Statistica Neerlandica (1996) Vol. 50, nr. 1, pp. 111–135

Fifty years of measurement and scaling in the Dutch social sciences

P. G. M. van der Heijden and K. Sijtsma*

Department of Methodology and Statistics, FSW, Utrecht University, PO Box 80140, 3508 TC Utrecht, The Netherlands

It is evident from recent issues of journals like Psychometrika and Applied Psychological Measurement that the Dutch contribution to the development of the area of measurement and scaling is considerable. In the first part of this paper, we try to answer the question how this could happen. In the early development De Groot in Amsterdam and Van de Geer in Leiden created an academic climate that made it possible for many researchers to develop their skills. Some other possible reasons are sketched for the flourishing of this area in the Netherlands, but it is difficult to assess which reasons were most important. In the second part of the paper, we analyze the international impact of the Dutch contribution by a citation analysis in four fields of measurement and scaling: factor analysis, test theory including item response theory, latent class analysis, and optimal scaling and multidimensional scaling. It appears that Dutch researchers not only publish a lot, but also that the impact of their publications is at the same level as publications produced world-wide.

Key Words: bibliometric analysis, history of statistics, measurement, psychometrics, scaling.

1 Introduction

Why is the contribution of Dutch social scientists in the area of measurement and scaling so large? Is this due to the stimulating influence of some early researchers, or are other factors also important? Has the impact by Dutch researchers on certain topics in measurement and scaling been particularly large? These are some of the questions that we try to answer in this paper, written for the special issue of Statistica Neerlandica celebrating its 50th anniversary.

Measurement and scaling covers a large part of the statistical methodology used and developed in the social sciences (cf. MOLENAAR, 1995). Briefly typified, the area deals with the construction of variables that are often operationalizations of hypothetical constructs such as intelligence, the attitude towards euthanasia, and the preference for particular brands of beer. These variables are used to measure individual and group differences and to scale stimuli such as the items in an intelligence test, the questions in an attitude questionnaire, or brands of beer in a taste

* sijtsma@fsw.ruu.nl

preference experiment for marketing purposes. Because this is such an important and large area in social science statistics, and because we had to restrict the topic of this paper, we will not deal with the development of statistical methods that are used for data analysis after the measurement and scaling problems have been resolved; for example, methods based on analysis of variance, regression analysis, and categorical data analysis.

The development of statistical methods in the social sciences in the Netherlands started only a few decades ago. A broader interest in statistical methods was rare until the late 1960s. A growing number of researchers was attracted towards this new field during the 1970s. The 1980s showed its full bloom and a consistent trend towards maturity. Fortunately, this trend has shown to be robust through the years since. Research results are now being published in leading international journals at a rate that might falsely suggest that development of statistical methods had already seen a long tradition in psychology, sociology, political science, and education.

We will explore and document the surprisingly fast development of the discipline of measurement and scaling, sometimes referred to as psychometrics, for short, in the period from 1945 until now in two parts. The first part is a historical overview of the origin and development of the interest in statistical methods in general, with a focus on the history of measurement and scaling from 1945 to 1980. Most attention is given to psychology where the majority of the developments took place. Of course, contributions from other social science disciplines are also noted.

Until now history about our field had been scattered around in bits and pieces. A first category is formed by official (that is, published) but rare and incomplete accounts such as by DRENTH (1965, 1975, chapter 1) and in a special issue of 'Tijdschrift voor Onderwijsresearch' ('Journal of Educational Research'; 1984, issue 5, dedicated to Van Naerssen, on the occasion of his retirement). An elaborate account of the development of methodology in psychology was given by DEHUE (1990). A second category are written sources such as letters, grant and other proposals, and minutes. Such sources are often in the private files of individuals and, therefore, difficult of access to the amateur historian. Even worse, they may have been thrown away or lost, or their existence may simply have been forgotten by the owner. Finally we have the memories in the minds of those involved in the development of the field. Memories may be less reliable but certainly easier of access and they were expected to provide valuable information that has never appeared in writing.

This overview is mostly based on interviews we held in the spring of 1995 with De Leeuw, Hagenaars, Heiser, Mellenbergh, Molenaar, Roskam, Ten Berge, Van de Geer, Van der Linden and Verhelst. We later consulted De Groot, De Gruijter, Mokken, Mooijaart, 't Hart, Van der Kamp and Van Rijckevorsel for additional information. Written sources were used for additional information or to check certain aspects of the verbal information. The precaution that we both took notes during the interviews has probably reduced random measurement error but can not prevent an occasional, unnoticed systematic bias due to memory failure in some of the interviewees. We believe, however, that the consistency of the information we

collected from different sources gives a generally valid impression of the development of the field. We have combined the verbal accounts with official accounts. Several cross checks were performed to boost reliability and validity. We are to blame for any remaining flaws.

The second part of this paper contains a bibliometric analysis of the contributions by Dutch social scientists engaged in the development of psychometric methods to the relevant international journals in the period from 1980 to 1993. For a citation analysis we divided the area of measurement and scaling into four fields: factor analysis, test theory including item response theory, latent class analysis, and optimal scaling and multidimensional scaling. The analysis showed that Dutch contributions have an impact that is comparable with international standards. This is a fine result considering that publishing in international outlets was an exception to the rule less than 20 years ago. The bibliometric results have been extended with some results from other sources than the reasonable but incomplete set of journals and the limited period of time covered.

2 Part One: Historical Overview

2.1 The Period after the Second World War

Although in the 1940s and 1950s several psychologists were oriented toward empirical research inspired, for example, by the earlier work of Heymans (Groningen) at the beginning of the century and Révész (Amsterdam) in the 1930s and 1940s, psychology and also sociology in the Netherlands were, in general, less strongly oriented towards empirical research than in the USA. For example, important influences in psychology came from phenomenology and existential philosophy (Utrecht), and from psychoanalysis (Amsterdam). Sociological thinking was much influenced by reflections about cultural phenomena and the organization of society. Empirical research in sociology often had the appearance of qualitative case studies and in psychology terms like intuition, empathy, and 'Verstehen' more adequately typified the spirit of the time than they do today.

This was much in contrast to the developments in the USA where there had been a long tradition of measurement and scaling. This tradition went back to the First World War when the aptitudes of massive numbers of recruits were assessed and to the ongoing flow of new immigrants whose abilities and aptitudes were measured when they entered the country. The first serious measurement models were proposed by THURSTONE (1927) in psychology and GUTTMAN (1944, 1950) and LAZARSFELD (1950) in sociology. The first journal devoted to quantitative psychology, Psychometrika, saw the light in 1935, followed by another journal, Educational and Psychological Measurement, in 1941. The interest in quantitative measurement methods in the UK even dates back further (SPEARMAN, 1910). MELLENBERGH (1984) mentions EDGEWORTH (1888) as the oldest paper in psychological measurement known to him. A comprehensive review of such developments can be found in DRENTH (1975, chapter 1).

Given the much different climate in the Netherlands, 'measurement' of individual differences on psychological traits was often more subjective impression formation on the basis of the responses of individuals to stimuli presented by the psychologist. Such stimuli could be suggestive cartoons or pictures, or the well known or perhaps notorious inkblots from the Rorschach test. Another possibility was that subjects drew pictures which were then used to infer aspects of the subject's personality. Much valued for this purpose was also the subject's handwriting (DEHUE, 1990, pp. 34–72; DRENTH, 1975, p. 55). This has long remained the reason why job application letters were required in handwriting.

However, there were also more serious attempts at measurement. In the early decades of the century there had been a change in the climate in the industry which led to more job differentiation and a stronger need for the objective assessment of the individual worker's skills. DEHUE (1990, pp. 37–41) describes the early attempts in the 1910s and 1920s by Brugmans and Prak, inspired by this change of climate, to assess applicants for jobs by means of systematic measurement of senso-motor and cognitive abilities. KOUWER (1952) published a monograph in which he briefly discussed the statistical requirements on measurement by means of psychological tests. SNIJDERS and SNIJDERS-OOMEN (1958, SNIJDERS-OOMEN, 1943) constructed a well-known non-verbal intelligence test. Performance and personality tests on which the subject's responses could be objectively evaluated and numerically scored were translated from English into Dutch. Such translations can alter the psychological meaning of the test results as well as the statistical or psychometric characteristics of the measurement instrument. However, the skills to use the more advanced statistical methods were still lacking in most researchers and, as a result, they were not able to sufficiently justify the use of these tests by collecting data followed by thorough statistical analysis.

The same can be said of sociology and political science, where sometimes questionnaires were used in research. This use was partly inspired by marketing research and American voters research. In addition, descriptive statistics were already used in empirical research in the 1950s. However, there was hardly any interest in the development of measurement and scaling tools at that time. More generally, inferential statistics had not yet entered the stage.

