Netherlands Official Statistics

Volume 15, Winter 2000

Publisher
Statistics Netherlands
R.L. Vellekoop
Prinses Beatrixlaan 428
2273 XZ Voorburg

Printed by
Statistics Netherlands
Facility Services

Information
Phone: # (31) 45 570 70 70
Fax: # (31) 45 572 62 68
E-mail: infoservice@cbs.nl

Where to order
E-mail: verkoop@cbs.nl

Internet
www.cbs.nl

Prices do not include postage and
administration costs.
Annual subscription: NLG 42.00 (€ 19.06)
per year
Price per copy: NLG 20.00 (€ 9.08)
Key-figure: A-125/2000
ISSN 0920-2048

Statistics Netherlands

Contents

The best of two worlds: total design method and new Kontiv design. An operational model to improve respondent co-operation

Ger Snijkers and Martin Luppes

Abstract

In 1999 a programme was started at Statistics Netherlands to increase response rates for self-administered business surveys by improving the communication with respondents. The traditional approaches based on simple communication strategies (one stimulus for all units at the same moment, traditional reminder approaches using authority principles) give serious problems with response time, net response and response quality. The goal of the new programme is to develop measures that change the traditional, formal and passive contact strategies into active, respondent-driven and motivational approaches.
In the past several measures have already been implemented, like a cognitive laboratory to improve the wording of advance letters and questionnaires, and a guide for form-design standards to improve the layout of letters and questionnaires.
This programme is based on the philosophies of Dillman (Total Design Method) and Brög (New Kontiv Design). Where the NKD design allows for undefined respondent behaviour, the TDM approach is based on the more general standardised survey approach of well-defined respondent behaviour. In fact, both TDM and NKD are quite similar in their (respondent driven) paradigms, but quite different in their operational approach. And both have proven their benefits. Therefore the challenge is to create strategies and tactics that incorporate the best of both perspectives.
Key words: response burden, business surveys, communication strategy.

1. Introduction

Statistics Netherlands has carried out several projects to improve different parts of the communication strategies in self-administered establishment surveys [1)]. These projects are aimed at improving the quality of advance letters and the effects of active strategies on response time and at the reduction of reminder calls. The success of the New Kontiv Design in the Dutch Mobility Survey (see section 3.2), which is a household survey using an active, respondent-oriented communication strategy to raise the low response rates [2)], triggered the idea of using similar approaches and tactics in establishment surveys.

Communication with respondents includes many closely interrelated aspects. However much effort is put into the planning of the survey process, the sampling and the standardisation of the questionnaires, if not enough attention is given to the position of the respondent, the form and design of the measurement instruments to be used and (most of all) what is needed to motivate respondents, the result will be low response rates and a relatively high number of time-consuming reminder calls (Dillman, 1978; Brög, 1997; Moritz and Brög, 1999). A good communication strategy should incorporate all these aspects, thus resulting in a consistent set of recommendations, procedures and instruments.
The goal of active respondent communication is twofold. First, seen from the respondent's point of view costs, time, effort and number of people involved have to be minimised. But seen from the point of view of our stakeholders and customers, statistical information should be adequate, accurate, and delivered on time. It is clear that there is a tension between the customer demand and respondent's willingness. Therefore we have developed a respondent-oriented approach in which both interests are represented and balanced, based on the Total Design Method (Dillman, 1978) and the New Kontiv Design (Brög, 1997).

In this article we present the outline of this respondent-oriented communication strategy for business surveys. This strategy is based upon the insights acquired from several projects in the area of establishment surveys in the past years, as well as the work done in the area of social surveys. In section 2 we give a short review on the assumptions of the traditional and modern respondent approaches in surveys. Section 3 presents some results of case studies on respondent-oriented communication. Based on these results, as well as the TDM and the NKD approach, in section 4 we give an outline of measures to be taken to improve survey participation in business surveys. In section 5 we present conclusions and some issues for further discussion.

2. The assumptions of survey design reviewed

Traditionally, survey design follows the 'one size fits all' strategy, which means that one design is used with one instrument for the whole sample. This approach is based on the idea that in order to get comparable response over the units, the units should receive a standardised stimulus, i.e. a stimulus in which environmental factors are controlled for. Standardisation – and not the respondent – is central in the approach, and by the same token the orientation is on the process. The respondent is in fact a 'standard' respondent (all respondents can be approached in similar standardised ways without taking into account their specific situations), who is sensitive to authority. Elements of this passive approach are: mandatory participation, based on authority and tradition, one – usually paper – questionnaire for all respondents, sent out at the same time, and no contact with respondents until the deadline has passed.

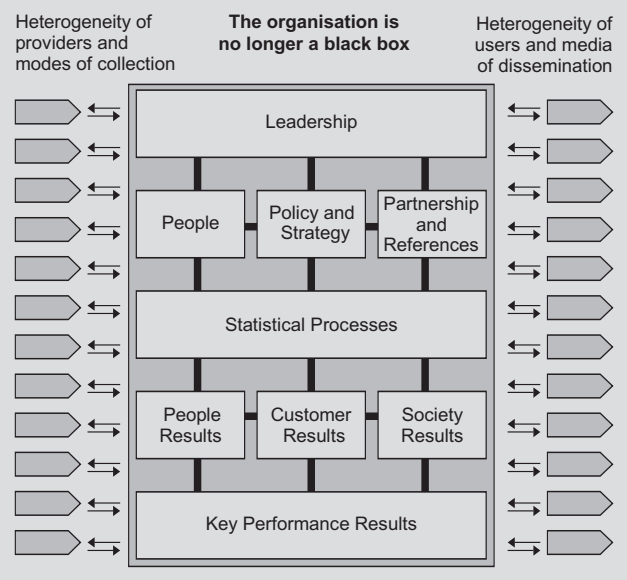
The 'one size fits all' survey approach was improved by the Total Design Method introduced by Dillman in 1978. The TDM approach is based on the premises of social exchange theory in which the compliance principle of reciprocity is used, not the principle of authority (Groves, Cialdini and Couper, 1992). The TDM assumed that maximum response could be generated by rewarding the respondent, reducing the costs for the respondent, and by establishing a relationship built on trust. This could be achieved by a coherent system using standardised, well-tested questionnaires and an appropriate and user friendly design (including the advance letters, brochures and other enclosures), based on carefully worded and respectfully formulated requests and instructions. This approach was applied to both mail and telephone interviews. Although this approach is respondent oriented, the basic assumption was still the 'one size fits all' strategy.

In the modern situation, the survey organisation is no longer in the position to set the standards; it has to adapt to the requirements set by the environment. The consequence for the survey organisation is that the focal point shifts from the survey process to the position and circumstances in which respondents respond, regardless of whether the survey population is the business or the household sector. Nowadays the general policy focuses on reducing response

burden, both actual and perceived, by minimising costs, time, effort and the number of people involved. At the same time the respondents' behaviour should be influenced in such a way that an accurate response is received, i.e. that all relevant data are collected accurately and on time. This means that the conditions under which the respondent has to provide information should be the focal point in the survey strategy.

The respondent-oriented survey design has been developed by Brög in his New Kontiv Design (see e.g. Moritz and Brög, 1999). In his latest adjustment of the Total Design Method, Dillman also focuses on the respondent in this way, still using 'TDM', but referring to it as Tailored Design Method (Dillman, 2000). These approaches fit in the now generally accepted ideas about Total Quality Management (TQM). The basic principle of TQM is that the quality of the product or service is not only based on internal process control, but on control of all aspects that influence the outcome of a production process [3]. The modern situation is presented in figure 1.

Figure 1. The modern survey organisation



In order to get the best results, i.e. minimum respondent burden and maximum customer satisfaction, it is necessary to include different aspects of the survey organisation in the communication strategies. In the specific situation of communication strategies in business surveys, this means that the surveying organisation should:

– Establish a relationship with providers on the basis of mutual respect and trust.
– Motivate them to participate by explaining why their response is important and by rewarding their co-operation (using incentives). Also show that you, as a survey organisation, feel that their participation is important.
– Make partnerships with intermediate organisations like umbrella organisations, branch organisations and so on, whose interest in accurate statistical information in their field makes them receptive for motivating their members to provide information.
– Identify subgroups of providers. Use relevant information on conditions and circumstances of the respondent in order to tailor the survey design: customise questionnaires, ask questions that can be answered, pick the right moment of survey, offer the appropriate mode of collection.
– Ask only relevant information, i.e. information that is actually requested by customers, and ask this information only once for the specific time base of the survey [4].
– At the same time invest in the use of registers instead of primary surveys.
– Give relevant feedback on the performance of the organisation (many users of statistical information in establishments are also providers of statistical information).

As for the output of the statistical process, quality of statistical information (survey estimates, figures, and so on) was traditionally defined in methodological terms, like reliability, validity, measurement error, sampling bias, and so on, in accordance with the 'one size fits all' approach. Less attention was given to aspects like timeliness of information, consistency of data sources, comparability with other types of information, accessibility, and so on. The call for a reduction of the response burden coincides with doubts about the necessity of certain types of information and the request for new sorts of information (primarily based on integration of existing information, using the high-end functionality of ICT). This change of perspective resulted in a more balanced set of quality indicators. Reviewing the general discussions at Statistics Netherlands in past years, three quality indicators can be distinguished which determine the overall quality of statistical information:

– Accuracy of the information: minimal mean square error in the statistical information. Or in non-technical terms: is the information plausible, valid and reliable?
– Timeliness of the statistical information: information has to be available as quickly as possible about the most recent events.
– Relevance of the statistical information: information has to be of some value for customers or users.

With respect to the accuracy of the information, the existing standard statistical procedures, methods and models, such as random sampling and multivariate estimation procedures, are applied to ensure minimal mean square error. Nevertheless, some assumptions underpinning the general practices are questioned here. The basic question in surveys is whether stimuli, varying according to specific circumstances of respondents, can generate the desired same response? In other words: is it necessary to treat every respondent with the same procedure(s) and instruments, in order to minimise response error? Or is response error a consequence of neglecting the specific circumstances in which respondents have to answer questions? During the 1970's research was initiated to increase the validity of survey data, resulting in the CASM movement (Cognitive Aspects of Survey Methodology; Jabine, et al., 1984; Hippler, et al., 1987). The present insight is that adapting questionnaires (wording of questions, instructions) and interview strategies (clear description of interviewer and/or respondent roles) to specific circumstances in which respondents have to answer questions will lead to better survey results (Tanur, 1992; Morton-Williams, 1993; Sudman, et al., 1996; Groves and Couper, 1998; Dillman, 2000; Sirken, et al., 1999).

