroft GW (1996) Limbic dysfunction in schizophrenia and mania - A study using F-18-labelled fluorodeoxyglucose and positron emission tomography. *Br J Psychiatry* 169: 509–516.
- Altman DG (1991a) *Practical statistics for medical research*. Chapman and Hall, London.
- Altman DG (1991b) Statistical methods in medical journals: developments in 1980s. *Stat Med* 10: 1897–1913.
- Altman DG (1998) Statistical reviewing for medical journals. *Stat Med* 17: 2661–2674.
- Altman DG (2000) Statistics in medical journals: some recent trends. *Stat Med* 19: 3275–3289.
- Altman DG, Gore SM, Gardner MJ & Pocock SJ (2000) Statistical guidelines for contributors to medical journals. In: Altman DG, Machin D, Bryant TN & Gardner MJ (ed) *Statistics with confidence. Confidence intervals and statistical guidelines*. British Medical Journal, Bristol, 171–190.
- Ando J, Ono Y, Yoshimura K, Onoda N, Shinohara M, Kanba S & Asai M (2002) The genetic structure of Cloninger's seven-factor model of temperament and character in a Japanese sample. *J Pers* 70: 583–609.
- Apfel RJ & Sifneos PE (1979) Alexithymia: Concept and measurement. *Psychother Psychosom* 32: 180–190.
- Arajärvi R, Haukka J, Varilo T, Suokas J, Juvonen H, Suvisaari J, Muhonen M, Suominen K, Tuulio-Henriksson A, Schreck M, Hovatta I, Partonen T & Lönnqvist J (2004) Clinical phenotype of schizophrenia in a Finnish isolate. *Schizophr Res* 67: 195–205.
- Arbuckle JL (1997) *Amos user's guide version 3.6*. SmallWaters Corporation, Chicago.
- Aschauer HN, Meszaros K, Willinger U, Fischer G, Berger P, Reiter E, Lenzinger E & Berger K (1994) Tridimensional Personality Questionnaire - some results on the validity of a German version of the questionnaire. *Neuropsychiatrie* 8: 177–181.
- Åstedt-Kurki P, Friedemann ML, Paavilainen E, Tammentie T & Paunonen-Ilmonen M (2001) Assessment of strategies in families tested by Finnish families. *Int J Nurs Stud* 38: 17–24.
- Bagby RM, Parker JDA & Taylor GJ (1991) Reassessing the validity and reliability of the MMPI Alexithymia scale. *J Pers Assess* 56: 238–253.
- Bagby RM, Parker JDA & Taylor GJ (1994a) The twenty-item Toronto Alexithymia Scale-I. Item selection and cross-validation of the factor structure. *J Psychosom Res* 38: 23–32.

- Bagby M & Taylor G (1997) Construct validation. Disorders of affect regulation. In: Taylor GJ, Bagby RM & Parker JDA (eds) *Alexithymia in medical and psychiatric illness*. Cambridge University Press, Cambridge, 46–66.
- Bagby RM, Taylor GJ & Atkinson L (1988) Alexithymia: A comparative study of three self-report measures. *J Psychosom Res* 32: 107–116.
- Bagby RM, Taylor GJ & Parker JDA (1994b) The twenty-item Toronto Alexithymia Scale-II. Convergent, discriminant, and concurrent validity. *J Psychosom Res* 38: 33–40.
- Bagby RM, Taylor GJ, Parker JDA & Loiselle C (1990) Cross-validation of the factor structure of the Toronto Alexithymia Scale. *J Psychosom Res* 34: 47–51.
- Bagby RM, Taylor GJ & Ryan DP (1986a) The measurement of alexithymia: Psychometric properties of the Shalling-Sifneos Personality Scale. *Compr Psychiatry* 27: 287–294.
- Bagby RM, Taylor GJ & Ryan DP (1986b) Toronto Alexithymia Scale. Relationship with personality and psychopathology measures. *Psychother and Psychosom* 45: 207–215.
- Baillie AJ & Mattick RP (1996) The benzodiazepine dependence questionnaire: Development, reliability and validity. *Br J Psychiatry* 169: 276–281.