2.2 The Early Years of Measurement and Scaling Amsterdam

The 1950s would bring the first signs of change. Early work in what is now known as classical test theory was done by Van der Giessen and Wiegersma (DRENTH, 1975, pp. 54–55). From sociology came the early work of Stouthard inspired by Guttman and others. Most changes, however, came from psychology. The interviewees agreed in their opinion that one of the two great initiators was the psychologist A. D. de Groot, then professor of applied psychology at the University of Amsterdam. The other was the psychologist J. P. van de Geer from Leiden to whom we will return later.

De Groot was only incidentally engaged in the development of psychometric methods himself, but he stimulated many others to do so. He emphasized that in psychology ideas had to be tested, and conclusions to be based on empirical evidence, and stimulated those in his environment to do empirical research and use statistical methods to analyze the data. He was influenced by MEEHL's (1954) monograph 'Clinical versus statistical prediction'. This study demonstrated the superiority of statistical prediction over prediction by the clinician based on his diagnostic skills and intuition but without the help of formalized methods. De Groot was also familiar with the famous monograph by GULLIKSEN (1950) on the psychometric theory of mental test scores. This book was the basis for his course in psychometrics, which was presumably the first course in psychometrics taught in the Netherlands.

De Groot's acquaintance with psychological measurement by means of tests already stemmed from the 1940s when he was head of the psychology section at Philips Eindhoven (LUIJTEN, 1993). There he was involved in the testing of personnel. Through contacts in the company he became involved in testing for educational selection purposes. These contacts were continued when he became a professor at the University of Amsterdam. In 1957 he founded the 'Research Instituut voor de Toegepaste Psychologie' (RITP; 'Research Institute for Applied Psychology'), that initially resided in his home and later became affiliated with the University of Amsterdam. After De Groot's retirement in 1979, the educational research task of RITP was taken over by the present SCO/Kohnstamm Institute. At RITP, where most of the work consisted of test construction for education and selection, the 'Amsterdamse schooltoetsen' ('Amsterdam schooltests') were constructed. These were the predecessors of the nowadays famous and feared 'Cito Basisschooltoetsen' ('Elementary School-Leaving Tests').

In 1958 De Groot visited Educational Testing Service (ETS; Princeton, NJ) and became much impressed by the technological lead of the Americans in the area of test construction. His experiences at ETS and the activities at RITP were very important for De Groot's later efforts to found the 'Centraal Instituut voor Toetsontwikkeling' (Cito; National Institute for Educational Measurement) at Arnhem in 1968. The success of this effort was also based on the insight to choose the right moment to convince the Dutch government of the usefulness of having an institution for educational measurement in the Netherlands (LUIJTEN, 1993).

Another influence that facilitated the founding of Cito and, in general, the growth of Dutch psychometrics was De Groot's 'Afdeling Examen Techniek' (AET; 'Division in Educational Testing'), which was located at the faculty of psychology of the University of Amsterdam. AET provided the theoretical foundation of the work done at RITP such as achievement test construction. Head of AET became R. F. van Naerssen, attracted and appointed by De Groot after having supervised the former's doctor's dissertation. Van Naerssen, also a psychologist by training (Leiden), was according to several of the interviewees the first Dutch psychometrician by modern standards. Mellenbergh was trained by De Groot and Van Naerssen and worked at AET. He thinks he was one of the first Dutchmen who used advanced

statistical and psychometric methods in psychological test construction. During the 1970s AET transformed into the Department of Methodology where Mellenbergh is nowadays in charge.

One of the many merits of De Groot was that everything done at AET would have to lead to publications. In his own words, paraphrased by Mellenbergh: 'Research that has not been published was never performed'. The hundreds of AET memoranda (VAN HEERDEN and HOOGSTRATEN, 1984) were the basis for papers and dissertations such as Mellenbergh's (1971), and were discussed in weekly meetings by the AET staff. In a time when the publication of research results was rare in general, this unusual activity set several young researchers on the tracks that finally led to the volume of output that characterizes Dutch psychometrics today.

Van Naerssen set an important example by publishing a large number of papers between 1958 and 1984, mainly in Dutch journals such as special issues of 'Nederlands Tijdschrift voor de Psychologie' ('Netherlands Journal of Psychology') and 'Tijdschrift voor Onderwijsresearch', which was founded by Van Naerssen and De Groot. Van Naerssen's dissertation (1962), on decision-making using psychological test scores, was much inspired by the work of CRONBACH and GLESER (1957) on decision-making using psychological tests. A chapter (VAN NAERSSEN, 1965) based on his dissertation was published in the next edition of Cronbach and Gleser's (1965) 'Psychological Tests and Personnel Decisions'. Although he mainly published in Dutch journals and has not been cited very often (VAN HEERDEN and HOOGSTRATEN, 1984), his value lies in the inspiration he was for others who were trying to climb onto the stage in those days.

Another initiative by Van Naerssen in the late 1960s was the informal forum 'Objectieve Studietoetsen' ('Study Group on Objective Educational Testing') which transformed into the much acclaimed 'Werkgroep Meetmethoden' ('Study Group Measurement Methods') a few years later. Here several researchers interested in psychometric methods, such as Van Naerssen and Mellenbergh (Amsterdam), Ten Berge, Hofstee and Molenaar (Groningen), and Van der Kamp and De Gruijter (Leiden), met on a regular basis. In the early seventies this forum transformed again into the 'Nederlandse Stichting voor Psychometrie' ('Netherlands Foundation for Psychometrics'), which stayed active until the mid-eighties. Very popular were the meetings every three months in Amsterdam where lectures were held and recent papers from the 'Tijdschrift voor Onderwijsresearch' were discussed. This was also the birthplace of the 'Werkgroep Raschmodellen' ('Study Group Rasch Models') in late 1981 which brought together the Dutch effort to develop item response theory (IRT). The international breakthrough of Dutch psychometrics was then about to take place.

Another important influence came from Amsterdam in the person of R. J. Mokken. He was a political scientist by training. Like many others he was influenced by De Groot and Duijker (Amsterdam). In the first half of the 1960s he was at the Mathematical Centre in Amsterdam to specialize in mathematical statistics and to prepare his dissertation. Mokken was familiar with Rasch's (1960) monograph and

had contacts with Birnbaum who, like Rasch, did very important theoretical work in IRT. Although much impressed by their work, he was looking for less restrictive IRT models that could replace GUTTMAN'S (1944, 1950) deterministic scalogram model and that could account for most of the data structure. Along the lines of the Amsterdam school of mathematical statistics he set himself to develop distribution-free or nonparametric models to this end. To paraphrase Mokken: 'Items were scarce and we could not afford to throw many away as a result of using restrictive psychometric models'. He finished his dissertation in 1970 and in 1971 it was published as a monograph (MOKKEN, 1971). The book used mathematical statistics to deal with IRT. One of its topics was the Rasch model, very popular among Dutch psychometricians in the 1980s and 1990s; but the larger part was devoted to a nonparametric formulation of IRT.

Several of the interviewees mentioned Mokken for different reasons. According to Heiser (Leiden) Mokken was the first to import ideas about modern (IRT) measurement models into Dutch psychometrics. He was also the first to introduce mathematical statistics and inferential statistics into Dutch psychometrics. Molenaar (Groningen), who is mathematician by training and held a job at the Mathematical Centre in the 1960s, met Mokken there and was much inspired by the latter's ideas about measurement and scaling. Mellenbergh (Amsterdam) regards Mokken as the founding father of Dutch IRT and claims that Mokken's monograph actually was ahead of its time (this is illustrated by the citation pattern of his monograph, see Table 4).

As to international outlets Mokken and his associates published their ideas outside the psychometric circuit and taught them at the yearly summerschool of the European Consortium for Political Research at the University of Essex, Colchester, UK. In the 1970s there was already a computer program for scale analysis that found its way to researchers in the Netherlands and West-Germany and that has been updated at regular intervals ever since (MOLENAAR et al., 1994). Mokken did not publish his ideas in international psychometric outlets until the 1980s (MOKKEN and LEWIS, 1982).

Although Mokken's work was a rich source of inspiration for several researchers (for example, Gadourek, Stokman, Van Schuur and Molenaar) till the present day, it was pushed aside by FISCHER'S (1974; Vienna) monograph which made a strong case for the Rasch model. The parametric formulation of this model and its philosophical foundation somehow had a stronger appeal for many researchers. This was corroborated by Roskam (Nijmegen), Van der Linden (Twente) and Verhelst (Cito) who all mentioned the fact that the Rasch model is an exponential family model and should therefore be preferred over other models. In the early 1980s there was even a discussion between opponents and proponents of the Mokken approach to IRT. This was a source of inspiration to researchers such as Verhelst and Molenaar.