Plausibility of statistical information refers to its external validity. The consistency of time-series is particularly important in this respect. Macro-economic figures may vary substantially over time. Whenever there is a major change in these figures, there has to be a plausible explanation. If the researcher fails to find such an explanation, this might be an indication of errors or failures in data collection, entry or analysis. Non-response, in particular non-response that varies over time and over subgroups (resulting in selective response), complicates the assessment of the plausibility of statistical information.

Timeliness of information is very important. Practically every customer satisfaction survey on the quality of statistical information mentions timeliness as one of the most important dimensions. Statistics Netherlands uses the so-called one-to-one rule, which means that statistical information has to be published within the time period following the time period for which the survey took place. For example, figures from a monthly survey in month T have to be published in month T+1, figures from a quarterly survey in quarter T

have to be published in quarter T+1, and so on. Questionnaires received after closing dates of surveys are processed at a later stage (publication of definite figures), or not at all. Respondents who return questionnaires too late cannot be considered as refusals or as non-respondents because they do not refuse and they do respond. Nevertheless, their response behaviour does have a negative effect on the quality of statistical information.

It is merely stating the obvious that statistical information has to be of some relevance for customers or users. In many countries, the national statistical institutes (NSI) have some form of programme, which prevents collection and dissemination of irrelevant information. For purposes of fine tuning, ongoing customer satisfaction surveys, focus groups with stakeholders and shareholders, and last but not least analysis of general trends and developments generate the necessary specifications for concepts, definitions and variables (Kavaliunas & Luppes, 1998).

High response rates have a positive effect on these quality indicators. Accuracy of information is improved by smaller variances and, given valid instruments and procedures, minimal bias. Timeliness of information is improved because more response is received within the same units of time. The basis of high response rates in establishment surveys is an efficient and active communication with providers of information [5)], which takes into account the specific circumstances of the provider and his or her organisation.

3. Case studies of respondent-oriented communication

3.1 Introduction

As we pointed out in section 2, respondent-oriented communication targets three dimensions of response. First of all, the response rate: as many respondents as possible should respond. This is influenced by how respondents are requested to participate. Furthermore, – and this is the second aspect: timeliness – if it is to be of any value for users, information must be made available on time, which means the communication strategy should also optimise, or to be more precise, reduce response time (the time between sending out the questionnaire and receiving it back). The third dimension is the accuracy of information, in which non-sampling error plays an important role (validity and selectivity of response). All aspects of the means and ways of communication with the respondent are reflected in the communication strategy and there is just one moment in time where the respondent decides whether or not to participate. Given the fact that non-response varies over subgroups of the population, mixed modes of contact strategies and questionnaires are ways to minimise non-response in subgroups. However, they will increase the costs of data collection. As Groves (1989) pointed out, it helps if we know which arguments respondents take into consideration in deciding whether or not to participate. These arguments can help the survey organisation not only to redefine data-collection procedures for subsequent surveys, but also to adjust the non-response error: '…realistic models for statistical adjustment and survey administration require theories of survey participation.' (ibid. p. 237).

Several projects carried out at Statistics Netherlands have addressed different theories or ideas with respect to human behaviour related to survey participation. Business surveys differ from household surveys with respect to sampling error because of substantial coverage problems, very skewed distributions within the target populations and volatile units, as well as non-sampling error, for example as a result of accessibility to relevant archives (Cox and Chinnappa, 1995). However, we believe that the non-response problem in business surveys can be described and analysed with the very same theories on survey participation as used in household surveys. In business surveys, as in household surveys (Snijkers, et al., 1999), it is people that have to be persuaded to participate in the survey, and it is people that read the advance letter and have to complete the questionnaire. If these parts of the communication are not optimised (e.g. refusal conversion by addressing the right person, personalisation of the letter, and overcoming cognitive difficulties with questionnaires), a refusal to participate or to complete the questionnaire is more likely to occur (Luppes, 1998).

In this section we shall give an outline of the first approaches in this field, starting with a description of a project on the Dutch Mobility Survey, which is a household survey. The results of this project triggered the rethinking of the communication procedures in business surveys, although some initiatives had been taken in business surveys prior to this project. In this context it should be mentioned that in 1992 the Questionnaire Lab was founded for cognitive testing of questionnaires (Snijkers, 1997a).

3.2 Project 'Improving response rate in the Mobility Survey' (1997)

At Statistics Netherlands a major impulse for respondent-oriented communication came with the redesign of the Dutch Mobility Survey on the basis of the New Kontiv Design (NKD), developed by *Socialdata* from Munich, Germany (Brög, 1997). The Mobility Survey is held among 60,000 households. The redesign was necessary because response rates had dropped from just over 50% in 1985 to about 35% in 1998 (Moritz and Brög, 1999). The basic philosophy in the NKD is that the respondent should be regarded as a customer, to whom interviewers have to adapt all their communication instead of the other way round. In 1997 a controlled field experiment based on the NKD design was conducted in order to establish whether a significant response improvement could be achieved. This indeed proved to be the case, the response rate in the NKD sample (n=1,000) was 74%, while in the control sample it was 44% (n=1,032).

The NKD is set up as a PAPI survey (self-completion diaries), with a telephone motivation of respondents and (possible) subsequent follow-up surveys for more detailed data in subgroups. An important advantage of this PAPI approach is that this mode of collection imposes low burden upon on the respondent. The respondents are called shortly after they have received the survey material and are motivated to complete the questionnaire and diaries. This motivation call is an important feature of the strategy. The telephone is not used to carry out the survey, but merely as an instrument to motivate the respondents.

The questionnaire itself is kept as user friendly as possible, which means as simple as possible. Basically, respondents may answer the questions in their own words, and only clearly defined and understandable categories for mode and purpose of trips are given. Pre-coded answers, explanations or definitions in the questionnaire may lead to confusion, so the design aims to put the burden of investigation on the survey organisation itself, rather than on the respondent. If the data from questionnaires are incomplete or require some clarification, additional data and information are collected by telephone. The same principles, partially structured questions and graphically well designed (i.e. comprehensible and readable) questionnaires apply to the diaries used. The basic idea behind the diary design is to obtain information on all out-of-home activities, not only those predefined by the researcher. This leads to quite an open structure.

3.3 Project 'Improving advance letters of 41 business surveys on annual production' (1996)

Advance letters are quite a common element of surveys and it is standard policy at Statistics Netherlands to send an advance letter

to alert the provider or respondent of the forthcoming call, questionnaire or interview. The quality of advance letters has often been discussed, but the lack of an appropriate theory and models lead to subjective decisions, based merely on what is considered to be appropriate. Based on the work of Cialdini (1990) and Groves, Cialdini and Couper (1992) on the compliance principles [6] underpinning requests to participate in a survey, an analytical tool was developed for content analysis of advance letters (Luppes, 1995). This tool makes it possible to describe the information content of an advance letter in relation to the psychological principles, used to convince the respondent.

This tool was applied in an internal study (unpublished) on the quality of 41 different advance letters used in annual establishment production surveys. The content analysis of these letters, performed by three independent coders, gave the following results and recommendations:

1. The enormous differences in length of the advance letters is not explained by a necessity to provide additional information that could help respondents to make a decision. On the contrary, a lot of the information, addressing definitions and detailed explanations of the survey in the longer letters probably causes confusion instead of clarity. Extra information should be given in enclosures, not in the letter. Especially the subtle differences between anonymity and confidentiality may be confusing (Luppes, 1994).
2. Most letters give a telephone number where respondents can obtain more information about the survey, but only 22 of the 41 letters give the name of a contact person. Personalised contacts are more effective in the provision of extra information about the survey.
3. In 31 of the 41 letters no information is given about the surveying organisation (Statistics Netherlands). Eleven of the 41 letters do not address anonymity and confidentiality-related responsibilities or guarantees on the part of the survey organisation. Information on the significance of the survey is given in only nine of the 41 letters, and only 17 contain general information about the influence survey results might have on government policy. Practically all letters address the costs of the survey in terms of information requested, the deadlines to be met and other direct costs (free return envelopes and so on). None of the letters fully informed the respondents on basic issues such as the aims and objectives of the survey, the survey organisation, the costs and benefits and the issue of anonymity and confidentiality (including informed consent issues).

3.4 Project 'Provider-oriented communication in the survey on finances of enterprises' (1998)

The traditional field strategy in the survey of finances of non-financial enterprises (SFE) was subjected to a study of whether communication with the providers could be improved. The SFE is an annual survey among a stratified sample of enterprises with a balance sheet total of more than 25 million Dutch guilders [7]. Although response rates in the SFE are quite high (it is a compulsory survey), response time periods are quite long, many reminder calls have to be made, as well as a substantial number of calls for clarification (accuracy problems). It was hypothesised that the traditional, passive strategy of communication caused the high number of reminder calls and the long response time periods. A qualitative study was set up to investigate the following hypothesis (Oppeneer and Luppes, 1998):

Positive attention towards the provider, as expressed in (1) the advance letter based on the principle of reciprocity, (2) the moment of sending out the questionnaire depending on availability of information (annual reports of the enterprises), combined with (3) an active reminder strategy, will lead to shorter response time periods and a drop in the number of reminder calls.

A combined research strategy was used in which information of several sources was analysed:

1. quantitative description of response rates and response time per stratum;
2. content analysis of advance letters based on an adaptation of the model described in Luppes (1995);
3. focus groups with internal staff and field staff on issues of participation (especially their perception of reasons why providers would participate or not);
4. telephone interview with providers based on the results of the focus groups (their information was used to construct a topic list with respect to reasons of participation and non- participation).

Although no definite correlation between response times, number of reminder calls and the communication strategy was established, the qualitative data in the study indicate that changing the passive communication strategy into an active one will lead to reduction of response times and the number of reminder calls. In short, the active strategy should at least consist of:

– Well formulated advance letters, which present relevant information on the survey (purpose, costs, benefits and time to complete the questionnaire) and give the provider the feeling that he or she is important for the survey. Also, thoughtful use of compliance principles is helpful in motivating the provider.
– The moment of sending out the questionnaire should be related to the availability of information. In many cases providers do not yet have information from the annual accounts when they receive the questionnaire. This means that the survey organisation has to keep track of the dates the information becomes available at the enterprises, and should send out the questionnaire accordingly.
– Given the fact that many enterprises are in a panel, it is helpful to use a provider's profile in which relevant dates and information are registered and which helps to customise the complete communication strategy. In fact, data-based communication approaches should be used to optimise the communication (Luppes, 1998).