- Barrett PT, Petrides KV, Eysenck SBG & Eysenck HJ (1998) The Eysenck Personality Questionnaire: An examination of the factorial similarity of P, E, N, and L across 34 countries. *Pers Individ Dif* 25: 805–819.
- Barry MM & Crosby C (1996) Quality of life as an evaluative measure in assessing the impact of community care on people with long-term psychiatric disorders. *Br J Psychiatry* 168: 210–216.
- Bartholomew DJ (1987) *Latent variable models and factor analysis*, Charles Griffin & Company Limited, London.
- Bentler PM & Stein JA (1992) Structural equation models in medical research. *Stat Methods Med Res* 1: 159–181.
- Bentler PM (1995) *EQS. Program manual*. Multivariate Software, Encino, CA.
- Berghmans T, Meert AP, Mascaux C, Paesmans M, Lafitte JJ & Sculier JP (2003) Citation indexes do not reflect methodological quality in lung cancer randomised trials. *Ann Oncol* 14: 715–721.
- Bermond B, Vorst HCM, Vingerhoets AJJM & Gerritsen W (1999) The Amsterdam Alexithymia Scale: its psychometric values and correlations with other personality traits. *Psychother Psychosom* 68: 241–251.
- Blackler D & Endicott J (2000) Psychometric properties: concepts of reliability and validity. In: *Handbook of psychological measures*. American Psychiatric Association, Washington, DC.
- Bollen KA (1989) *Structural equations with latent variables*. John Wiley & Sons, Inc, New York.
- Brändström S, Richter J & Przybeck T (2001) Distributions by age and sex of the dimensions of temperament and character inventory in a cross-cultural perspective among Sweden, Germany, and the USA. *Psychol Rep* 89: 747–758.
- Brändström S, Schlette P, Przybeck TR, Lundberg M, Forsgren T, Sigvardsson S, Nylander PO, Nilsson LG, Cloninger RC & Adolfsson R (1998) Swedish normative data on personality using the Temperament and Character Inventory. *Compr Psychiatry* 39: 122–128.
- Bulpitt CJ (1987) Confidence intervals. *Lancet* i: 494–497.
- Cardno AG, Murphy KC, Jones LA, Guy CA, Asherson P, De Azevedo MH, Coelho IM, de Macedo e Santos AJ, Pato CN, McGuffin P, Owen MJ & O'Donovan MC (1996) Expanded CAG/CTG repeats in schizophrenia. A study of clinical correlates. *Br J Psychiatry* 169: 766–771.
- Chen EYH, Lam LCW, Chen RYL & Nguyen DGH (1996) Negative symptoms, neurological signs and neuropsychological impairments in 204 Hong Kong Chinese patients with schizophrenia. *Br J Psychiatry* 168: 227–233.
- Chen WJ, Chen HM, Chen CC, Chen CC, Yu WY & Cheng AT (2002) Cloninger's Tridimensional Personality Questionnaire: Psychometric properties and construct validity in Taiwanese adults. *Compr Psychiatry* 43: 158–166.
- Child D (1990) *The essentials of factor analysis*. Cassel, London.
- Cloninger CR (1986) A unified biosocial theory of personality and its role in the development of anxiety states. *Psychiatr Dev* 4: 167–226.
- Cloninger CR (1987) A systematic method for clinical description and classification of personality variants. A proposal. *Arch Gen Psychiatry* 44: 573–588.

- Cloninger CR, Przybeck TR & Svrakic DM (1991) The Tridimensional Personality Questionnaire: U.S. Normative Data. *Psychol Rep* 69: 1047–1057.
- Cloninger CR, Przybeck TR, Svrakic DM & Wetzell RD (1994) The Temperament and Character Inventory (TCI): A guide to its development and use. Center for Psychobiology of Personality, Washington University, St Louis.
- Cloninger CR, Svrakic DM & Przybeck TR (1993) A psychobiological model of temperament and character. *Arch Gen Psychiatry* 50: 975–990.
- Cohen K, Auld F, Demers L & Catchlove R (1985) Alexithymia. The development of a valid and reliable projective measure (the objectively scored Archetypal9 Test). *J Nerv Ment Dis* 173: 621–627.