A relatively large part of the Dutch IRT research thus went in the direction of the Rasch model and other Rasch-based exponential family models. Roskam had met Andersen (Copenhagen), a well-known young coworker of Rasch, at the

Mathematical Psychology Group Meeting in Regensburg (West-Germany, 1974), and Fischer (Vienna), who had been one of the first to proceed with Rasch's psychometric heritage, at the educational measurement conference in Montreux (Switzerland, 1975). These contacts led to the visit of Van den Wollenberg, a student of Roskam, to Fischer in Vienna in 1976. Van den Wollenberg wrote a dissertation (VAN DEN WOLLENBERG, 1979) about the Rasch model that inspired many other researchers such as Molenaar and Glas (Cito). He also published his results in international journals (VAN DEN WOLLENBERG, 1982a, 1982b).

After Mokken had left the university in the late 1970s for a job at Statistics Netherlands (CBS), he was succeeded by Saris at the faculty of political science of the University of Amsterdam. In the 1970s Saris was one of the first to study covariance-structure analysis. Others were Stouthard (Tilburg), Van der Kamp and Mellenbergh. In this context, he had met Jöreskog (Sweden) and had worked together with Satorra (Spain). Saris was also mentioned as an important initiator of the 'Sociaal-Wetenschappelijke Sectie' ('Social Science Section') of the 'Vereniging Voor Statistiek' (VVS; Netherlands Society for Statistics and Operations Research). In the late 1970s he took the initiative to start the 'Methoden en Data Nieuwsbrief' ('Newsletter on Methods and Data') which later became the VVS' unofficial outlet 'Kwantitatieve Methoden' ('Quantitative Methods'). According to several interviewees these initiatives were important for the incorporation of statistics in the social sciences.

Leiden

The interviewees agreed that the second important initiator of the development of statistical methods in the Dutch social sciences was J. P. van de Geer from Leiden University. Like De Groot, Van de Geer was psychologist by training. In 1948 he became the assistant of Chorus, who was the first professor of psychology in Leiden, and had to teach statistics to psychology students. His interest in statistics was stimulated by a monograph written by Thurstone about factor analysis. To understand this book he started to consult other statistics texts. Another source of inspiration came from a review by Kouwer (Utrecht) of Chorus' book on intelligence. Chorus asked Van de Geer to prepare a reply. The latter was thus forced to think about Chorus' quantitative concept of item difficulty which had been heavily criticized by Kouwer.

The really important breakthrough for Van de Geer was the visit of Coombs to the Department of Psychology at the University of Amsterdam in 1955 (VAN DE GEER and BEZEMBINDER, 1988). There Van de Geer followed a course taught by Coombs and became much impressed by the latter's teaching skills, his systematic treatment of data types, and the fact that statistics appeared to be something one could do for a living rather than an obligatory subject to teach to students. A few years later Van de Geer visited Coombs in Ann Arbor (Michigan, USA). In 1963 Coombs taught a course in Leiden. In 1975 he became *doctor honoris causa* at Leiden University. It may be noted here that Coombs was also a source of inspiration to many others, such as Bezembinder (Nijmegen) who visited him for a year in 1962–1963.

Already in 1957 Van de Geer had applied an advanced form of analysis of variance in his dissertation, which was unusual in those days, and a few years later he was one of the first in the Netherlands to do a canonical correlation analysis. Based on work by Kruskal he also became interested in non-metrical analysis of distance data. The American statistical literature was his only source of information.

Van de Geer told us that it had never been his intention to found a school, and that his liberal attitude towards the work of others simply ruled out such ambitions. He did not have many contacts in the Netherlands outside Leiden because his colleagues were interested in different topics. For this same reason he did not attend the meetings of the Werkgroep Meetmethoden and other discussion forums. However, he had many contacts outside the Netherlands, in particular in the USA.

It seems safe to conclude that Van de Geer's work and ideas were compelling enough to compensate for his lack of ambitions. His influence was two-fold: substantively and facilitating (see the *Liber Amicorum* published in honour of Van de Geer's retirement; CROMBAG, VAN DER KAMP and VLEK, 1987). Substantively he did influential work in working out a geometric approach to multivariate analysis. He wrote a book in English (VAN DE GEER, 1971) on this topic that is probably the most often cited Dutch social science statistics publication (see Part Two of this paper). This book was non-standard in the sense that very limited attention was given to statistical testing, distributions and the like, and that a large part of the book was devoted to matrix algebra (for a quantitative comparison of this book with other multivariate analysis books, see GIFI, 1990, Chapter 1). This influence can be traced in the two undergraduate text books called "Methoden en technieken van psychologisch onderzoek" ('Methods and Techniques of Psychological Research') (two parts) written by a group of authors from the Department of Methods and Techniques for Psychology that coined themselves MEERLING (1980). In the spirit of Van de Geer, these books mostly avoided discussing statistical testing. This can also be traced in the work of De Leeuw and Gifi to be discussed below. A debate focussing on this descriptive data-analytic as opposed to inferential statistical attitude was held in 1987, and is documented in DE LEEUW (1987; see also MOLENAAR, 1987).

The facilitating influence of Van de Geer is that he provided an intellectual climate that was very fruitful for many young researchers. A first example was the organization of the Nuffic summer schools in The Hague in the 1960s, which were mentioned by many interviewees. The Nuffic summer schools were important events in the development of Dutch psychometrics. Van de Geer organized the first two in 1964 and 1966 on decision-theory and measurement theory, respectively. The 1966 summer school was particularly important, with teachers such as Rasch, Coombs and Torgerson. The goal of the summer schools was to bring young researchers into contact with well known (non-Dutch) researchers. The 1966 conference had approximately 100 participants, mostly European and relatively many Dutchmen. Here it was that Fischer met Rasch for the first time. Because in the 1970s Fischer rather than Rasch himself was the most important proponent of the Rasch model this acquaintance proved to be very important for the group of Dutch IRT researchers

who studied the Rasch model in later decades. This and other summer schools were extremely stimulating for many people. Here young researchers had the possibility to meet leading international researchers; several were inspired to choose psychometrics as their future trade.

A second example stems from dissertations that were written under his supervision. One of the first dissertations was Roskam's (1968), like Van de Geer a psychologist by training from Leiden. In the late 1950s, Roskam already used factor analysis (published in 1959). In case of problems he consulted Van de Geer. He was also influenced by the work of Van Naerssen and by Bezembinder to whom he attributes his interest in measurement. The famous papers by KRUSKAL (1964a, 1964b) inspired Roskam to program the non-metrical multidimensional scaling method proposed there. After Bezembinder had asked him to come to Nijmegen he also programmed non-metrical versions of additive conjoint measurement, unfolding and factor analysis. Together with multidimensional scaling these topics formed his dissertation. Another important psychometrician who wrote his dissertation under the supervision of Van de Geer was TEN BERGE (1977) (Groningen) on the topic of the matching of loadings matrices in factor analysis.

A third example of his inspiration was the foundation of the Department of Data Theory in approximately 1970. Van de Geer had the important insight that data analysis was not related to particular substantive areas such as psychology. As the former-dean of the Faculty of Social Sciences he was able to set up this department, that had the mission to do research from this perspective. The department had no teaching obligations. The study of methods related to psychological research was done in another department where Van der Kamp became the head. This department was also responsible for much of the statistics training for psychology students.

De Leeuw was one of the first researchers at the Department of Data Theory. As he described in the preface of GIFI (1990), Van de Geer gave him the opportunity to develop his own line of research in optimal scaling. Like so many others, at first De Leeuw was not interested at all in quantitative methods during his psychology training. However, this changed quickly after he became acquainted with the work of Coombs and after his visit to the Nuffic summer school on measurement and scaling. From that moment on he was much oriented towards the American literature on statistics and data analysis.

Unlike several others, however, De Leeuw always felt a strong need to write everything down to better direct his thinking. To him writing was also an end in itself and was not limited to scientific work. Initially, he did not publish his results in journals because his writing was far too voluminous for that and, besides, publication would take too much time anyway. He rather wrote internal reports which attracted the attention of Kruskal at Bell Laboratories (New Jersey, USA). As a result, in 1973 he was invited to visit Kruskal and became acquainted with the American publication policy. He also met Young and Takane at that time and together they started a series of publications that appeared in the second half of the 1970s and expanded many of the ideas from the internal reports De Leeuw wrote earlier in Leiden. This work mainly

dealt with known techniques such as regression analysis, principal component analysis and common factor analysis, which were extended with optimal scaling features.

De Leeuw was a charismatic person who succeeded in founding a school on optimal scaling and multidimensional scaling. In the 1970s he motivated young researchers such as Heiser, Meulman, Stoop, Van der Burg and Van Rijckevorsel, and together with the effort of others (see GIFI, 1990) this culminated in two series of computer programs and a much cited internal report (GIFI, 1981; see Table 4). The first series was the ALS series for nonlinear multivariate analysis, parts of which were recently included in SPSS. The first FORTRAN version of HOMALS was programmed by Van Rijckevorsel in the mid-seventies. The second series was the SMACOF series for unfolding and multidimensional scaling programmed by Heiser.