3.5 Project 'An efficient reminder strategy in business surveys on commercial services' (1998)

In a qualitative study (Cörvers, 1998) on improving the reminder strategy in commercial services surveys, concepts developed in the New Kontiv Design (see project 1) and the Total Design Method were used to formulate some standards and rules for improving response rates. The reason for this study were the relatively low response rates in these surveys, which could only be raised to an acceptable level by an extensive number of costly and time-consuming reminder calls, letters and duplicate questionnaires. The study resulted in the following recommendations for improving the efficiency of reminder strategy:

1. Start with questionnaires which are simple to complete, which look attractive and which use concepts and questions which respondents recognise and use;
2. Reduce the number of written reminders and use the telephone more. Although this may result in higher costs, it is more effective than and reminder letters;
3. Personalise the contacts between the respondent and the survey staff as much as possible. Reducing the psychological distance (i.e. establishing a relation of trust) is also possible by collecting data in a joint venture with boards of trade. The benefits of such joint ventures make the survey more important for the providers.
4. Use an active reminder strategy in which a frequent and regular contact with providers is established. Also the quality of the contact should be high, which implies continuous training of field staff.
5. Make it clear in the case of obligatory surveys that completing and returning the questionnaire on time is a legal requirement. Just stating this requirement in the communication leads to

higher response rates (Paxson, Dillman and Tarnai, 1995), although the effectiveness of this authority principle is questioned in Oppeneer and Luppes (1998).

3.6 Project 'Redesign of the annual establishment production survey' (1999–2000)

In the second half of 1998 Statistics Netherlands started a major redesign of the annual establishment production surveys. In this redesign of over 130 surveys the input, throughput and output processes are integrated into one survey, using standardised and harmonised questionnaires, consistent data collection strategies and macro-editing procedures. The redesign was triggered by a growing dissatisfaction with the relatively low response rates, the accuracy and the coherence of statistical information. Early in 1999 an extensive inventory was made of customer demands, using focus groups of stakeholders and regular customers, as well as desk research. Based on these findings, the questionnaire was redesigned early in 2000 and cognitively tested in five focus groups with providers [8)].

Although the groups were small (about four people per group), the results were in accordance with earlier results; Snijkers, 1997b:

– The advance letter did not provide an answer to all the questions the providers had about the survey. For example, it did not state why the survey was conducted and what the data are used for. Providers said they would appreciate it if some results of the survey could be sent back to them (feedback).
– Providers pointed out that in former survey, the questionnaire was sent out at the wrong time: in March, whereas the requested data become available in June. Also in June they have more time to complete the questionnaire.
– The former questionnaire asked for a lot of detailed information, which took too much time to provide. As such items were often left open, many follow-up phone calls were necessary and this probably contributed to high item non-response rates. The corresponding items in the new questionnaire were easier to answer.
– The ordering of some items in the new questionnaire was not in accordance with the administration of some businesses. For these businesses, a lot of effort was needed to obtain the requested data. Others had no problems with the ordering. In this way subgroups of providers can be identified.
– Many providers complained that they received a large number of questionnaires every year, and, as they pointed out, they are not employed to fill in forms. What is worse, they had to provide the same kind of information for several surveys. They were not impressed by the mere fact that the surveys are mandatory.
– While at the beginning of the discussion, respondents were a bit sceptical about the aims of the meeting (improving the questionnaire), afterwards they felt it had been very useful both for them and for Statistics Netherlands. They felt they had got to know Statistics Netherlands a little better.

4. Measures to improve respondent co-operation

Given the results of the case studies described in section 3, and taking into account the features of the Total Design Method and the New Kontiv Design, we define a number of measures to improve survey participation in business surveys. We present this list of measures (that may not be complete) within the design phase of the survey and data collection phase.

Design phase:

– *Contact person:*
  Make an effort to find the right contact person within each business, i.e. the person who has access to the requested information and is authorised to provide that information. Use all information available on past response behaviour and availability of information within the establishment to customise and personalise the communication. Databases with provider profiles form the basis for effective communication. It goes without saying that properly defining the target population and picking the sample is the first step.

– *Advance letter:*
  - Always use a personalised advance letter, well formulated and written in the right tone; make the respondent feel that he or she is important.
  - Make sure the letter is clear and short; attractive to look at and well designed, without typing errors; the tone should stimulate and motivate participation and the language should be neutral and non-directive, without official jargon.
  - The letter should contain information on the organisation conducting the survey; the survey itself (what is it about, who has been requested to participate, an actual request to participate, including liability); why it is important to participate; the costs and benefits of participation; the issue of anonymity and confidentiality (including informed consent issues); who to contact in the case of questions, i.e. name and telephone number (personalisation of contact).
  - Any additional information about the survey should be given in an extra enclosure: more information on what the survey is about and what the data are used for.

– *The questionnaire:*
  - Only ask for relevant information that cannot be collected any other way, like primary data collection using Electronic Data Interchange (EDI) from business administrations, and secondary data collection using registers.
  - Only use tested, well-designed and attractive questionnaires (Dillman, 1999; Jenkins and Dillman, 1997) that are user-friendly, simple to complete, without complex routing, with questions and instructions worded in such a way that they are easy to understand.
  - Use questionnaires that are tailored to subgroups. Ask each subgroup only for the information that they are able to provide. If necessary use follow-up questionnaires.

– *Mixed mode design:*
  - Use a mixed mode design, in which respondents can choose how to provide the information, e.g. EDI, paper forms, by phone, by fax, by the Internet, or in a non-standardised way by allowing respondents to send in reports, etc. containing the requested information.
  - Only use well-tested data collection procedures and instructions.

Data collection (fieldwork):

– *Sending out the questionnaires:*
  - Make an effort to get the timing right: send out the advance letter and questionnaire at the moment the requested information is available and the respondent has time to provide that information. As many establishments participate in panel surveys information should be available to help determine the best moment for surveying each unit. As a basic rule, for every unit the availability of annual reports should be recorded in the provider profile, together with past response behaviour.
  - Motivate providers to participate by using incentives and by making personal contact at an early stage. Phone the provider immediately after the form has been sent out in order to motivate (make an agreement about when the data can be expected, thus not giving him or her the chance to say 'no'), to show the respondent that the survey and his/her participation is important, and to maintain interaction with the respondent (building a relationship). This is essential for providers who are known to return forms late, and for those whose data are

essential to the survey (e.g. large enterprises in a survey on volume of trade).
- Try to get the data as soon as possible after the questionnaire has been sent out. This is especially important for establishments that are essential within the survey. This makes publication of accurate estimates at an early stage possible.

– *Reminding and maintaining interaction (follow up):*
  - Use the phone to collect missing data (reducing item non-response), immediately after the completed questionnaire has been received. Use e-mail instead of the phone if providers prefer to communicate by e-mail. Although little is known about the real effects of e-mail communication on response behaviour, it is generally accepted that e-mail is a very efficient means of information exchange.
  - As for reminders (reducing unit non-response), switch modes: for example, use the phone, fax or the Internet instead of mail as a first reminder. This will remind late providers more effectively to send in the requested information. Using a tailored communication strategy based on a provider profile (using all information that is known about the provider in the phone call) will increase effectiveness of the reminder strategy. In general the time between two contacts should not be too long, but the moment of follow-up contact should always be arranged beforehand.
  - Sending back results from the survey to providers helps to maintain interaction and to build up a relationship of trust and respect. This information should preferably be of use to the individual provider (e.g. information on his line of business in his own region, or benchmark information). This is especially important in the case of panel surveys.

5. Conclusions

We started with a discussion on the position of providers of information and customers of this information in relation to the survey organisation in the past and at present. We have seen that in modern society the survey organisation is no longer in a position to tell providers what to do and customers what to expect as statistical products. The 'one size fits all' approach is no longer appropriate, neither for the input nor for the output of the statistical process. Dillman's Total Design Method and Brög's New Kontiv Design have improved this approach on the input side of the survey process. Today, the organisation has to adapt to input and output demands, making a 'mixed mode' approach necessary. These demands are:
– Input: reduced response burden and increased response rates.
– Output: increased timeliness and accuracy of relevant statistical information.

Of course, internal demands with regard to the throughput can also be put forward, such as efficient use of statistical modelling, applying efficient sampling, stratification and weighting methods, and using other techniques to speed up the production process, for example macro-editing and optical character recognition (OCR). As for the input and output goals, we feel that the demands can be achieved by using an active, respondent-driven and tailored communication strategy.

The respondent-oriented approach implies optimisation of the communication by using customised or tailored questionnaires and contact strategies, based on the specific conditions and circumstances of the respondent. Basically this comes down to asking the *right person* (the person who has access to the requested information and is authorised to provide that information) for the *right information* (the data that are really needed, nothing less and nothing more) at the *right moment* (when the data are available and the contact person has the time to complete the questionnaire) with the *right mode* (the mode that suits the respondent best).

However, implementation of an active, provider-oriented communication is not without consequence for the survey organisation. Apart from rising costs, and without being complete, these consequences are:
– Homogenous subgroups of providers should be identified in advance on the basis of information about which questionnaire to send and when, the preferred interviewing mode, their significance within the survey, and so on. This enables specific communication towards specific subgroups and optimises the process .
– A contact administration at the level of the provider is necessary, keeping track of all contacts with all contact persons.
– A facility for testing questionnaires and data-collection modes should be present in order to optimise questions, procedures and instructions.
– Mixed-mode designs make the statistical process much more complex since parallel processes have to be developed, thus making a detailed planning of the statistical process and co-ordination of logistics necessary. They also have consequences for the type of statistical models used.
– A call centre is needed, staffed by enthusiastic employees, well trained in handling different types of verbal communication strategies.

But above all, we would like to stress that putting a respondent-oriented approach into practice is more than just adapting tools and procedures. In order to be effective, the internal culture of the survey organisation should reflect the values of modern society. Staff not only have to be trained in the use of new tools and procedures, but also in the principles of bilateral human (interviewer-respondent) interaction, in which the special relationship between a survey organisation and the respondent is reflected. Therefore, we firmly believe that the most successful organisations will be those which determine the perceptions, needs and wants of the sample population best, and which minimise response burden through the design, delivery, and communication of appropriate and comprehensible requests.