- Cole DA (1987) Utility of confirmatory factor analysis in test validation research. *J Cons Clin Psychol* 55: 584–594.
- Comrey AL & Lee HB (1992) A first course in factor analysis. Lawrence Erlbaum Associates, Inc., New Jersey.
- Costa PT & McCrae RR (1985) The NEO Personality Inventory manual. Psychological Assessment Resources, Odessa, FL.
- Cox BJ, Parker JDA & Swinson RP (1996) Confirmatory factor analysis of the fear questionnaire with social phobia patients. *Br J Psychiatry* 168: 497–499.
- Cox BJ, Wessel I, Norton GR & Swinson RP (1994) Citation patterns in anxiety disorders research in 14 journals: 1990 - 1991. *Am J Psychiatry* 151: 933–936.
- Cronbach LJ (1951) Coefficient alpha and the internal structure of tests. *Psychometrika* 6: 297–334.
- Cummings P & Rivara FP (2003) Reporting statistical information in medical journal articles. *Arch Pediatr Adolesc Med* 157: 321–324.
- Dahlstrom WG, Welsh GS & Dahlstrom L (1982) An MMPI Handbook, Volume 1, Clinical Interpretation. University of Minnesota Press, Minneapolis.
- Davis M & Lipson L (1996) A contribution to the history of family therapy in Australia and New Zealand: a bibliometric perspective. *Aust N Z J Family Therapy* 17: 9–18.
- De Fruyt F, Van De Wiele L & Van Heeringen C (2000) Cloninger's psychobiological model of temperament and character and the five-factor model of personality. *Pers Individ Dif* 29: 441–452.
- Dijkstra A, Coleman M, Tomas C, Välimäki M & Dassen T (2003) Cross-cultural psychometric testing of the Care Dependency Scale with data. *J Adv Nurs* 43: 181–187.
- Duijsens IJ, Spinhoven P, Goekoop JG, Spermon T & Eurelings-Bontekoe EHM (2000) The Dutch Temperament and Character Inventory (TCI): dimensional structure, reliability and validity in a normal and psychiatric outpatient sample. *Pers Individ Dif* 28: 487–499.
- Eggho L & Rousseau R (1990) Introduction to informetrics. *Quantitative Methods in Library. Documentation and Information Science*. Elsevier, Amsterdam.
- Ekelund J, Lichtermann D, Järvelin MR & Peltonen L (1999) Association between novelty seeking and the type 4 dopamine receptor gene in a large Finnish cohort sample. *Am J Psychiatry* 156: 1453–1455.
- Everitt BS & Landau S (1998) The use of multivariate statistical methods in psychiatry. *Stat Methods Med Res* 7: 253–277.
- Eysenck HJ & Eysenck SBG (1975) Manual of the Eysenck Personality Questionnaire. Educational and Industrial Testing Service, San Diego.
- Faryna A, Rodenhauer R & Torem M (1986) Development of an Analog Alexithymia Scale: testing in a patient population. *Psychother Psychosom* 45: 201–206.
- Finch JF & West SG (1997) The investigation of personality structure: statistical models. *J Res Pers* 31: 439–485.
- Garfield E (1986) Which medical journals have the greatest impact? *Ann Intern Med* 105: 313–320.
- Goldin J, Zhu W & Sayre JW (1996) A review of the statistical analysis used in papers published in *Clinical Radiology* and *British Journal of Radiology*. *Clin Radiol* 51: 47–50.
- Grabe HJ, Spitzer C & Freyberger HJ (2001) Alexithymia and the temperament and character model of personality. *Psychother Psychosom* 70: 261–267.

- Gutiérrez F, Torrens M, Boget T, Martín-Santos R, Sangorrín J, Pérez G & Salamero M (2001) Psychometric properties of the Temperament and Character Inventory (TCI) questionnaire in a Spanish psychiatric population. *Acta Psychiatr Scand* 103: 143–147.
- Hakko H, Räsänen P, Tiihonen J & Nieminen P (2002) Use of statistical techniques in studies of suicide seasonality, 1970 to 1997. *Suicide Life Threat Behav* 32: 191–208.