An internal report on nonlinear multivariate analysis, first written in Dutch in 1980, and rewritten and translated into English one year later, was published under the author name of Albert Gifi. Few people know that Gifi was the man-servant of Sir Francis Galton. The author name of Gifi actually represented a group of authors and, according to De Leeuw, was chosen as a result of feelings of equality among the authors, the need to avoid a large number of authors on the cover of the report, and plain snobism. The report showed a strong preference for geometrical representations, which was where the influence of Van de Geer was felt most strongly. The Gifi-project, that started around 1968, ended around 1982. Different participants report different dates, however, and it could even be argued that the project still continues today. The Gifi-project saw a large number of researchers participating and being trained in it throughout the years. The author Gifi reappeared once in 1990 when the Gifi report was finally edited into an official book by Heiser, Meulman and Van den Berg.

The impact of De Leeuw has been much larger than through the Gifi-project. He is co-author of often cited papers of many other PhD students (see Part Two of this paper), among which are Kroonenberg, working on methods for the decomposition of three-way matrices, and Bekker (now in economics), working on identification. Other contributions were in multi-level modelling and in the estimation of the Rasch model. In 1987, De Leeuw left Leiden for the University of California at Los Angeles.

2.3 Early Publishing Policy

To understand the unofficial or implicit publishing policy of Dutch psychometricians through the years it is important to know that in the 1960s and earlier only few papers from Dutch social and behavioural scientists were published in national, let alone international outlets. The earliest exception to this rule was experimental psychology. The next exception was to become psychometrics in the late 1970s. In the 1960s Van Naerssen, among others, had already set an example at the national level and De Groot's earlier mentioned quote was unequivocal in its intentions. However, these were exceptions.

We have the impression that publication of research results was initially highly dependent upon the individual's motivation and circumstances and that the majority at first was not inclined to publish. Perhaps this down-to-earth or even somewhat

vulgar activity did not fit into the contemplative atmosphere of the Dutch universities at that time. Another reason mentioned sometimes was the lack of communication between researchers from different universities that seemed characteristic of the 1960s and earlier (e.g., SWANBORN, 1995). Thus many were doing their research in relative isolation, although Mellenbergh remembers that already in the 1960s several flourishing contacts existed between the Amsterdam school of De Groot, and Van der Kamp (Leiden), Kouwer and Hofstee (Groningen) and Fokkema and Drenth (Vrije Universiteit, Amsterdam). It also appears that the moment a researcher came into contact with a foreign, often American, researcher in psychometrics this was to have a decisive influence on the former's attitude towards publication.

In many of the interviews this hypothesis was supported directly or indirectly. Roskam had his first paper in a Dutch journal in 1959, but published his first international papers about multidimensional scaling around 1970 after he had visited Coombs and Lingoes at the University of Michigan at Ann Arbor. This also lead to the well-known MINISSA computer program. Heiser claimed that the climate in Leiden in the 1970s was that results should be documented somehow, be it in internal reports and, if necessary, in chalk on a blackboard, but international publication somehow was not highly valued. This was corroborated by Van de Geer, who thought initially that his work was of interest only to a small community and that sending internal reports was a faster way of communication than publishing papers in journals. The contacts of De Leeuw with Carroll, Takane and Young in the 1970s certainly contributed to a change in the climate. He started publishing in international outlets in the second half of the 1970s. Molenaar visited Novick in 1977 and 1978 following Novick's visit to Groningen in 1976. At approximately that time he started publishing in international outlets on a regular basis. Several others also started to publish in international journals in the second half of the 1970s. Examples are Mellenberg and Van der Linden on decision making using educational test scores, and Ten Berge on factor analysis and reliability. Ten Berge, stimulated to publish by Hofstee, published his first international paper in 1972 (TEN BERGE, 1972).

Van der Linden and others noted the inspiration they drew from the courses Molenaar organized in the second half of the 1970s, starting with Novick in 1976. Other internationally well-known researchers such as Jöreskog (Sweden), Fischer (Austria) and Cox (England) were invited to teach about specialized topics, and for many Dutch researchers this was their first acquaintance with the international forum. Others had already visited the Nuffic conferences held in the 1960s. Another source of inspiration were the conferences on educational measurement in 1975 in Montreux and in 1977 in Leiden. The organizers Crombag, De Gruijter and Van der Kamp invited many researchers of great stature such as Cronbach, Fischer, Lord and Novick (see VAN DER KAMP, LANGERAK and DE GRUIJTER, 1980). Here Dutch researchers became acquainted with topics such as Bayesian statistics and generalizability theory. Paraphrasing Van der Linden: 'These encounters gave us (Dutchmen) the confidence that we could also play a role in the international arena'.

2.4 Why the Netherlands?

To many outside and also inside the Netherlands it is somewhat of a mystery why psychometrics could develop into such a strong discipline in this small country. The interviews gave some food for thought but did not really carry us far beyond speculation.

De Groot and Van der Kamp both mentioned the optimistic expectations about the contribution of psychology and other social science disciplines to post-war Dutch society, which led to the need for basic and applied research in areas such as educational and personnel selection. With this came the need for statistical methodology. These developments took place in universities and in other research institutions, such as research divisions of Philips Eindhoven, Nederlandse Spoorwegen (Netherlands Railways), PTT (nowadays KPN; this is the Dutch postal and communications services), and army, airforce and navy. One example that combines both is the advisorship of Van de Geer at the 'Instituut voor Zintuigfysiologie' (nowadays 'TNO Technische Menskunde'; TNO Human Factors Research Institute) in Soesterberg. Since the post-war requirement to have a well-educated population that could rebuild the country and its economy was far from unique for Dutch post-war society in comparison with other societies, however, it does not seem to be a sufficient explanation for the unusual development of Dutch psychometrics. But on the other hand, in the Dutch intellectual climate testing was much more acceptable than, for example, in France, where the emphasis in psychology was not on empirical research but on psychoanalysis.

Another reason for the growth of psychometrics is that by Dutch law methodology and statistics had to be part of a university training in psychology. This was already true in the early days of De Groot and Van de Geer. It may explain why the discipline of statistics grew much earlier and faster in psychology than in sociology where such official requirements did not exist. It may also have led to chairs in psychology departments for methodology, statistics and psychometrics. This would then have facilitated the training of students and the hiring of personnel. Another reason is the classification of psychology in five basic disciplines (DUIJKER, 1959), one of which was methodology. This may have given the discipline a justification from within psychology that was also acceptable for all those psychologists who had a totally different orientation in which 'numbers' did not fit in.

A final reason for the bloom of psychometrics could be the Dutch educational system in which all students at the educational level that prepares for academic training are brought into contact with mathematics at age 12 for a consecutive period of at least three years. In addition, all psychology training programs in the Netherlands require that the student has had a full mathematics (in particular, algebra) training for a period of six years. Students who stayed below this level are required to catch up. This climate guarantees a minimal preparedness and a possible affinity with statistics and psychometrics. Since we are not familiar with educational systems in other countries we find it hard to evaluate the validity of this argument.

Each of these reasons may hold a part of the truth but in retrospect it is hard to tell exactly why psychometrics was able to emerge in the Netherlands more powerfully than in most other countries. It is our impression that in the beginning a few individuals were motivated and inspired to start a new psychology much based on what they had read and seen with their own eyes of American psychology. Their decidedness and enthusiasm inspired a few others and this slowly changed the climate from contemplation to quantification in larger groups. This was probably facilitated by the necessity dictated by law to train students in elementary methodology and statistics. But why did it take until approximately 1980 before there was a sudden increase in the number of papers in international outlets? This is an interesting question because the intellectual switch had already been made in the 1960s and 1970s. As Heiser puts it: 'The level of psychometric thinking in the 1970s was not lower than in the 1980s. it was simply less visible in the journals.' He remembers that at the end of 1970s and the first years of the 1980s many Americans became interested in Europe and wanted to collaborate.

Perhaps part of the explanation is remote from educational goals and social ideals. Molenaar told us that in 1978 a faculty research committee investigated whether research was being done at the Faculty of Social Sciences in Groningen. Note that its mission was not to stimulate research. but to see if it existed. Paraphrasing his comments: 'Remember that the 1970s were coming to an end and that the government increased pressure on the universities to come up with results. Many of us felt that the times were changing.' This was also the moment that it was no longer self-evident that university budgets kept growing. This was due to decreasing economic growth which had already been slowing down considerably in the late 1970s. In other words. from that moment on Dutch universities were expected to develop into more performance-oriented organizations which could show a positive balance by the end of the year. Because psychometrics had already made the intellectual transition necessary for the new productive climate it could adapt so easily to the demands of higher productivity.