For further information, please contact Ger Snijkers: gsks@cbs.nl or Martin Luppes: mlps@cbs.nl.

References

Brög, W. (1997), *Raising the Standard.* Keynote Paper presented at the International Conference on Transport Survey Quality and Innovation, Grainau, Germany.

Cialdini, R.B. (1990), *Deriving psychological Concepts relevant to survey Participation from the Literature on Compliance, Helping and Persuasion.* Paper presented at the 1st International workshop on Household Survey Non-response, Stockholm.

Cörvers, F. (1998), *Communicatie met berichtgevers in de commerciële dienstverlening: aanbevelingen voor efficënter rappeleren bij de sector HDV* (Communication with respondents in commercial services: recommendations for more efficient reminder strategies; in Dutch). Internal Report Statistics Netherlands, Heerlen.

Cox, B., and B.N. Chinnappa (1995), Unique Features of Business Surveys. In: Cox, B., et al (eds.), *Business Survey Methods*, New York: Wiley.

Dillman, D.A. (1978), *Mail and Telephone Surveys: The Total Design Method.* New York: Wiley.

Dillman, D.A. (2000), *Mail and Internet Surveys: The Tailored Design Method.* New York: Wiley.

Groves, R.M. (1989), *Survey Errors and Survey Costs*. New York: Wiley.

Groves, R.M., R.B. Cialdini and M.P. Couper (1992), Understanding the Decision to Participate in a Survey. In*: Public Opinion Quarterly*, Volume 56.

Groves, R.M., and M.P. Couper (1998), *Non-response in Household Interview Surveys*. New York: Wiley.

Hippler, H.J., N. Schwarz and S. Sudman (1987), *Social Information Processing and Survey Methodology*. New York: Springer.

Jabine, T.B., M.L. Straf, J.M. Tanur and R. Tourangeau (1984), *Cognitive Aspects of Survey Methodology: Building a Bridge between Disciplines*. Report of the Advanced Research Seminar on Cognitive Aspects of Survey Methodology. Washington: National Academy press.

Jenkins, C.R., and D.A. Dillman (1997), Towards a Theory of Self-Administered Questionnaire Design. In: Lyberg, L., et al. (eds.), *Survey Measurement and Process Quality*, New York: Wiley.

Kavaliunas, J., and M. Luppes (1998), *Measuring Customer Satisfaction: Progress and Best Practices of National Statistical Agencies.* Invited paper presented at the Joint IASS/IAOS Conference at Aguascalientes, Mexico.

Luppes, M. (1994), *Interpretation and Evaluation of Advance Letters.* Paper presented at the 5th Workshop on Household Survey Non-response, Ottawa, Canada.

Luppes, M. (1995), Content Analysis of Advance Letters in Budget Surveys from Seven Countries. *Journal of Official Statistics*, Volume 11, No. 4.

Luppes, M. (1998), *Communication with Respondents and Customers, Different Goals, Similar Strategies.* In: Proceedings of the GSS Methodology Conference 29 June 1998, Government Statistical Service, London, UK (invited paper).

Moritz, G., and W. Brög (1999), *Redesign of the Dutch Travel Survey: Response Improvement.* Paper presented at the Transportation Research Board Conference on Person Travel, Washington DC.

Morton-Williams, J. (1993), *Interviewer Approaches.* Aldershot, United Kingdom: Dartmouth Publishing Company.

Oppeneer, M., and M. Luppes (1998), *Improving Communication with Providers of Statistical Information in Business Surveys.* In: Proceedings of the GSS Methodology Conference 29 June 1998, Government Statistical Service, London, UK.

Paxson, M.C., D.A. Dillman and J. Tarnai (1995), Improving Response to Business Mail Surveys. In: Cox, B., et al (eds.), *Business Survey Methods*, New York: Wiley.

Sirken, M.G., D.J. Herrmann, S. Schechter, N. Schwarz, J.M. Tanur and R. Tourangeau (1999), *Cognition and Survey Research*. New York: Wiley.

Snijkers, G. 1997a, Computer-Assisted Qualitative Interviewing: A Method for Cognitive Pre-testing of Computerised Questionnaires. In: *Bulletin De Methodologie Sociologique*, June 1997, No. 55, pp. 93–107.

Snijkers, G., 1997b, *Herontwerp Enquêteformulieren en invoering euro: Focusgroepen met berichtgevers en buitendienst-medewerkers* (Redesign of questionnaires and introduction of the euro: focus groups with respondents and field staff; in Dutch). Internal report Statistics Netherlands, Heerlen.

Snijkers, G., J. Hox and E.D. de Leeuw (1999), Interviewers' Tactics for Fighting Survey Non-response. In: *Journal of Official Statistics*, Vol. 15, No. 2, pp. 185–198.

Sudman, S., N.M. Bradburn and N. Schwarz (1996), *Thinking about Answers. The Application of Cognitive processes to Survey Methodology*. San Francisco: Jossey-Bass.

Tanur, J.T. (1992), *Questions about Questions. Inquiries into the Cognitive Bases of Surveys*. New York: Russell Sage Foundation.

Notes

1) A communication strategy in a general sense is defined as performing activities and using appropriate means, necessary within a survey to get an accurate response from a respondent within a predetermined time.
2) In the Netherlands, non-response is a major problem in survey research, both in household and business surveys. Although business surveys are obligatory, the response rates of these surveys are considered to be too low. For example, the response rates of annual establishment production surveys (including number of employees, turnover, revenues and costs) vary between 50 and 85%. Low response rates not only increase confidence intervals, which could lead to rather meaningless point estimates, but also cause non-response bias in the estimates (Groves, 1989; Groves and Couper, 1998).
   Furthermore, the response time, i.e. the time between sending out and receiving completed questionnaires is quite long, which effects the timeliness of the statistical information.
3) Based on the ideas of Total Quality Management, the European Foundation for Quality Management introduced the EFQM model. This model is a non-prescriptive framework of nine areas for which to assess an organisation's performance: Leadership, People, Policy and Strategy, Partnership and References, Processes, People Results, Customer Results, Society Results, and Key Performance Results. The first five areas are 'Enablers', the other four are 'Results'. The 'Enablers' cover what an organisation does, the 'Results' what an organisation achieves. 'Results' are caused by 'Enablers'. The 'Enablers' learn and improve by evaluation of the 'Results'. In figure 1 the EFQM model (slightly adapted from the original model: turned on its side) is chosen as a model to describe the modern, open organisation.
4) In establishment surveys double questionnaires and/or overlap in questions do occur. This complaint, voiced by many of our providers of information, was one of the reasons statistical processes are being reorganised at Statistics Netherlands.
5) In the case of establishment surveys we prefer to speak of providers of information instead of respondents. The providers are the actual persons providing the information whereas the respondent is the business unit or establishment about which data are collected.
6) Groves, Cialdini and Couper (1992) describe six compliance principles which can be used in the request to participate in a survey. Of these six principles the compliance principle of authority (compliance based on power difference) and the compliance principle of reciprocity (based on mutual exchange of values) are probably most commonly found in advance letters.
7) Information on small enterprises is collected on basis of tax registers.
8) It is difficult to organise focus groups with establishments. First of all, it is difficult to find the right person (i.e. the person who completed the questionnaires before), and secondly it is not easy to make an appointment. We had about 20% response for recruitment; and of those who said they would come, only about half actually showed up.

Volume measurement of education

Paul Konijn and Foske Kleima*

Abstract

In order to be able to determine better the contribution of non-market services to overall economic growth, Statistics Netherlands has been looking at possibilities to improve the volume measurement of government services in general, and education in particular, following European initiatives in this respect. The present article describes the investigations of Statistics Netherlands into the main issues, namely how to measure the output and quality of education. It analyses some alternative methods for volume measurement using various indicators and finishes with conclusions and some ideas for future developments.

Key words: education quality, education output, volume measurement, volume indicators, public services.

1. Introduction

The economic growth rate is one of the most important indicators calculated in the national accounts. As it is equal to the change of total value added generated in the whole economy adjusted for price changes, the value of each goods or service transaction is decomposed into a price and a volume component. However, for public services like education such a decomposition is not easy to achieve, since there is no genuine market price. In the national accounts the output of public education is by convention equal to what the government spends on it, while at the same time the same government is the consumer of this output. For the calculation of the volume of education output many countries use a similar convention: the volume of output is equal to the volume of the inputs (a so-called 'input method'). Thus, the change in the volume is equal to the change in real costs. For example, an increase in the number of teachers will lead to an increase in the volume of output. However, the results of such an input method give no information on possible productivity changes in education, because inputs and outputs are equal by convention, and hence productivity does not change. There will also be a bias in the figure for overall economic growth if productivity is systematically under- or overestimated.

One important aspect of volume measurement is the valuation of changes in quality. Changes in the characteristics of a product (i.e. the 'quality') have to be included in the volume component. If a better quality product can be bought for the same price, or produced at the same costs, the value to the consumer, or the output of the producer increases (see e.g. SNA, 1993). The current methods for volume measurement of education do not take quality changes in education into account.

Statistics Netherlands is therefore investigating possibilities to improve the volume measurement of government services in general, and education in particular, following European initiatives in this respect. In 1998 Eurostat set up three Task Forces devoted to the volume measurement of non-market services, aiming at improvement of the reliability and the comparability of volume data. Statistics Netherlands participated in the Task Force on education (Task Force, 1998). At the meeting of the National Accounts Working Party of February 1999, the member states agreed to continue practical research into the implementation of the recommendations of the Task Force, in particular on the measurement of quality changes.

The present article describes the investigations of Statistics Netherlands, including the views of various education experts on the issues concerned. In the course of the study, experts from the Ministry of Education, from the Inspectorate for Education and from universities were sounded out, and there were also discussions in the Advisory Committee for Education Statistics. These discussions yielded many ideas that were explored further. The main issues discussed were how to measure output and quality of education. This report analyses some alternative methods for volume measurement, using different indicators. It should be noted that in the international literature on education several studies have also appeared (e.g. Merkies, 2000).

Section 2 discusses the definition of output of education and ways to incorporate quality changes. In section 3, possible indicators are compared for the following school types: primary education, lower general education (mavo), vocational colleges and university education. These school types were selected to illustrate the differences between possible volume indicators. Other types of education are not discussed here, since similar results were obtained (see Konijn and Kleima, 2000). Section 4 gives the cost shares of each type of education in the Dutch education system, thus including the types of education that are not discussed in section 3. Subsequently, these cost shares are used as weights in the calculation of the overall volume indices. Section 5 analyses the various derived indices on their effect on labour productivity and GDP growth. Section 6 concludes and gives some ideas for future developments.