- Hansenne M, Le Bon O, Gauthier A & Ansseau M (2001) Belgian normative data of the Temperament and Character Inventory. *European Journal of Psychological Assessment* 17: 56–62.
- Harvey CA, Curson DA, Pantelis C, Taylor J & Barnes TRE (1996) Four behavioural syndromes of schizophrenia. *Br J Psychiatry* 168: 562–570.
- Haviland MG (1998) The validity of the California Q-set Alexithymia Prototype. *Psychosomatics* 39: 536–539.
- Haviland MG, Shaw DG, MacMurray JP & Cummings MA (1988) Validation of the Toronto Alexithymia Scale with substance abusers. *Psychother Psychosom* 50: 81–87.
- Herbertz H (1995) Does it pay to cooperate – a bibliometric case-study in molecular-biology. *Scientometrics* 33: 117–122.
- Herbst JH, Zonderman AB, McCrae RR & Costa PT (2000) Do the dimensions of the Temperament and Character Inventory map a simple genetic architecture? Evidence from molecular genetics and factor analysis. *Am J Psychiatry* 157: 1285–1290.
- Hintikka J, Honkalampi K, Lehtonen J & Viinamäki H (2001) Are alexithymia and depression distinct or overlapping constructs?: A study in a general population. *Compr Psychiatry* 42: 234–239.
- Hokanson JA, Bryant SG, Gardner RJ, Luttman DJ, Guernsey BG & Bienkowski AC (1986) Spectrum and frequency of use of statistical techniques in psychiatric journals. *Am J Psychiatry* 143: 1118–1125.
- Honkalampi K, Koivumaa-Honkanen H, Tanskanen A, Hintikka J, Lehtonen J & Viinamäki H (2001) Why do alexithymic features appear to be stable? A 12-month follow-up study of a general population. *Psychother Psychosom* 70: 247–253.
- Horowitz M, Sonneborn D, Sugahara C & Maercker A (1996) Self-regard: A new measure. *Am J Psychiatry* 153: 382–385.
- Hosmer DW & Lemeshow S (1989) *Applied Logistic Regression*. Wiley, New York.
- Hulley SB & Cummings SR (1988) *Designing Clinical Research*. Williams & Wilkins, Baltimore.
- Isohanni M, Isohanni I & Veijola J (2002) How should a scientific team be effectively formed and managed. *Nord J Psychiatry* 56: 157–162.
- Jamart J (1992) Statistical tests in medical research. *Acta Oncol* 37: 723–727.
- Jöreskog KG & Sörbom D (1979) *Advances in factor analysis and structural equation models*. Abt Books, Cambridge, USA.
- Jöreskog KG & Sörbom D (1996) *LISREL 8 user's reference guide*. Scientific Software International, Chicago.
- Jolliffe IT & Morgan BJ (1992) Principal component analysis and exploratory factor analysis. *Stat Methods Med Res* 1: 69–95.
- Kendler KS (1996) Parenting: A genetic-epidemiologic perspective. *Am J Psychiatry* 153: 11–20.
- Kendler KS, Eaves LJ, Walters EE, Neale MC, Heath AC & Kessler RC (1996) The identification and validation of distinct depressive syndromes in a population-based sample of female twins. *Arch Gen Psychiatry* 53: 391–399.
- Kijima N, Tanaka E, Suzuki N, Higuchi H & Kitamura T (2000) Reliability and validity of the Japanese version of the Temperament and Character Inventory. *Psychol Rep* 86: 1050–1058.
- Kim J-O & Mueller CW (1978a) Introduction to factor analysis, what it is and how to do it, Sage University paper series on quantitative applications in the social sciences, 07-013, Sage Publications, Inc, Beverly Hills and London.
- Kim J-O & Mueller CW (1978b) Factor analysis, statistical methods and practical issues, Sage University paper series on quantitative applications in the social sciences, 07-014, Sage Publications, Inc, Beverly Hills and London.
- Kleiger JH & Kinsman RA (1980) The development of an MMPI alexithymia scale. *Psychother Psychosom* 34: 17–24.