3 Part Two: A Citation Analysis for the Period 1980–1993

In the first part of the paper we have presented a report of the history of measurement and scaling in the Netherlands, with special attention to the period up to and including the seventies. We relied for a large part on the interviews that we held with a number of people who were important in this period. Starting around 1980, however, it became much more usual to find Dutch publications in international journals. An account of the work performed by Dutch social scientists in the area of measurement and scaling can be obtained from lists of publications and citations, which are available on CD-ROM. This makes it feasible to use a quantitative approach to describe the recent history.

In particular we would like to answer the question what the influence of Dutch researchers is, and has been, world-wide, in the area of measurement and scaling. The

generally held belief is that this influence has been large (for example, HAMBLETON and SWAMINATHAN, 1985). There are many indications that this belief is correct. To mention a few: one way would be to count the number of publications in Psychometrika, which is, together with Sociological Methods and Research, the journal with the highest impact factor in our area (impact factor is 1). MOLENAAR (1995) recently reported that, in the period 1970–1979 the Dutch share of papers of Psychometrika was 2%, but for the period 1983–1994 this percentage rose to 20%. Another possibility is to calculate the percentage of associate editors of journals in this field, and for Psychometrika this percentage is at the moment 20%. The Netherlands even had the honour to have Molenaar as editor of Psychometrika in the mid-eighties; the present editor of Psychometrika, Heiser, is also Dutch. Several Dutch psychometricians have been and presently are members of the editorial boards of Applied Psychological Measurement, Journal of Educational and Behavioral Statistics, and Psychometrika. De Leeuw was President of the Psychometric Society in 1987. A few interviewees remembered a session on Dutch contributions to psychological test theory ('Recent developments in test theory in the Netherlands') at the AERA conference of 1985 (Chicago) for which there was so much interest that the auditorium did not have enough seats. And last but not least, in a recent ranking of Dutch psychologists based on the number of publications, author position and quality of the journal, it turned out that the psychometricians Kiers, Ten Berge and Peter Molenaar were fifth, sixth and ninth (the source used was Psychological Abstracts; see KNIPPENBERG and VAN LUIJTELAAR, 1995, for details).

We studied the recent impact of the Dutch social sciences in measurement and scaling by means of a citation analysis. We divided the area of measurement and scaling into the fields of factor analysis, test theory including item response theory, latent class analysis, and optimal scaling and multidimensional scaling (see below for a more precise description). The basic idea of citation analysis is that the number of times a publication is cited is a measure of the impact of this publication. A citation can be considered as a quantitative judgement by one or more peers of the relevance of this publication.

The assessment of science by means of citation analysis is not beyond discussion. However, alternatives are not beyond discussion either. In a survey of the merits of the contributions of peer review and bibliometric methods to evaluate science, NEDERHOF (1988) concludes that "the outcomes of peer review are usually more than fifty per cent due to chance. It is much easier to cope with random error in bibliometric evaluation." He also states, after a summary of objections against using citations, that "many of the objections which have been raised against citation counts and indicators based upon citations, have little empirical basis, lack coherence, or can be met by using proper methods." (ibid., p. 206). Our own impression is that, although citation analysis is not very reliable and valid for the assessment of the impact of specific publications of individuals, more reliable measures can be obtained if citations of many publications are added up. This is what we did: we did not want to assess the impact of specific publications (although we present the top 25 of

publications later to satisfy the curiosity of the reader), but rather of four fields in the area of measurement of scaling. We believe that citation analysis can provide a good impression of the Dutch impact in this field. Note that we did not aim to evaluate individuals or research groups. For a discussion of the validity of the impact-measures that we define below, see MOED et al. (1992). We agree with these authors that these measures should all be used with care. They concluded that "The quality or international position of a group may be an important factor in determining the publication output and citation rate of a group, but it is definitely not the only factor". For introductions, more details and a discussion of the validity and usefulness of measures used in citation analyses, see NEDERHOF (1988) and MOED et al. (1992). For a discussion of citation patterns in the journals of statistics and probability, and the difficulties encountered in coming to valid conclusions, see STIGLER (1994). We will now first describe the methods employed, and then discuss the results.

3.1 Methods

Except for the results in Table 4, all citation analyses were performed by the Centre for Science and Technology Studies (CWTS, Leiden University). The citation counts were made from citation indices constructed by the Institute for Scientific Information (ISI) in Philadelphia, USA. Data bases used were three citation indices: the Social Science Citation Index (SSCI), the Science Citation Index (SCI), and the Arts and Humanities Citation Index (A&HCI).

The procedure to go from publications to citation analyses had some arbitrary elements that we will describe next. We think, however, that this did not seriously damage the impact of the four fields that we discerned.

The steps we took were:

1. The CWTS ordered from the ISI a database of journal citation data of all publications that originated in the Netherlands between 1980 and 1993 (included). Several sources of error have been eliminated from this database (see MOED et al., 1992). We specified journals that had to serve as a reference for the citation analysis. A problem was that we wanted to include Applied Measurement in Education, Computational Statistics Quarterly, Methodika, and Statistica Neerlandica, but these are not ISI-journals, and therefore they are not included. The social science journals that served as a base in the analyses are, in alphabetical order, Applied Psychological Measurement, British Journal of Mathematical and Statistical Psychology, Educational and Psychological Measurement, Journal of Educational Statistics, Journal of Educational Measurement, Journal of Mathematical Psychology, Multivariate Behavioural Research, Psychological Bulletin, Psychologische Beiträge, Psychometrika, Quality & Quantity, Sociological Methods & Research, and Sociological Methodology (not all volumes). The non-social science journals included are Applied Statistics, Computational Statistics & Data Analysis, Econometrica, Journal of the American Statistical Association, Journal of Classification, Journal of Econometrics and Journal of Multivariate Analysis.

2. The CWTS next provided a list of all publications (only 'articles', 'letters', 'notes' and 'reviews', but usually not 'comments' and publications in 'computational sections') that were published in these journals in the period chosen, where at least one of the authors provided a Dutch address. This implies, for example, that publications written by Lewis while he worked in Groningen are part of our data base, whereas publications by de Leeuw written while he worked at UCLA are not part, if he does not have a co-author working in the Social Sciences in the Netherlands.
3. Using the list, we skipped
—all publications that do not have at least one author working in a Dutch Social Science Faculty, Cito or RPD.Advies (the former RPD; a service that assesses personnel for the Dutch civil service); so this paper does not discuss the contribution of, for example, neuropsychology, econometrics and CBS.
—all publications that are not in the area of measurement of individuals and scaling of stimuli.
Only in a few cases we checked the contents of papers, but most of the time we judged the contents from the title or from knowledge of the contents of a paper. We classified publications into one of the four fields:
i. FA (Factor Analysis); includes topics such as structural equation modelling, (three-mode) principal component analysis, facet theory, and rotation;
ii. TT (Test Theory, including item response theory); includes topics such as test design, decision theory, agreement, and (confirmatory) unfolding;
iii. LCA (Latent Class Analysis); includes topics such as latent markov models and mixture models;
iv. OS (Optimal Scaling, including multidimensional scaling); includes optimal scaling versions of principal component and factor analysis, including correspondence analysis, and multidimensional scaling of (dis)similarities, and (exploratory) unfolding.
When publications could be classified in more than one field (for example, dealing with a comparison of IRT and factor analysis), a decision was made by flipping a coin. This was done only a few times.
4. The result of this classification was analyzed by the CWTS. For each of the publications on this list the CTWS provided the number of citations from all ISI journals (not only the journals in 1). A 'variable citation window' was used, i.e. the citation window was related to the year in which publications appeared (for a publication in 1985 the window is 9 years).
5. Given these results we performed a further citation analysis to study whether the influence of Dutch publications grows or diminishes (trend analysis). For this analysis we chose a so-called 'fixed citation window' of five years, including the year that the publication appeared. That is, for a publication published in 1985, the number of citations in 1985 up to and including 1989 was counted. Thus we controled for the fact that publications that are older can be cited during a longer time. A drawback is that publications published in 1990 and later are not included in this analysis because the citation window is based on five years. When studying annual

results, it should be noted that sometimes a publication published in, for example, a journal's last issue of 1985 will be counted in subsequent analyses (see 4) under 1986 if this is the year the publication appears in the SSCI.

6. A serious drawback of the above undertaking is that only the impact of publications in journals is studied (see for example NEDERHOF et al., 1993). This bothered us especially in the field of optimal scaling where many papers appeared as chapters in books. To remedy this, we therefore collected a sample of books and papers in books that seemed important to us, and counted the number of citations from the paper version of the SSCI.