2. Defining education output

2.1 The Task Force definition

The Eurostat Task Force defined the output of education as follows:

'*The educational output is the quantity of teaching received by the students, adjusted to allow for the qualities of the services provided, for each type of education. The quantities should be weighted together using data on the costs of (or prices for) the education provided*.'

With regard to the implementation of this definition, the Task Force goes on to say:

'*(..) that the preferred measure of educational output, for both market and non-market education services, is the number of hours spent by students in being taught*.'

The Task Force therefore explicitly opts for an output method, using the number of pupil-hours as an output indicator. It does not recommend an input method because of the impossibility of measuring productivity independently.

In practice, numbers of pupils could be used instead of numbers of pupil-hours, assuming that the number of hours that a pupil receives teaching is constant over time, or at least does not change substantially. It is rather difficult to verify this assumption, because of the problem of defining the number of hours of teaching. For example, a few years ago new teaching methods were introduced in Dutch secondary education, in which pupils receive less teaching in class and work more hours out of class on projects. In any case, the

* Paul Konijn has left Statistics Netherlands and is now working at Eurostat.

formal study load (class teaching plus project work) is reasonably constant over time in the Netherlands. There are no data on the actual study load.
The Task Force attaches great importance to a detailed stratification of data on pupil numbers. The rule should be that types of education that have significantly different costs per pupil – which serve as the weights to aggregate the change in pupil numbers of different types of education – have to be distinguished. Hence, an increase in the number of students in higher education will have a greater influence on the total volume change than a similar increase in primary education.

2.2 Quality

The Task Force definition states clearly that changes in the quality of teaching should also be taken into account. The example most often used to illustrate the importance of quality adjustments is that of a reduction in class size. If the volume of education output is defined solely as the number of pupils, a reduction in class size (for example by an increase in the number of teachers) would be registered as a productivity decline (since productivity = output/input = pupils/teachers). However, class size reductions might be an important instrument for quality improvement.
A good method of volume measurement should capture quality changes. Thus if a reduction in average class size leads to better teaching, this should be registered. If the teaching has indeed improved, this should be reflected in the pupils' achievements: they should have learned more in the same amount of time. Thus, one could say that the quality of education is determined by the efficiency in the transfer of knowledge to the pupils. One could therefore restate the definition of output of education as follows:

The output of education is equal to the total amount of knowledge transferred to pupils and students by schools in a certain period.

One could argue that this definition is limited in the sense that it does not reflect other functions of education. However, in the end transfer of knowledge is the basic function of education. Hence, it is transfer of knowledge that we would like to measure. In principle, the measure should be the difference in knowledge of pupils between the end and the beginning of the year. Such a measure would take changes in the starting level of knowledge into account. However, while there are abundant data on final examination results or similar indicators for achievements, there are hardly any data on the knowledge that pupils possess when they start school. Final examination results will then have to be used as an approximation for the difference.
The approach followed here could be labelled the 'user-value' approach to quality. The change in quality is measured by the value the user attaches to the improved product, whatever extra costs were involved to produce it. Increasing the resources for education will only yield extra quality if it can be shown that they will increase the effectiveness in transferring knowledge.
An alternative approach – which we could label the 'resource-cost' approach – would be to assume that society consumes education and thus determines, by means of the political system, how much tax money is spent on education and in which way. It is then assumed that an increase in resources will increase the quality of education. Quality is therefore measured by the extra costs that are made to produce teaching, irrespective of the actual increase in effectiveness of knowledge transfer. An input method follows this reasoning. Using this approach would make it impossible to actually evaluate the effect of government policy on the quality of education.

2.3 Indicators for quality-adjusted output

In practice it is not possible to measure transfer of knowledge directly and we shall have to look for indicators that give a reasonable approximation.

The first suggestion is the simple use of pupil numbers or, preferably, numbers of pupil-hours, under the assumption that there are no changes in the quality of teaching.
One indicator that will be affected by the quality of teaching is the number of pupils or students who graduate. It should be remembered that such an indicator is an approximation for output and not output itself: schools produce 'transfer of knowledge', not diplomas; they prepare pupils for examinations and the better they do so, the more pupils will pass. That is why the number of diplomas or graduates may be a better indicator of the amount of knowledge transferred than the number of pupils.
However, there are some drawbacks to the use of diplomas as an output indicator:
– the teaching provided to pupils who do not graduate does not count as output;
– there are large differences in the number of diplomas between schools and over time;
– there is a time lag between quality improvements and a possible increase in the number of graduates (the number of graduates in a certain year reflects the quality of teaching in a number of years preceding that year);
– the time that pupils take to reach their graduation should be taken into account as well: output will decrease if pupils need more time.

Some of these drawbacks could be mitigated by taking the number of pupils who successfully complete the present year of their course. Large parts of Dutch education (at least primary and secondary education) are organised in school years. For each year there is a minimum level of education that the pupil should have attained to be allowed to move up to the next year. Whether or not the pupil moves up is determined by his or her results during the year. There is no yearly central examination system, exams are taken only for final graduation: so schools themselves decide which pupils move up. The number of pupils who move up each year is an alternative output indicator that is affected by the quality of the teaching. Of course, there are other factors that influence the moving up ratios as well, for example a change in standards.
If such an indicator is used, the assumption is that a pupil who has moved up has obtained all the knowledge required for that year, while a pupil who does not move up is assumed to have obtained none of the required knowledge. The big advantage of such a method is that there is no time lag problem as with the use of numbers of graduates: the moving up ratios reflect the quality of the teaching in that year.
For secondary education quite detailed information is available on the number of pupils who move up each year and the number who will have to do the year again. For primary education such information is also available. However, some types of education, in particular higher education, are not organised in this way.
Another alternative that follows that same kind of logic, is to look at the 'expected' number of years that the current pupils need to obtain their graduation. The expected study duration is calculated by means of a model of the flow of students through the school system. For vocational education such calculations have been carried out (see below), but for university education such calculations are more complicated because of variations in the allowed study duration.
Data on expected study duration can be used to adjust numbers of pupils using the following formula:

$$\text{adjusted \# pupils} = \text{\# pupils} * \left(\frac{\text{nominal study duration}}{\text{expected study duration}}\right)$$

If, for example, a course normally takes four years, but the average pupil needs five years to complete it, then each year on average only 4/5 = 80% of the knowledge that should have been transferred, is actually transferred. Therefore, the number of pupils should be adjusted by that factor. This adjustment factor is in fact an approximation for the yearly moving up ratio.
If only the number of graduates is known, and no information is available on moving up ratios or study duration, the following interpretation may help. Suppose the number of pupils is distributed

equally over the respective school years, e.g. in a four-year course each year will have a quarter of the total number of pupils. Calculate the graduation ratio as the number of graduates divided by the number of pupils in the graduation year. Suppose subsequently that the moving up ratios for each year are equal to this graduation ratio. Under these assumptions the change in the number of graduates is equal to the change in the number of pupils who move up to their next year.
Another possible source of quality indicators are final examination scores. For secondary education, for example, the final examination is centrally organised: it would therefore be possible to analyse the trend in the final examination results. It is doubtful, however, whether there will be such a trend. Final examination results seem better suited for comparison among schools than for a comparison in time.

3. Results for some types of education

In the following sections, the available data for four types of education have been analysed to illustrate the results using different output indicators.
The main data source is Statistics Netherlands' education statistics and the data pertain to numbers of pupils and graduates, as well as on moving up ratios. In a few cases, indicated where applicable, data were used from the Ministry of Education. At some points it was necessary to complement the information. In the following tables, the figures printed in italics are estimates often based on extrapolation. All volume indices presented below have reference year 1995 (1995=100).
Pupil numbers for a certain year are determined on a reference date. For most types of schools, this is 1 October, for university education it is 1 December. The number of pupils in, for example, calendar year 1997 is defined as the number of pupils on the reference date of that year. These data refer in fact to the school year 1997/1998. The data on graduates and moving up ratios in 1997, however, in fact refer to school year 1996/1997. Moving up ratios include the results of the final year of education, i.e. the final examination scores. Pupils who flow into the next year of a lower level of secondary education are also counted as moving up. Ideally, the information on school years should be recalculated to calendar years. An initial calculation of the effect of this on the outcomes showed that for an individual year the volume growth of total education may change by 0.2 or 0.3 of a percentage point, although the overall picture does not change. Furthermore, such a recalculation would make it difficult to trace back the figures taken from the education statistics. Therefore, for the moment this recalculation is not carried out. In the end, when implementing the methodology in the national accounts, it is recommended that all series be formulated in calendar years.

3.1 Primary Education

For primary education information is available on:
– the total number of pupils;
– the composition of the 'pupil stock';
– the percentage of pupils who move up each year;
– the scores of the 'level tests' (*CITO* tests).

The advantages and disadvantages of these various sources of information are discussed below.

Number of pupils
The number of pupils in primary education in the nineties is shown in Table 1.

Primary education for adults is aimed at teaching basic and social skills to Dutch and immigrant adults, most importantly 'Dutch as a second language'. A rough calculation showed that the number of hours of teaching received per student is about 10% of the number of hours that children in primary education receive. Therefore, this factor is used for the number of adults in total primary education.
The resulting volume index shows an average increase in the number of pupils of 1.2% per year in this period. As stated in the introduction, using the number of pupils does not allow the observation of quality changes. It should also be noted that if the output of primary education is measured by pupil numbers it will be influenced to a large extent by demographic factors. These are very important indeed: more pupils means a need for more teaching, and vice versa.

Composition of the pupil stock
A volume index that uses only pupil numbers, does not take into account changes in the composition of the pupil stock, for example an increase in immigrant children. If the starting level of knowledge of the pupils declines, the schools will have to make a larger effort to maintain the final level. By distinguishing between different types of pupils, changes in the composition of the pupil stock can be included. This is possible since schools receive more resources for pupils who are behind in their development.
Native Dutch pupils with a lower level of education are called '1.25' pupils. Slower pupils with immigrant parents are called 1.9 pupils. Table 2 shows the composition of the pupil stock.