- Kline P (1994) *An easy guide to factor analysis*. Routledge, London.
- Kooiman CG, Spinhoven P & Trijburg RW (2002) The assessment of alexithymia. A critical review of the literature and a psychometric study of the Toronto Alexithymia Scale-20. *J Psychosom Res* 53: 1083–1090.
- Koskelainen M, Sourander A & Vauras M (2001) Self-reported strengths and difficulties in a community sample of Finnish adolescents. *Eur Child Adolesc Psychiatry* 10: 180–185.
- Kozeny J, Kubicka L & Prochazkova Z (1989) Psychometric properties of the Czech version of Cloninger's three dimensional personality questionnaire. *Pers Individ Dif* 10: 1253–1259.
- Krystal JH, Giller El & Cicchetti DV (1986) Assessment of in posttraumatic stress disorder and somatic illness: introduction of a reliable measure. *Psychosom Med* 48: 84–94.
- Kurmis AP (2003) Understanding the limitations of the journal impact factor. *J Bone Joint Surg Am* 85: 2449–2454.
- Landis JR & Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33: 159–174.
- Laor N, Wolmer L, Mayes LC, Golomb A, Silverberg DS, Weizman R & Cohen DJ (1996) Israeli preschoolers under scud missile attacks - a developmental perspective on risk-modifying factors. *Arch Gen Psychiatry* 53: 416–423.
- Laukkanen E, Halonen P, Aivio A, Viinamäki H & Lehtonen J (2000). Construct validity of the Offer Self-Image Questionnaire in Finnish 13-year-old adolescents: differences in the self-images of boys and girls. *Nord J Psychiatry* 54: 431–435.
- Lenzenweger MF & Dworkin RH (1996) The dimensions of schizophrenia phenomenology - not one or two, at least three, perhaps four. *Br J Psychiatry* 168: 432–440.
- Lichtermann D, Ekelund J, Peltonen L & Järvelin MR (2001) Genetic architecture of temperament. *Am J Psychiatry* 158: 1339.
- Lindeman M, Keskiavaara P & Roschier M (2000) Assessment of magical beliefs about food and health. *J Health Psychol* 5: 195–209.
- Lloyd GG & Fletcher AE (1989) Citation trends of general psychiatric journals. *Psychol Med* 19: 15–18.
- Loas G, Otmami O, Verrier A, Fremaux D & Marchand MP (1996) Factor analysis of the French version of the 20-item Toronto Alexithymia Scale (TAS-20). *Psychopathology* 29: 139–144.
- Loehlin JC (1987) *Latent variable models. An introduction to factor, path, and social structural analysis*. Lawrence Erlbaum Associates, Inc. Hillsdale, NJ.
- Luwel M (1999) Is the science citation index US-biased? *Scientometrics* 46: 549–562.
- MacCallum RC & Austin JT (2000) Applications of structural equation modeling in psychological research. *Annu Rev Psychol* 51: 201–226.
- Madden PAF, Heath AC, Rosenthal NE & Martin NG (1996) Seasonal changes in mood and behavior - The role of genetic factors. *Arch Gen Psychiatry* 53: 47–55.
- Manfredonia MG, Gasperini M, Sciuto G, Bellodi L & Smeraldi E (1991) The distribution of dimensional personality traits in an Italian general population. *New Trends in Experimental and Clinical Psychiatry* 7: 113–129.
- Mariani L & Marubini E (2000) Content and quality of currently published phase II cancer trials. *J Clin Oncol* 18: 429–436.
- Marsh HV, Balla JR & McDonald RP (1988) Goodness-of-fit indexes in confirmatory factor analysis: the effect of sample size. *Psychol Bull* 103: 391–410.
- Maziade M, Bouchard S, Gingras N, Charron L, Cardinal A, Roy MA, Gauthier B, Tremblay G, Cote S, Fournier C, Boutin P, Hamel M, Merette C & Martinez M (1996) Long-term stability of diagnosis and symptom dimensions in a systematic sample of patients with onset of schizophrenia in childhood and early adolescence. 2. Positive/negative distinction and childhood predictors of adult outcome. *Br J Psychiatry* 169: 371–378.