In the citation analysis the following measures were calculated by the CWTS (see also MOED et al., 1992):

—*P* is the number of publications;

—*C* is the number of times the *P* publications are cited;

—*CPP* is the mean number of citations per publication, calculated as *C*/*P*. This is a measure for the mean impact of a group of publications;

—*CPPex* is the mean number of citations, self-citations excluded. This measure shows the mean impact of a group of publications on researchers different from the authors;

—*JCSm* is the mean Journal Citation Score. It can be used as a baseline for comparison of performance of a publication in the same journal in the same year. Thus in this paper we can compare Dutch publications with publications world-wide. *JCSm* is calculated as follows. First, publications are disaggregated into normal articles, letters, notes, and reviews. Consider a normal article in Psychometrika in 1990. Assume that in 1990 Psychometrika has 80 normal articles that lead in 1990–1993 to 200 citations. Then for the normal articles from Psychometrika 1990 *JCSm* = 200/80 = 2.5. Thus the typical normal article in Psychometrika in 1990 has a *JCS* of 2.5. Assume now that there is a group of five publications, consisting of two 'articles' published in Psychometrika in 1990 and three 'notes' in Applied Psychological Measurement in 1991 (with a *JCSm* of 1.4 for the period 1991–1993). Then the estimated *JCS* value of that set is $2 \times 2.5 + 3 \times 1.4 = 9.2$, and the *JCSm* for a publication, the 'expected impact', is 9.2/5 = 1.84). The CWTS has JCSm scores for most of the ISI journals. These JCSm scores are calculated using 'variable citation windows', and including self citations, because expected impacts are only available including self-citations (however, this percentage was checked by the CWTS, and it turned out that in our fields self-citations did not occur significantly more often than in world-wide publications).

A criticism of this measure is that, by comparing the impact of publications with the impact of all comparable publications in the same journal, a comparison of *CPP* with *JCSm* ignores the level of the journals in which the publications of a particular field are published. Thus publishing in prestigious journals is not encouraged. For this reason the next measure is used:

—*FCSm* is the mean Field Citation Score. FCSm can be used as a baseline for comparison of performance of a publication in the same ISI field in the same year. It

Table 1. Bibliometric Statistics for the total year period 1980–1993; abbreviations are explained in the Method Section

	P	*C*	*CPP*	*CPPex*	*JCSm*	*FCSm*
FA	78	431	5.5	4.3	9.1	5.9
TT	128	575	4.5	3.4	4.4	4.3
LCA	6	14	2.3	2.0	2.3	2.5
OS	29	102	3.5	2.5	3.9	3.8
Total	241	1122	4.6	3.5	5.8	4.7

is the mean number of citations received by all articles published in specific ISI fields. The ISI has grouped journals on about 150 topics, such as 'economics', 'business, finance', 'sociology', and 'history'. *FCSm* is calculated similar to *JCSm*.

3.2 Results

The key findings are summarized in Table 1. We were surprised by the low number of articles in LCA and OS. For OS this low number may be due to relatively many publications appearing as internal reports, publications in books, articles in non-ISI journals, or ISI-journals not in our list. For LCA our impression is that in the early 1980s only Mooijaart worked in this field, but his interest shifted to structural equations models when, during a stay in Uppsala, nobody showed much interest in the topic. The (Dutch) thesis of Hagenaars was published in 1985 (its translated version appeared as a book in 1990, see below), and after this his group in Tilburg started to become active.

The 241 publications in total were cited 1122 times. All publications cited at least 10 times (self-citations excluded) are listed in Table 2. The paper by Mokken and Lewis on Mokken-scale analysis as it was later called is cited 48 times. Then there are two papers on covariance structure analysis, one by Boomsma and one by Satorra and Saris. De Leeuw's name appears 6 times in Table 2 in three of the four fields. For the rest many names are encountered that already appeared in the first part of the paper. In the top-25 there are 13 IRT papers, 9 FA papers, and 3 OS-papers. We leave a further interpretation of the results to the reader.

Being cited is not trivial. Of the 164 papers that were published before 1990 (these are the papers that have had some time to 'generate' citations) 35 papers were never cited, 29 papers once, 27 twice, 11 thrice, and then, from four to nine times cited, the number of papers was 13, 7, 4, 8, 3, and 2 (for papers being cited at least 10 times, see Table 2).

Although means are not good measures to describe distributions when the distributions are as skewed as here, we go on with a discussion of CPPex. On average FA-papers were cited most often (4.3 times). For FA a typical paper in the same journal is cited 9.1 times, but in the journal group 5.9 times. Therefore, FA-papers were published in the higher-impact-journals of the journal group, but the number of citations of the FA-papers published is approximately equal to the expected number of citations in the journal group (5.5 versus 5.9). For TT the papers were cited

Table 2. Papers cited at least 10 times, self-citations excluded. Area codes are TT (test theory), FA (factor analysis) and OS (optimal scaling)

TT	48	MOKKEN, R. J. and C. LEWIS (1982), A nonparametric approach to the analysis of dichotomous item responses, *Applied Psychological Measurement*, **6**, 417–430.
FA	43	BOOMSMA, A. (1985), Nonconvergence, improper solutions, and starting values in LISREL maximum-likelihood estimation, *Psychometrika*, **50**, 229–242.
FA	39	SATORRA, A. and W. E. SARIS (1985), Power of the likelihood ratio test in covariance structure analysis, *Psychometrika*, **50**, 83–90.
TT	32	VAN DEN WOLLENBERG, A. L. (1982), Two new test statistics for the Rasch model, *Psychometrika*, **47**, 123–140.
FA	31	KROONENBERG, P. M. and J. DE LEEUW (1980), Principal component analysis of 3-mode data by means of alternating least-squares algorithms, *Psychometrika*, **45**, 69–97.
FA	27	TEN BERGE, J. M. F. (1986), Rotation to perfect congruence and the cross-validation of component weights across populations, *Multivariate Behavioral Research*, **21**, 41–64.
TT	25	MOLENAAR, I. W. (1983), Some improved diagnostics for failure of the Rasch model, *Psychometrika*, **48**, 49–72.
OS	24	VAN DER HEIJDEN, P. G. M. and J. DE LEEUW (1985), Correspondence-analysis used complementary to loglinear analysis, *Psychometrika*, **50**, 429–447.
TT	23	KELDERMAN, H. (1984), Loglinear Rasch model tests, *Psychometrika*, **49**, 223–245.
FA	20	PIETERS, J. P. M. (1983), Sternberg's additive factor method and underlying psychological processes—some theoretical considerations, *Psychological Bulletin*, **93**, 411–426.
TT	18	THEUNISSEN, T. J. J. M. (1985), Binary programming and test design, *Psychometrika*, **50**, 411–420.
FA	15	ARRINDELL, W. A. and J. VAN DER ENDE (1985), An empirical test of the utility of the observations-to-variables ratio in factor and components-analysis, *Applied Psychological Measurement*, **9**, 165–178.
FA	14	TAKANE, Y. and J. DE LEEUW (1987), On the relationship between item response theory and factor analysis of discretized variables, *Psychometrika*, **52**, 393–408.
TT	13	STOKMAN, F. and W. VAN SCHUUR (1980), Basic scaling, *Quality & Quantity*, **14**, 5–30.
TT	13	ZEGERS, F. E. and J. M. F. TEN BERGE (1985), A family of association coefficients for metric scales, *Psychometrika*, **50**, 17–24.
TT	12	TEN BERGE, J. M. F., T. A. B. SNIJDERS and F. E. ZEGERS (1981), Computational aspects of the greatest lower bound to the reliability and constrained minimum trace factor analysis, *Psychometrika*, **46**, 201–213.
TT	12	VAN DER LINDEN, W. J. and E. BOEKKOOI-TIMMINGA (1989), A maximin model for test design with practical constraints, *Psychometrika*, **54**, 237–247.
OS	12	TAKANE, Y., F. W. YOUNG and J. DE LEEUW (1980), An individual-differences additive-model—an alternating least-squares method with optimal-scaling features, *Psychometrika*, **45**, 183–209.
FA	11	KELDERMAN, H., G. J. MELLENBERGH and J. J. ELSHOUT (1981), Guilford's facet theory of intelligence—an empirical comparison of models, *Multivariate Behavioral Research*, **16**, 37–61.
OS	11	VAN DER HEIJDEN, P. G. M., A. DE FALGUEROLLES and J. DE LEEUW (1989), A combined approach to contingency table analysis using correspondence-analysis and log-linear analysis, *Applied Statistics*, **38**, 249–292.
TT	11	MOLENAAR, I. W. (1981), On Wilcox's latent structure model for guessing, *British Journal of Mathematical and Statistical Psychology*, **34**, 224–228.
FA	10	MOLENAAR, P. C. M. (1985), A dynamic factor model for the analysis of multivariate time-series, *Psychometrika*, **50**, 181–202.
TT	10	MELLENBERGH, G. J. and P. VIJN (1981), The Rasch model as a loglinear model, *Applied Psychological Measurement*, **5**, 369–376.
TT	10	VAN DER LINDEN, W. J. (1982), A latent trait method for determining intrajudge inconsistency in the Angoff and Nedelsky techniques of standard setting, *Journal of Educational Measurement*, **19**, 295–308.
TT	10	DE LEEUW, J. and N. D. VERHELST (1986), Maximum-likelihood estimation in generalized Rasch models, *Journal of Educational Statistics*, **11**, 183–196.