Table 1
Numbers of pupils in primary education

Reference date 1 Oct., x 1,000	1990	1991	1992	1993	1994	1995	1996	1997	1998
Pupils in primary education	1,392	1,408	1,415	1,427	1,451	1,477	1,502	1,520	1,534
Adults in primary education	115	121	134	145	139	125	118	110	116
Total, inc. 10% of adult education	1,404	1,420	1,428	1,442	1,465	1,490	1,514	1,531	1,546
Volume index pupils in primary education	94.2	95.3	95.9	96.8	98.3	100.0	101.6	102.8	103.8

Table 2
Numbers of 1.0, 1.25 and 1.9 pupils in primary education

Reference date 1 Oct., x 1,000	1990	1991	1992	1993	1994	1995	1996	1997	1998
1.0 pupils	–	766	782	801	846	896	949	1,042	1,069
1.25 pupils	–	480	468	452	424	396	363	285	267
1.9 pupils	–	156	162	170	176	181	184	189	194
Volume index using composition	–	95.4	96.2	97.1	98.5	100.0	101.4	102.2	103.1

The total number of pupils in Table 2 differs slightly from the total given in Table 1, since a few thousand so-called 1.4 and 1.7 pupils have been left out. The numbers of pupils per type are weighted together using cost factors which reflect the average relative amounts of extra resources granted per pupil of a certain type. A rough estimation of the weighting factors based on information from the Ministry of Education resulted in the following factors: 1.09 for a 1.25 pupil and 1.42 for a 1.9 pupil. The calculated cost factors are based on approximate data for 1999 provided by the Ministry of Education. As it is assumed that the cost factors do not significantly change in time, the cost factors for 1999 have been applied to the years 1991–1998. Although according to the Ministry of Education this assumption is not entirely correct, Statistics Netherlands does not have the data to calculate the cost factors more accurately. Again, 10% of the number of adult students in primary education is added to the weighted total number of pupils, resulting in the volume indices shown above. The index shows an average increase of 0.96 % per year and thus differs slightly from the one based on the number of pupils. However, the refinement could be of increasing importance in the future. Because of the uncertainty in the cost factors, for the time being the weighted index will not be used for the calculation of the overall index for education.

Pupils that move up each year

It appears that only 11% of pupils (with weight 1.0) in primary education need nine years instead of the nominal eight years to complete primary school. Practically no pupils need more than nine years. This results in a moving up ratio of 0.986 per year. For 1.25 and 1.9-pupils the moving up ratios per year are 0.979 and 0.967 respectively (Inspection of education, 1999). Since these factors are almost constant over the years and very close to one, a correction of the numbers of pupils with these factors will not have a significant effect.

CITO test

A large proportion of Dutch children take the CITO test, a national test, in the final year of primary education. The aim of this test is to give an indication of the level of secondary education appropriate for the pupil. The average test scores of all pupils did not change in the period 1995–2000. No CITO scores are available by type of pupil, so they are not really useful as an indicator for quality changes. Since two years ago, introduction tests are also organised in the sixth and seventh years of primary school. In the future it may be possible to compare the results of the introduction tests with those of the final test, and thus draw some conclusions about the 'amount of transferred knowledge' in the last years of primary school.

In conclusion, the extra available information concerning primary school education does not lead to an improvement of the simple volume index based on numbers of pupils.

3.2 Secondary education: lower general secondary education (mavo)

Mavo is the lower level of general secondary education, comprising four years. This section only refers to the last two years of *mavo*, plus part-time *mavo*, which is mostly adult education. There are no data on the number of hours of teaching received by the part-time pupils. It is therefore assumed that part-time pupils receive 50% of the amount of teaching per year that full-time pupils receive. The number of graduates consists of graduates from both full-time and part-time courses, both counting for full (see Table 3).

The *mavo* stream showed a decrease in the number of pupils and graduates. Also the percentage of pupils not moving up decreased. The difference between the index based on numbers of pupils and the one based on numbers of pupils moving up is not very large. The number of graduates dropped spectacularly in 1991 and 1992, after which it was reasonably stable until 1997. In 1998 there was again a considerable fall. Seen over the whole period, the decreases in the various indices are not too far apart (see Figure 1). In the various overall indices one of the three indices for *mavo* is used.

Figure 1. Comparison of the volume indices for mavo

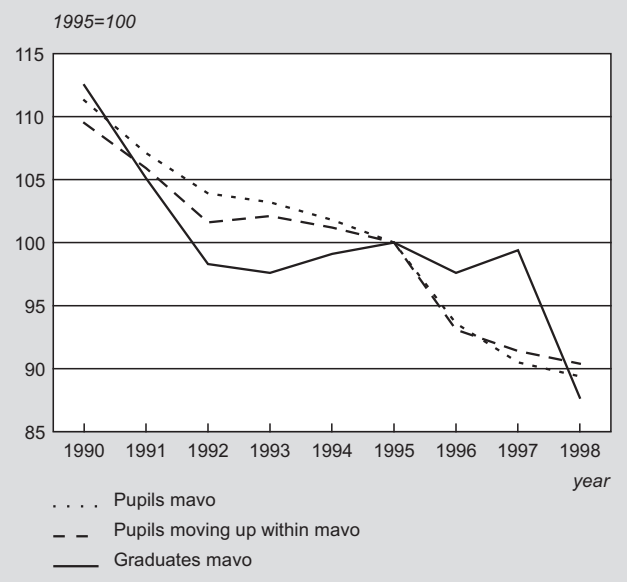


Table 3
Numbers of pupils and graduates in the last two years of mavo

Reference date 1 Oct., x 1,000	1990	1991	1992	1993	1994	1995	1996	1997	1998
Pupils full-time mavo	131	127	123	121	120	118	111	102	100
Pupils part-time mavo	53	49	48	50	48	47	43	52	53
Total, with part-time pupils counting 50%	158	152	147	146	144	142	133	128	127
Volume index pupils mavo	111.3	107.1	103.9	103.2	101.8	100.0	93.6	90.5	89.4
Percentage of pupils not moving up to next year	9	8.5	9.5	8.5	8.0	7.5	8.0	6.5	6.5
Number of pupils moving up	143	139	133	134	132	131	122	120	118
Volume index pupils mavo moving up	109.5	105.9	101.6	102.1	101.2	100.0	93.1	91.4	90.4
Number of graduates	61	57	54	53	54	54	53	54	48
Volume index graduates mavo	112.5	105.1	98.3	97.6	99.1	100.0	97.6	99.4	87.7

3.3 *Vocational colleges (hbo)*

The vocational colleges (*hbo*) are the highest form of vocational education in the Netherlands. The nominal study duration is four years. Data are available on numbers of students and graduates. There is a clear increase in the number of students at vocational colleges, but an even larger increase in the number of graduates. This indicates an improvement in performance.
The Ministry of Education calculates a series of the expected study duration of the student population in a certain year for the vocational colleges. However, this series only goes back to 1993. It shows a clear decrease of the average study duration, which again points to efficiency gains.

If the number of students is adjusted for this decrease in study duration we find a stronger increase of the output, but still not as strong as the increase in the number of graduates, as is shown in the Figure 2.

Because the information concerning study duration is not available for the whole period under study only the indices based on number of pupils and number of graduates can be used in the overall index.

3.4 *University education (wo)*

At universities, the number of students fell considerably while the number of graduates increased (except in 1997 and 1998). In principle, this could be interpreted as an improvement of performance. However, it should be noted that in this period rules for study grants changed: students had to complete their studies in a shorter period of time in order not to lose their grants. On the one hand, the total number of enrolled students decreased because the study duration per student decreased, while on the other hand, students eligible for grants under the old and the new rules graduated at the same time, causing for example a peak in the number of graduates in 1996.

Figure 2. Comparison of volume indices for hbo

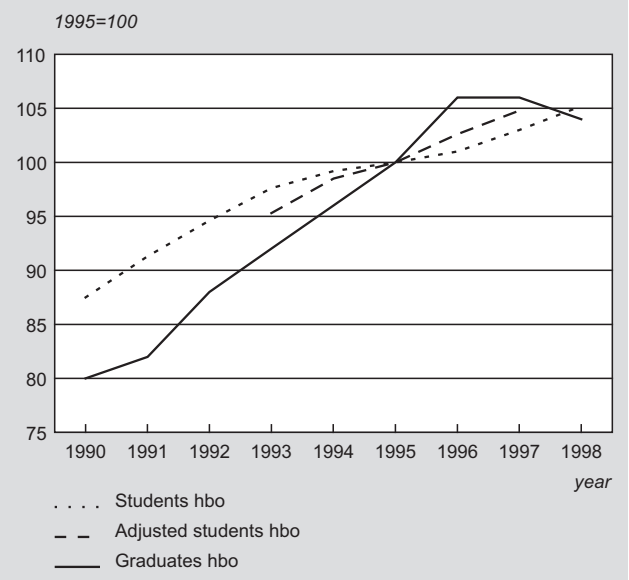


Figure 3. Comparison of the volume indices for wo

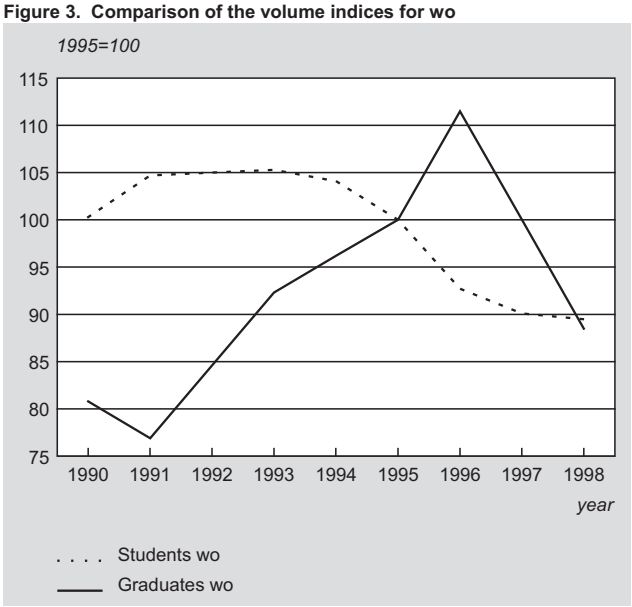


Table 4
Numbers of students and graduates for hbo

Reference date 1 Oct., x 1,000	1990	1991	1992	1993	1994	1995	1996	1997	1998
Students full-time hbo	194	205	214	222	228	231	233	237	241
Students part-time hbo	53	50	49	48	44	42	43	45	48
Total, with part-time students counting 50%	221	230	239	246	250	252	255	260	265
Volume index students hbo	87.5	91.3	94.6	97.6	99.2	100.0	101.0	103.0	105.2
Number of graduates	40	41	44	46	48	50	53	53	52
Volume index graduates hbo	80.0	82.0	88.0	92.0	96.0	100.0	106.0	106.0	104.0