- McCrae RR & Costa PT (2004) A contemplated revision of the NEO Five-factor Inventory. *Pers Individ Dif* 36: 587–596.
- McGuigan SM (1995) The use of statistics in the *Br J Psychiatry*. *Br J Psychiatry* 167: 683–688.
- Metsämuuronen J (2003) Tutkimuksen tekemisen perusteet ihmistieteissä. *International Methelp*. Helsinki.

- Miettunen J, Kantojärvi L, Veijola J, Järvelin M-R & Joukamaa M (2003) Finnish normative data and international comparison of Cloninger's temperament dimensions (abstract). *Laeknabladid* 48 (Suppl.): 72.
- Morgan GA, Gliner JA & Harmon RJ (2001) Measurement validity. *J Am Acad Child Adolesc Psychiatry* 40: 729-731.
- Morlino M, Lisanti F, Gogliettino A & de Girolamo G (1997) Publication trends of papers on schizophrenia. *Br J Psychiatry* 171: 452-456.
- Mulaik SA (1986) Factor analysis and Psychometrika: major developments. *Psychometrika* 51: 23-33.
- Müller J, Bühner M & Ellgring H (2003) Is there a reliable factorial structure in the 20-item Toronto Alexithymia Scale? A comparison of factor models in clinical and normal adult samples. *J Psychosom Res* 55: 561-568.
- Müllner M, Matthews H & Altman DG (2002) Reporting on statistical methods to adjust for confounding: a cross-sectional survey. *Ann Intern Med* 136: 122-126.
- Muthén L & Muthén B (1998) *Mplus: statistical analysis with latent variables*. Muthén and Muthén, Los Angeles.
- Nakagawa A, Marks IM, Takei N, DeAraujo LA & Ito LM (1996) Comparisons among the Yale-Brown obsessive-compulsive scale, compulsion checklist, and other measures of obsessive-compulsive disorder. *Br J Psychiatry* 169: 108-112.
- Naughton M, Oppenheim A & Hill J (1996) Assessment of personality functioning in the transition from adolescent to adult life: Preliminary findings. *Br J Psychiatry* 168: 33-37.
- Nieminen P (1995) Statistical content of published therapeutic community research. *Therapeutic Communities* 16: 239-251.
- Nieminen P (1996) Type of empirical research reports as an explanatory factor in citation performance of psychiatric research. *Scientometrics* 35: 309-320.
- Nieminen P & Isohanni M (1997) The use of bibliometric data in evaluating research on therapeutic community for addictions and in psychiatry. *Subst Use Misuse* 32: 555-570.
- Nieminen P & Isohanni M (1999) Bias against European journals in medical publication databases. *Lancet* 353: 1592.
- Norton NC (1989) Three scales of alexithymia: do they measure the same thing? *J Pers Assess* 53: 621-637.
- Norušis MJ (1993) *SPSS for Windows, Professional Statistics, Release 6.0*. SPSS Inc., Chicago.
- Nummenmaa T, Kontinen R, Kuusinen J & Leskinen E (1997) Tutkimusaineiston analyysi. WSOY, Porvoo.
- Otter C, Huber J & Bonner A (1995) Cloninger's Tridimensional Personality Questionnaire: reliability in an English sample. *Pers Individ Dif* 18: 471-480.
- Pandey R, Mandal MK, Taylor GJ & Parker JDA (1996) Cross-cultural alexithymia. Development and validation of a Hindi translation of the 20-item Toronto Alexithymia Scale. *J Clin Psychol* 52: 173-176.
- Parker G, Cheah YC & Parker K (2003a) Properties of the Temperament and Character Inventory in a Chinese sample. *Acta Psychiatr Scand* 108: 367-373.
- Parker G, Hadzi-Pavlovic D, Parker K, Malhi G, Mitchell P, Wilhelm K & Austin MP (2003b) An Australian validity study of the temperament and character inventory. *Acta Psychiatr Scand* 108: 359-366.
- Parker JDA, Bagby RM, Taylor GJ, Endler NS & Schmitz P (1993) Factorial validity of the 20-item Toronto Alexithymia Scale. *Eur J Pers* 7: 221-232.