Table 3. Number of papers published from 1980 to 1993 in the four fields

	80	81	82	83	84	85	86	87	88	89	90	91	92	93
FA														
# publ.	5	2	2	5	7	11	6	6	2	8	1	11	8	4
# citat.	18	6	5	16	11	67	18	16	0	12				
TT														
# publ.	3	8	5	8	11	9	15	11	7	7	16	8	16	4
# citat.	8	23	24	40	27	13	40	17	6	16				
LCA														
# publ.	0	0	0	0	0	0	0	1	1	0	1	1	1	1
# citat.	0	0	0	0	0	0	0	3	3	0				
OS														
# publ.	1	0	2	0	3	1	3	1	2	1	2	3	2	8
# citat.	3	0	2	0	2	0	20	1	3	11				

approximately as often as the typical paper in the same journal or in the same journal group (4.5 versus 4.4 and 4.3). Similar results hold for LCA and OS.

The conclusion from Table 1 is that, for all four fields, the impact of Dutch papers is approximately equal to the impact of other papers (i.e. having a world-wide origin) in the same journal or journal group. Dutch researchers not only publish a lot, but what they publish can compete with comparable publications produced world-wide.

The results of a trend analysis, using a citation window of five years, can be found in Table 3. The number of publications in our ISI-journal group in the field FA seems to be more or less stable over time, whereas the number of publications in the other three fields seems to increase. Using a five-year citation window the number of citations of the publications are also reported in Table 3. No clear trend in impact can be discerned in the number of citations in any of the fields. This also holds for the JCSm and the FCSm measures, which we do not report here.

The three important researchers in the early history of the field that we identified in Part One were De Groot, Van de Geer and Mokken. In Table 4 the number of citations of one book of each of them is reported on a yearly basis. The book "Methodology" of De Groot was cited 57 times. Mokken's book on his scaling model was cited 151 times, and it is still often cited. The book of Van de Geer on linear multivariate analysis published in 1971 was cited 257 times!

Kroonenberg's thesis was cited 52 times, and Saris and Stronkhorst's book (sometimes misspelled in the SSCI as "Casual modelling ... .") was cited 56 times. Hagenaars' recently published book is cited 11 times. The internal report by GIFI (1981) was cited 103 times. The publication trend of GIFI (1990) also looks promising.

All in all, we think that the publications in Table 4 show for the OS-field that the impact has been impressive too, and is much larger than is suggested by Tables 1 and 2. The impact of GIFI (1990) will probably grow larger due to the recent inclusion of programs such as HOMALS in SPSS.

Table 4. Citations (including self-citations) of some books and papers in books, until 1993

De Groot, A. D. (1961), *Methodologie* and (1969) *Methodology*, Mouton, Den Haag. Starting 1969: 1, 3, 5, 2, 3, 5, 1, 2, 4, 2, 2, 4, 2, 2, 1, 0, 0, 1, 1, 1, 4, 3, 2, 4, 2 (total 57).

De Gruijter, D. N. M. and L. J. T. van der Kamp (1984), *Statistical models in psychological and educational testing*, Swets and Zeitlinger, Lisse, 0, 2, 3, 1, 0, 1, 0, 0, 0, 0, (total 7).

De Leeuw, J. and W. Heiser (1980), Multidimensional scaling with restrictions on the configuration, in: P. R. Krishnaiah (ed.), *Multivariate analysis V*, North Holland, Amsterdam, 2, 0, 2, 3, 4, 2, 2, 3, 2, 4, 2, 2, 2, 7 (total 37).

De Leeuw, J. and W. Heiser (1982), Theory of multidimensional scaling, in: P. R. Krishnaia and L. N. Kanal (eds.), *Handbook of Statistics 2*, North Holland, Amsterdam, 0, 1, 3, 1, 1, 0, 1, 0, 1, 1, 1, 1 (total 11).

Gifi, A. (1981), *Non-linear multivariate analysis*, Department of Data Theory, Leiden, 1, 3, 10, 8, 5, 13, 12, 15, 16, 13, 6, 1, 0 (total 103).

Gifi, A. (1990), *Non-linear multivariate analysis*, Wiley, New York, 1, 7, 11, 16 (total 35).

Hagenaars, J. A. (1990), *Categorical longitudinal data*, Sage, Los Angeles, 0, 2, 5, 4 (total 11).

Heiser, W. J. (1981), *Unfolding analysis of proximity data*, Unpublished PhD thesis, Department of Data Theory, Leiden, 0, 0, 1, 2, 3, 2, 2, 5, 1, 4, 1, 1, 0 (total 22).

Kroonenberg, P. M. (1983), *Three-mode principal component analysis*, D.S.W.O.-Press, Leiden, 1, 1, 4, 6, 3, 3, 9, 8, 10, 4, 3 (total 52).

Mokken, R. J. (1971), *A theory and procedure of scale analysis*, Mouton, Den Haag, De Gruijter, Berlin. 0, 0, 3, 2, 3, 4, 3, 4, 1, 7, 7, 10, 9, 9, 10, 13, 11, 7, 13, 9, 8, 8, 10 (total 151).

Saris, W. E. and L. H. Stronkhorst (1984), *Causal modelling in nonexperimental research: an introduction to the LISREL approach*, SRF, Amsterdam, 0, 3, 7, 1, 7, 12, 14, 2, 4, 6 (total 56).

Van de Geer, J. P. (1971), *Introduction to multivariate analysis for the social sciences*, Freeman, San Francisco, 0, 1, 9, 11, 19, 22, 17, 24, 22, 20, 24, 14, 15, 9, 15, 8, 5, 6, 7, 4, 1, 1, 3 (total 257).

Van Rijckevorsel, J. L. A. and J. de Leeuw (eds.) (1988), *Component and correspondence analysis: dimension reduction by functional approximation*, Wiley, New York, 0, 2, 2, 0, 1, 0 (total 5).

4 Conclusion

The publication and citation analysis has supported two informal hypotheses that we have frequently heard mentioned in the past years. The first is that during the last 15 years Dutch authors have published a relatively high number of papers on psychometric topics in the international journals. The second is that the impact of Dutch papers is considerable. More specifically, the mean impact of these papers matches the international standards in the field.

A few final remarks may give an indication of the present state of affairs. First, we think that it is a fortunate development that there is a growing interest among psychometricians to produce user-friendly computer programs for their methods. We believe that such programs are a necessary condition to reach a considerable impact, not only among peers, but also among researchers in psychology, sociology, political science and education. Much of the programming effort and distribution is the hands of the interuniversity software house ProGAMMA located at the university of Groningen. In addition, we mention local initiatives such as by the group of Hagenaars (Tilburg) with respect to LCA, by Verhelst and others at Cito with respect to IRT, and by the Department of Data Theory with respect to OS.

Second, there is a number of successful forums in which Dutch psychometricians discuss their latest research. Internationally oriented are the yearly workshops on

IRT in Twente. Since the late 1970s the Psychometric Society has a European Meeting every two years. This conference was organized in Groningen in 1981, in Twente in 1987, and in Leiden in 1995.

Third, the ongoing development of psychometrics has lead to the foundation of the 'Interuniversitair Onderzoeksinstituut voor Psychometrie en Sociometrie' (IOPS: Interuniversity Graduate School of Psychometrics and Sociometrics) in 1987. IOPS was the first initiative to found a postgraduate school in the Dutch social sciences. In 1994 IOPS was officially recognized by the Royal Netherlands Academy of Arts and Sciences (KNAW) as a postgraduate school. At the moment, departments from six universities participate in IOPS and researchers from departments of other universities participate as associate members. The foundation and recognition of IOPS shows, together with many other accomplishments, that Dutch psychometrics has indeed reached the stage of full maturity.

There are also some developments that might lead to a decline. First, because of the shortening of universitary studies, the number and level of graduates in quantitative methods in the social sciences is declining: they simply do not have the time to choose this specialization, and the time to specialize is limited. A second point mentioned by a few interviewees is that the level of sophistication in quantitative methods is now so high that it has become very difficult for non-specialists to understand what is going on in our field.

It will be clear that our description of the history can be criticized. Our citation analysis can be criticized because the choices that we made are sometimes a bit arbitrary (see Section 3.1). Someone asked us why we did not count citations of individuals. This was not done because it would lead to an assessment of persons, and perhaps research institutions, and we have chosen to evaluate fields instead (some people worked in more than one field).

We have ignored that econometricians were internationally quite active for a long time. We were not able to trace the influence of this group on the social sciences. Our impression is that interest in measurement in econometrics (i.e. latent variables) is quite recent, and the focus has been on regression models.