Table 5
Numbers of students in hbo adjusted for expected study duration

	1993	1994	1995	1996	1997
Expected study duration hbo	4.59	4.51	4.48	4.41	4.40
4/expected study duration	0.87	0.89	0.89	0.91	0.91
Adjusted number of students	214	222	225	231	236
Volume index adjusted number of students hbo	95.3	98.5	100.0	102.6	104.8

The volume index based on the number of graduates varies considerably compared with the one based on the number of students (see Figure 3). One explanation is that the effect of demographic changes is stronger for the number of graduates than for the total number of students. In addition, the large relative variations are partly caused by the small population of graduates. A small absolute change might seem large in relative terms. This effect can be mitigated somewhat by taking a moving average.
For university education it would in principle also be possible to calculate expected study duration, as was done for the vocational colleges. However, such a calculation is hampered by the fact that , as stated above, the nominal study duration has changed and this change was not the same for each discipline. For example, nowadays students in technical physics receive a grant for five years, language students for only four. For each discipline an expected study duration should thus be calculated. At the moment only realised study duration per group of disciplines (so-called HOOP groups) can be calculated for students that have completed their courses. Although the Ministry of Education calculates expected study duration for each HOOP group, realised and expected study duration differ significantly. At the moment it is not possible to improve the volume index for university education by using expected study duration.
In the various overall indices either the number of pupils or the number of graduates is used.

4. Cost shares and overall indices

Cost shares

Table 7 lists government expenditure on the various types of education. The types of education discussed in section 3 have been shaded grey. Expenditure that cannot be attributed to a certain type of education, such as spending on general administration (e.g. the Ministry) are not included, so the total expenditure does not correspond with expenditure published in the education statistics or in the government budget. Since the outlays per type of education are used as weights in the construction of the overall volume index, it is mainly the share of each type of education in the total that is important, not the absolute numbers.
Some assumptions had to be made, in particular with respect to the costs for the various types of secondary education. In recent years many schools have merged, making it hard to allocate the costs to the various school types. For this reason, Statistics Netherlands publishes only the spending on secondary education, vocational education and adult education as a whole. In data from the Ministry, however, some more indications were found, in particular on costs per pupil. Not all the data matched completely, but a distribution was nevertheless made. For university education, the expenditure on research had to be separated, as did expenditure on student passes for public transport and study grants.

The share of each type of education in total expenses in year t-1 is the weighting factor for the volume index of that type of education in the overall volume index for year t (i.e. the volume change from year t-1 to year t). The overall volume index thus is a chain Laspeyres index.

Overall volume indices

The following volume indices for education as a whole have been analysed:
– Index of the number of pupils without weighting for each type of education;
– Index of the number of pupils with weighting with expenditure per pupil/student for each type of education;

Table 6
Government expenses on the various types of education

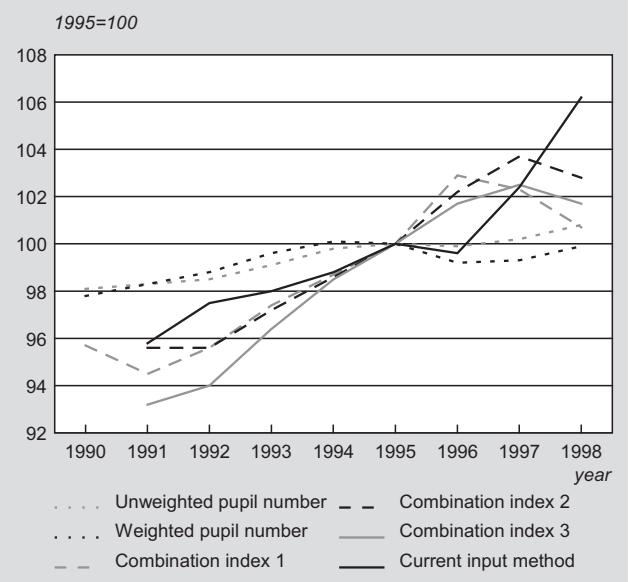
Total government expenses in billions of guilders	1990	1991	1992	1993	1994	1995	1996	1997
Primary education	6.0	6.2	6.6	6.6	7.0	7.3	7.5	7.6
Special education	1.3	1.4	1.5	1.6	1.7	1.7	1.7	1.8
First common years of secondary education	2.5	2.7	3.0	3.0	3.2	3.3	3.3	3.5
Pre-vocational education	1.3	1.3	1.3	1.4	1.4	1.4	1.5	1.5
Lower general secondary education	0.9	1.0	1.1	1.0	1.1	1.2	1.1	1.1
Higher general secondary education	0.9	0.9	1.0	1.0	1.1	1.2	1.2	1.3
Pre-university education	1.1	1.1	1.2	1.2	1.3	1.3	1.3	1.4
Senior vocational education	2.2	2.3	2.2	2.4	2.4	2.5	2.7	2.8
Vocational colleges	3.4	3.5	3.7	3.9	4.2	4.2	4.4	4.6
University education (i.e. excl. research)	2.6	2.7	2.9	2.8	2.9	3.1	3.1	3.1
Total	22.2	23.0	24.4	24.9	26.1	27.1	27.8	28.5

Table 7
Government costs per pupil for the various types of education

Total government expenses per pupil in 1000 guilders	1990	1991	1992	1993	1994	1995	1996	1997
Primary education	4.3	4.4	4.6	4.6	4.7	4.9	4.9	5.0
Special education	12.4	12.8	13.4	13.8	14.2	14.2	14.1	14.7
First common years of secondary education	6.6	7.1	7.8	7.8	8.5	9.0	9.2	9.4
Pre-vocational education	9.0	9.7	9.9	10.9	10.7	10.7	11.4	11.7
Lower general secondary education	6.0	6.4	7.2	7.1	7.8	8.3	8.4	8.7
Higher general secondary education	6.6	7.1	7.8	7.8	8.5	9.0	9.2	9.4
Pre-university education	7.3	7.7	8.5	8.5	9.2	9.7	9.9	10.2
Senior vocational education	5.9	6.0	6.0	6.4	6.3	6.7	7.3	7.7
Vocational colleges	15.5	15.3	15.5	16.0	16.7	16.7	17.3	17.7
University education (i.e. excl. research)	15.1	15.1	16.1	15.4	16.1	18.2	19.5	20.1

– Index of either the number of pupils or the number of graduates, depending on the type of education (“combination index 1”);
– Index of either the number of pupils or the moving average of graduates, depending on the type of education (‘combination index 2’);
– Index of either the number of pupils, pupils moving up or graduates, depending on the type of education (‘combination index 3’);
– Current volume index according to the input method. This is composed of a part before the recent revision of the national accounts and a part after revision. Before revision it is the volume index of the output of the branch ‘subsidised education’. This includes research at universities. After revision the volume index of the total supply of the product group ‘subsidised education’ is taken. This does not include research. Strictly speaking, the indices before and after revision are not totally comparable, but the difference will be minimal.

Figure 4. Comparison of the total volume indices for education



The exact composition of the combination indices is given in Table 8. Types of education discussed in section 3 are shaded grey. In Table 9 and Figure 4 the overall indices are shown.

We can draw the following conclusions from these data:
– The indices of pupil numbers show a clearly different trend than the combination indices. The current index on the basis of the input method is comparable with the combination indices until 1996 but increases rapidly in 1997 and 1998.
– The indices of numbers of pupils show hardly any variation. The effect of weighting with expenditure per pupil on the overall indices is very limited.
– The effect of taking the moving average of the number of graduates is also limited (compare combination indices 1 and 2). As expected, the index becomes a bit more stable. The peak in university graduations in 1996, for example, is flattened out.
– Combination indices 1 and 2, based to a large extent on numbers of graduates, first show a decrease, but then grow steadily until 1997 and 1998, when they decline again.
– Combination index 3, based to a large extent on numbers of pupils moving up, follows the trend of combination indices 1 and 2, but begins and ends a bit lower.
– The current index based on the input method shows a large deviation from the other indices in the years 1997 and 1998. This increase is caused by extra resources allocated to education, in particular intended to reduce class sizes (indeed many new teachers were recruited) and to increase the use of computers in schools. The effect of such investments will only be reflected – if at all – after a few years in the numbers of graduates or pupils moving up (but never in the numbers of pupils).

Table 8
Composition of the combination indices

	Combination index 1	Combination index 2	Combination index 3
Primary education	Pupils	Pupils	Pupils
Special education	Pupils	Pupils	Pupils
First common years of secondary education	Pupils	Pupils	Pupils moving up
Pre-vocational education	Graduates	(graduates t-1 + graduates t)/2	Pupils moving up
Lower general secondary education	Graduates	(graduates t-1 + graduates t)/2	Pupils moving up
Higher general secondary education	Graduates	(graduates t-1 + graduates t)/2	Pupils moving up
Pre-university education	Graduates	(graduates t-1 + graduates t)/2	Pupils moving up
Senior vocational education	Graduates	(graduates t-1 + graduates t)/2	Pupils
Vocational colleges	Graduates	(graduates t-1 + graduates t)/2	(graduates t-1 + graduates t)/2
University education (i.e. excl. research)	Graduates	(graduates t-1 + graduates t)/2	(graduates t-1 + graduates t)/2

Table 9
The different total volume indices for education

Volume indices, 1995=100	1990	1991	1992	1993	1994	1995	1996	1997	1998
Unweighted pupil numbers	98.1	98.3	98.5	99.1	99.8	100	99.9	100.2	100.8
Weighted pupil numbers	97.8	98.3	98.8	99.6	100.1	100	99.2	99.3	99.9
Combination index 1	95.7	94.5	95.6	97.4	98.7	100	102.9	102.3	100.7
Combination index 2		95.6	95.6	97.2	98.6	100	102.2	103.7	102.8
Combination index 3		93.2	94.0	96.4	98.5	100	101.7	102.5	101.7
Current input method		95.8	97.5	98.0	98.8	100	99.6	102.4	106.2

5. Effect on labour productivity and GDP growth

Effect on labour productivity

The national accounts contain the data shown in Table 10 on labour input in the branch 'subsidised education'. Labour input in education remained practically constant from 1990 to 1995. In later years, however, there is a significant increase. The data on labour input were revised considerably upwards (15%) during the revision. It is therefore uncertain to what extent the yearly changes are comparable before and after the revision.