- Parker JD, Taylor GJ & Bagby RM (2003c) The 20-Item Toronto Alexithymia Scale. III. Reliability and factorial validity in a community population. *J Psychosom Res* 55: 269-275.
- Parker JDA, Taylor GJ, Bagby RM & Thomas S (1991) Problems with measuring Alexithymia. *Psychosomatics* 32: 196-202.
- Pélissolo A & Lépine JP (2000) Normative data and factor structure of the Temperament and Character Inventory (TCI) in the French version. *Psychiatry Res* 94: 67-76.
- Peritz BC (1994) On the heuristic value of scientific publications and their design; a citation analysis of some clinical trials. *Scientometrics* 30: 175-186.



- Pincus HA, Henderson B, Blackwood D & Dial T (1993) Trends in research in two general psychiatric journals in 1969-1990: research on research. *Am J Psychiatry* 150: 135-142.
- Pulver A, Allik J, Pulkkinen L & Hämäläinen, M (1995) A Big Five personality inventory in two non-Indo-European languages. *European Journal of Personality* 9: 109-124.
- Puttonen S (1998) Cloningerin Temperament and Character Inventory (TCI): käsitelvaliditeetti ja psykometriset ominaisuudet. Unpublished MA thesis, Univ Helsinki, Dept Psychology.
- Rantakallio P (1988) A longitudinal study of the Northern Finland birth cohort for 1966. *Paediatr Perinat Epidemiol* 2: 59-88.
- de la Rie SM, Duijsens IJ & Cloninger CR (1998) Temperament, character and personality disorders. *J Personal Disord* 12: 362-372.
- Ritchev PN, Frank RA, Hursti U-K & Tuorila H (2003) Validation and cross-national comparison of the food neophobia scale (FNS) using confirmatory factor analysis. *Appetite* 40: 163-173.
- Romans SE, Martin J & Mullen P (1996) Women's self-esteem: A community study of women who report and do not report childhood sexual abuse. *Br J Psychiatry* 169: 696-704.
- Sackeim HA, Luber B, Katzman GP, Moeller JR, Prudic J, Devanand DP & Nobler MS (1996) The effects of electroconvulsive therapy on quantitative electroencephalograms - Relationship to clinical outcome. *Arch Gen Psychiatry* 53: 814-824.
- Salokangas RK, Honkonen T, Stengard E & Koivisto AM (2002) Symptom dimensions and their association with outcome and treatment setting in long-term schizophrenia. Results of the DSP project. *Nord J Psychiatry* 56: 319-327.
- SAS Institute (1999) SAS/STAT® user's guide, version 8. SAS Institute, Cary, NC.
- Sifneos PE (1973) The prevalence of alexithymic characteristics in psychosomatic patients. *Psychother Psychosom* 22: 255-262.
- Sifneos PE (1986) The Shalling-Sifneos Personality Scale-Revised. *Psychother Psychosom* 45: 161-165.
- Sifneos PE (1996) Alexithymia: Past and present. *Am J Psychiatry* 153 (Suppl. 7): 137-142.
- Spearman C (1904) General intelligence objectively determined and measured. *Am J Psychol* 15: 201-293.
- SPSS Inc. (2001) SPSS Base 11.0 for Windows User's Guide. SPSS Inc., Chicago.
- Stata Corporation (2001) Stata User's Guide. Stata Press, College Station, Texas.
- Sternberg RJ, Castejon JL, Prieto MD, Hautamäki J & Grigorenko ELAF (2001) Confirmatory factor analysis of the Sternberg Triarchic Abilities Test in three international samples: An empirical test of the triarchic theory of intelligence. *European Journal of Psychological Assessment* 17: 1-16.
- Sterne JAC & Smith GD (2001) Sifting the evidence - what's wrong with significance testing. *BMJ* 322: 226-331.
- Stevenson J, Thompson MJJ & Sonuga-Barke E (1996) Mental health of preschool children and their mothers in a mixed urban rural population. III. Latent variable models. *Br J Psychiatry* 168: 26-32.