Acknowledgements

The authors gratefully acknowledge the cooperation of J. M. F. ten Berge, T. Dehue, J. P. van de Geer, A. D. de Groot, D. N. M. de Gruijter, J. A. Hagenaars, H. 't Hart, W. J. Heiser, L. J. Th. van der Kamp, J. de Leeuw, W. J. van der Linden, G. J. Mellenbergh, R. J. Mokken, I. W. Molenaar, A. Mooijaart, A. J. Nederhof, A. F. J. van Raan, J. L. A. van Rijckevorsel, E. E. Ch. I. Roskam, and N. D. Verhelst. We are to blame for all remaining flaws in this text.

References

CROMBAG, H. F. M., L. J. TH. VAN DER KAMP and C. A. J. VLEK (eds.) (1987), *De psychologie voorbij: ontwikkelingen rond model, metriek en methode in de gedragswetenschappen*, Swets and Zeitlinger, Lisse.

CRONBACH, L. J. and G. C. GLESER (1957, 1965), *Psychological tests and personnel decisions*, University of Illinois Press, Urbana.

DEHUE, T. (1990), *De regels van het vak. Nederlandse psychologen en hun methodologie 1900–1985*, van Gennep, Amsterdam. Recent English translation (1995), *Changing the Rules: Psychology in the Netherlands 1900–1985*, Cambridge University Press, Cambridge.

DE LEEUW, J. (1988), Models and techniques, *Statistica Neerlandica* **42**, 91–98.

DRENTH, P. J. D. (1965), *De psychologische test*, Van Loghum Slaterus, Arnhem.

DRENTH, P. J. D. (1975), *Inleiding in de testtheorie*, Van Loghum Slaterus, Deventer.

DUIJKER, H. C. J. (1959), Nomenclatuur en systematiek der psychologie, *Nederlands Tijdschrift voor de Psychologie* **14**, 176–217.

EDGEWORTH, F. Y. (1888), The statistics of examinations, *Journal of the Royal Statistical Society* **51**, 599–635.

FISCHER, G. H. (1974), *Einführung in die Theorie psychologischer Tests*, Huber, Bern.

GIFI, A. (1990), *Nonlinear multivariate analysis*, Wiley, New York.

GULLIKSEN, H. (1950), *Theory of mental tests*, Wiley, New York.

GUTTMAN, L. (1944), A basis for scaling qualitative data, *American Sociological Review* **9**, 139–150.

GUTTMAN, L. (1950), The basis for scalogram analysis, in: S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star and J. A. Clausen (eds.), *Measurement and prediction*, Princeton University Press, Princeton.

HAMBLETON, R. K. and H. SWAMINATHAN (1985), A look at psychometrics in the Netherlands, *Nederlands Tijdschrift voor de Psychologie* **40**, 446–451.

KOUWER, B. J. (1952), *Test in de psychologische praktijk*, Erven J. Bijleveld, Utrecht.

KNIPPENBERG, H. and G. VAN LUIJTELAAR (1995), Een methode voor het identificeren van topauteurs in de Nederlandse psychologie, *De Psycholoog* **30**, 8–13.

KRUSKAL, J. B. (1964a), Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis, *Psychometrika* **29**, 1–27.

KRUSKAL, J. B. (1964b), Nonmetric multidimensional scaling: a numerical method, *Psychometrika* **29**, 115–129.

LAZARSFELD, P. F. (1950), The logical and mathematical foundation of latent structure analysis, in: S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star and J. A. Clausen (eds.), *Measurement and prediction*, Princeton University Press, Princeton.

LUIJTEN, T. (1993), *Het Cito tussen Schiermonnikoog en Maastricht*, Internal report of Cito, Arnhem.

MEEHL, P. E. (1954), *Clinical versus statistical prediction*, University of Minnesota Press, Minneapolis.

MEERLING (1980), *Methoden en technieken van psychologisch onderzoek Deel 1 en 2*, Boom, Meppel.

MELLENBERGH, G. J. (1971), *Studies in studietoetsen*, Psychologisch Laboratorium, Universiteit van Amsterdam (dissertation), Amsterdam.

MELLENBERGH, G. J. (1984), Van zielsmeting bezield, *Tijdschrift voor Onderwijsresearch*, **9**, 253–256.

MOED, H. F., R. E. DE BRUIN, J. BEERENS, J. A. SCHONEVELD and I. VRIEND (1992), A bibliometric system for the assessment of national research performance: database description and first application, *Report CWTS-92-03*, University of Leiden, CWTS, Leiden.

MOKKEN, R. J. (1971), *A theory and procedure of scale analysis*, Mouton, Den Haag, De Gruijter, Berlin.

MOKKEN, R. J. and C. LEWIS (1982), A nonparametric approach to the analysis of dichotomous item responses, *Applied Psychological Measurement* **6**, 417–430.

MOLENAAR, I. W. (1987), Formal statistics and informal data analysis, or why laziness should be discouraged, *Statistica Neerlandica* **42**, 83–90.

MOLENAAR, I. W. (1995), Statistiek en samenleving en het meten van mensen, *Kwantitatieve Methoden* **16**, 77–92.

MOLENAAR, I. W., P. DEBETS, K. SIJTSMA and B. T. HEMKER (1994), *User's manual MSP*, iecPROGAMMA, Groningen.

NEDERHOF, A. J. (1988), The validity and reliability of evaluation of scholarly performance, in: A. F. J. van Raan (ed.), *Handbook of quantitative studies of science and technology*, Elseviers, Amsterdam, 193–228.

NEDERHOF, A. J., R. F. MEIJER, H. F. MOED and A. F. J. VAN RAAN (1993), Research performance indicators for university departments. A study of an agricultural university, *Scientometrics* **27**, 157–178.

RASCH, G. (1960), *Probabilistic models for some intelligence and attainment tests*, Nielsen and Lydiche, Copenhagen.

ROSKAM, E. E. CH. I. (1968), *Metric analysis of ordinal data in psychology*, Drukkerij/Uitgeverij VAM, Voorschoten.

SNIJDERS-OOMEN, A. W. M. (1943), *Intelligentie onderzoek van doofstomme kinderen*, Berkhout, Nijmegen.

SNIJDERS, J. TH. AND N. SNIJDERS-OOMEN (1958), *Niet-verbaal intelligentieonderzoek van horenden en doofstommen: Snijders-Oomen niet-verbale intelligentieschaal SON*, Wolters, Groningen.

SPEARMAN, C. (1910), Correlation calculated from faulty data, *British Journal of Psychology* **3**, 271–295.

STIGLER, S. M. (1994), Citation patterns in the journals of statistics and probability, *Statistical Science* **9**, 94–108.

SWANBORN, P. G. (1995), A personal history of scaling techniques, in: J. J. Hox and W. Jansen (eds.), *Measurement problems in social and behavioural research*, SCO, Amsterdam.

TEN BERGE, J. M. F. (1972), Difficulty factors, distribution effects, and the least squares simplex data matrix solution, *Educational and Psychological Measurement* **32**, 911–920.

TEN BERGE, J. M. F. (1977), *Optimizing factorial invariance*, Unpublished doctoral dissertation, University of Groningen, Groningen, The Netherlands.

THURSTONE, L. L. (1927), A law of comparative judgement, *Psychological Review* **34**, 273–286.

VAN DE GEER, J. P. (1971), *Introduction to multivariate analysis for the social sciences*, Freeman, San Francisco.

VAN DE GEER, J. P. and TH. BEZEMBINDER (1988), In memoriam Clyde H. Coombs 1912–1988, *Nederlands Tijdschrift voor de Psychologie* **43**, 259–262.

VAN DEN WOLLENBERG, A. L. (1979), *The Rasch model and time-limit tests*, Stichting Studentenpers Nijmegen, Nijmegen.

VAN DEN WOLLENBERG, A. L. (1982a), Two new test statistics for the Rasch model, *Psychometrika* **47**, 123–140.

VAN DEN WOLLENBERG, A. L. (1982b), A simple and effective method to test the dimensionality axiom of the Rasch model, *Applied Psychological Measurement* **6**, 83–91.

VAN DER KAMP, L. J. TH., W. F. LANGERAK and D. N. M. DE GRUIJTER (1980), *Psychometrics for educational debates*, Wiley, New York.

VAN HEERDEN, J. and J. HOOGSTRATEN (1984), Gesprek met prof. dr. R. F. van Naerssen, *Tijdschrift voor Onderwijsresearch* **9**, 201–206.

VAN NAERSSEN, R. F. (1962), *Selectie van chauffeurs: onderzoekingen ten behoeve van de selectie van chauffeurs bij de Koninklijke Landmacht*, Wolters, Groningen.

VAN NAERSSEN, R. F. (1965), Application of the decision-theoretical approach to the selection of drivers, in: L. J. Cronbach and G. C. Gleser (eds.), *Psychological tests and personnel decisions*, University of Illinois Press, Urbana.

Received: April 1995. Revised: August 1995.