- Stewart ME, Ebmeier KP & Deary IJ (2004) The structure of Cloninger's Tridimensional Personality Questionnaire in a British sample. *Pers Individ Dif* 36: 1403-1418.
- Streiner DL (1995) Stayin' alive: an introduction to survival analysis. *Can J Psychiatry* 40: 439-444.
- Sung SM, Kim JH, Yang E, Abrams KY & Lyoo IK (2002) Reliability and validity of the Korean version of the Temperament and Character Inventory. *Compr Psychiatry* 43: 235-243.
- Svrakic DM, Przybeck TR & Cloninger CR (1991) Further contribution to the conceptual validity of the unified biosocial model of personality: US and Yugoslav data. *Compr Psychiatry* 32: 195-209.
- Svrakic DM, Whitehead C, Przybeck TR & Cloninger CR (1993) Differential diagnosis of personality disorders by seven-factor model of temperament and character. *Arch Gen Psychiatry* 50: 991-999.
- Takala M & Hakkarainen M (1953) Factor analysis and validity of the Wartegg-Drawing test/Ueber Faktorenstruktur und Validität des Wartegg-Zeichentests. *Annales Academiae Scientiarum Fennicae (Helsinki)* 81 Serie B: 95.

- Takeuchi M, Yoshino A, Kato M, Ono Y & Kitamura T (1993) Reliability and validity of the Japanese version of the Tridimensional Personality Questionnaire among university students. *Compr Psychiatry* 34: 273–279.
- Taylor GJ (2000) Recent developments in alexithymia theory and research. *Can J Psychiatry* 45: 134–142.
- Taylor GJ, Bagby RM & Parker JDA (1997) Disorders of affect regulation. Alexithymia in medical and psychiatric illness. Cambridge University Press, Cambridge, UK.
- Taylor GJ, Bagby RM & Parker JD (2003) The 20-Item Toronto Alexithymia Scale. IV. Reliability and factorial validity in different languages and cultures. *J Psychosom Res* 55: 277–283.
- Taylor GJ, Ryan D & Bagby RM (1985) Toward the development of a new self-report alexithymia scale. *Psychother Psychosom* 44: 191–199.
- Thomas P, Kearney G, Napier E, Ellis E, Leudar I & Johnston M (1996) The reliability and characteristics of the brief syntactic analysis. *Br J Psychiatry* 168: 334–337.
- Vazquez-Barquero JL, Lastra I, Nunez MJC, Castanedo SH & Dunn G (1996) Patterns of positive and negative symptoms in first episode schizophrenia. *Br J Psychiatry* 168: 693–701.
- Vogt WP (1993) Dictionary of statistics and methodology. A nontechnical guide for the social sciences. Sage Publications, Newbury Park.
- Vorst HCM & Bermond B (2001) Validity and reliability of the Bermond-Vorst Alexithymia Questionnaire. *Pers Individ Dif* 30: 413–434.
- Walter SD (1995) Methods of reporting statistical results from medical research studies. *Am J Epidemiol* 141: 896–906.
- Wang Q & Zhang B (1998) Research design and statistical methods in Chinese medical journals. *JAMA* 280: 283–285.
- West R & McIlwaine A (2002) What do citation counts count for in the field of addiction? An empirical evaluation of citation counts and their link with peer ratings of quality. *Addiction* 97: 501–504.
- Weyers P, Krebs H & Janke W (1995) Reliability and construct validity of the German version of Cloninger's Tridimensional Personality Questionnaire. *Pers Individ Dif* 19: 853–861.
- Wolff G, Pathare S, Craig T & Leff J (1996a) Community attitudes to mental illness. *Br J Psychiatry* 168: 183–190.
- Wolff G, Pathare S, Craig T & Leff J (1996b) Community knowledge of mental illness and reaction to mentally ill people. *Br J Psychiatry* 168: 191–198.
- Zohar AH, Lev-Ari L, Benjamin J, Ebstein R, Lichtenberg P & Osher Y (2001) The psychometric properties of the Hebrew version of Cloninger's Tridimensional Personality Questionnaire. *Pers Individ Dif* 30: 1297–1309.