For further comparison of the various indices, the effect on labour productivity is calculated by dividing the output volume indices by the above labour input index. The results are given in Table 11 and Figure 5.

Using an index based on pupil numbers will lead to a decrease of labour productivity after 1995, a direct consequence of the increase in labour input with a constant population of pupils. The other indices are grossly comparable, except for the present input method in 1997 and 1998. This shows a large increase in labour productivity simply because the volume of labour has increased. This is not a meaningful result.

Effect on GDP growth

Table 12 gives a tentative estimate of the effect that the various indices have on the GDP (market prices) growth figure. Government consumption of subsidised education contributes directly to GDP.

For 1996, growth is 0.1 of a percentage point higher according to combination indices 1, 2 and 3. For 1997 and 1998 the new indices have a lowering effect because the input method gives a strong growth for these years. For 1998 the effect is even 0.2 of a percentage point which can be called significant.

Figure 5. Effects on labour productivity changes

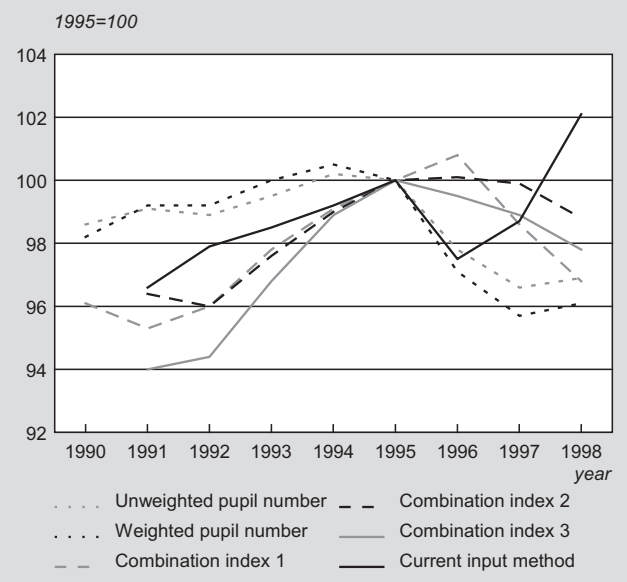


Table 12
The GDP growth figures

GDP growth figure	1996	1997	1998
Unweighted pupil numbers	103.0	103.7	103.5
Weighted pupil numbers	103.0	103.7	103.5
Combination index 1	103.1	103.6	103.5
Combination index 2	103.1	103.7	103.5
Combination index 3	103.1	103.7	103.5
Current figure	103.0	103.8	103.7

6. Conclusion

Which index gives the best representation of the trend in the output of education?

It will certainly not be the current index on the basis of an input method: the assumed relation between inputs and outputs has not been shown to exist, or at least not in the same year.

Neither will it be an index based on pupil numbers alone. In this index none of the various performance improvements we have encountered are taken into account.

Table 10
The volume of labour

x 1,000 labour years	1990	1991	1992	1993	1994	1995	1996	1997	1998
Volume of labour before revision	238	237	238	238	238	239			
Volume of labour after revision						274.9	280.8	285.1	286
Yearly changes		99.6	100.4	100.0	100.0	100.4	102.1	101.5	100.3
Volume index	99.6	99.2	99.6	99.6	99.6	100	102.1	103.7	104.0

Table 11
Labour productivity indices

Labour productivity indices, 1995=100	1990	1991	1992	1993	1994	1995	1996	1997	1998
Unweighted pupil numbers	98.6	99.1	98.9	99.5	100.2	100	97.8	96.6	96.9
Weighted pupil numbers	98.2	99.2	99.2	100.0	100.5	100	97.1	95.7	96.1
Combination index 1	96.1	95.3	96.0	97.8	99.1	100	100.8	98.6	96.8
Combination index 2		96.4	96.0	97.6	99.0	100	100.1	99.9	98.8
Combination index 3		94.0	94.4	96.8	98.9	100	99.5	98.9	97.8
Current input method		96.6	97.9	98.5	99.2	100	97.5	98.7	102.1

The indices using indicators such as graduates and pupils moving up do show these performance improvements. On the basis of the time lag argument, the index using pupils moving up (combination index 3) seems preferable. According to this index, the average yearly output volume change is 1.2 % over the years 1991-1998, the labour productivity change is 0.5 %.
Obviously, such indices remain crude descriptions of the developments in the educational system. Some assumptions are debatable (such as on the relationship between the moving up ratios and the quality of the teaching) and many aspects of education are ignored. Nevertheless, the three combination indices developed in this article most likely give better descriptions of educational developments than the present input method.
In the future, it might be possible to improve the volume indicators for primary education and university education. For primary education, results of the CITO tests among pupils in the last two years of primary school might give the opportunity to observe differences in knowledge of pupils between the beginning and the end of a school year. In addition, a more transparent system for the classification and funding of different types of pupils might allow variations in the pupil populations to be monitored. For university education only when study duration remains unchanged for several years, will it be possible to calculate expected study duration more accurately.

For further information, please contact Foske Kleima: fkla@cbs.nl.

References

Inspectie van het onderwijs, 1999. *Onderwijsverslag voor het jaar 1998* (Education report 1998).

Konijn, P. and F.J. Kleima, 2000. *Volume measurement of education*. Paper presented at the OECD national accounts meeting. Paris, 26–29 September 2000.

Merkies, A.H.Q.M., 2000. Economics of scale and consolidation in Dutch primary school industry. In: *Public provision and performance, contributions from efficiency and productivity measurement.* Blank, J.L.T. (ed.), Amsterdam, Elsevier. pp. 191–218 and references therein.

SNA, 1993. *System of National Accounts 1993* Brussels/ Luxembourg, New York, Paris, Washington D/C: CEC, IMF, OECD, UN&WB.

Task Force 'prices and volumes for education', 1998. *Final report of the Task Force 'prices and volumes for education'.*

Presenting quarterly growth rates

Gert Buiten and Ronald Janssen

Abstract

Macro-economic growth figures can be presented in a number of ways. The United States expresses economic growth in seasonally adjusted annual rates while most European countries measure growth by not seasonally adjusted year-on-year rates. This article takes stock of the three most used methods of presenting economic growth. For purposes of international comparison the European method is the most appropriate, while to determine the changes in direction and rate of growth the American method has significant advantages, although these are partly counteracted by the seasonal adjustment procedure it requires.

Key words: economic growth, growth rates, GDP, seasonal adjustment

1. Quarterly rates as indicators for annual rates

The best way to keep track of recent economic developments is to compile quarterly figures. The most important macro-economic indicator for economic development is economic growth, and is usually expressed as the volume change of the gross domestic product (GDP) in a certain year compared with the previous year. It is the link between quarterly figures on the one hand and annual figures on the other that creates the problems surrounding the use of quarterly rates: how can the GDP change in a certain quarter best be translated to an annual growth rate? The main reason underlying this dilemma is that is more difficult to compare quarters with each other than years. Firstly because the four quarters differ within one and the same a year, among other things because of differing numbers of working days and because the seasons affect potential production and spending patterns. Secondly, there may also be differences between the same quarters in different years, for example because of different numbers of working days.
In practice three methods are used to present quarterly growth rates, each with its own way of coping with the differences between quarters and each with its own way of identifying turning points and changes in the rate of growth. The methods are set out in Section 2.
Two of the methods require a separate procedure to adjust for seasonal effects; the main problems of seasonal adjustment are dealt with in Section 3.

2. Three representations

There are three main methods of calculating quarterly economic growth: measuring the change compared with the corresponding quarter in the previous year, measuring the growth since the previous quarter and extrapolating the growth since the previous quarter to calculate an annual growth rate. The first method is used by most European countries, who use the second method to give supplementary information. The third method is used by among others the US and Japan.

A. Changes on one year previously

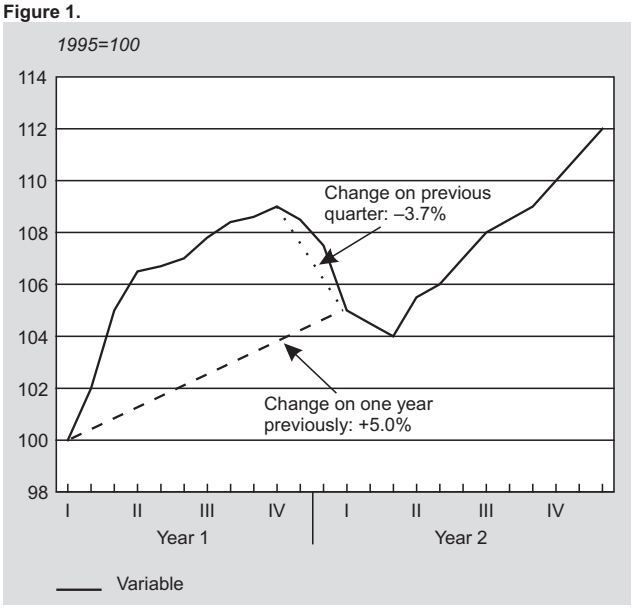
The first method of presenting quarterly growth rates is to calculate the percentage of increase of GDP in a certain quarter compared with the corresponding quarter in the previous year. This is the method used in most European countries, and is referred to as the y.o.y. (year-on-year) or n.s.a. (not seasonally adjusted) rate.

The method comprises an implicit adjustment for seasonal effects: a winter quarter is compared with a previous winter quarter, etc., but no correction is made for differences in the number of working days.

The advantage of this method as an indicator for the ultimate annual growth is that the growth rate is in the same order of magnitude as the 'real' annual growth. For example, a growth rate of 3.1% for the first quarter and 2.8% for the whole year.
Furthermore the method is also sensitive to incidental effects. In its purest form, it does not correct for things like an extra working day, an extremely cold or a mild winter, a swine fever epidemic etc. This sensitivity is amplified further by the occurrence of 'echoes'. If the economy underachieves in a given quarter because of, say, adverse weather conditions, a 'normal' quarter one year later will give an extra high growth rate. The annual rates, too, are subject to incidental and echo effects. But the effect of one quarter is less significant for the year as a whole.

The method is relatively slow to spot turning points and changes in growth rates. As Figure 1 shows, it may still indicate growth while the value of a variable is already decreasing. Policymakers depending on this information therefore run the risk of changing course too late.
Another drawback in this respect is that incidental factors may hamper the view of more structural developments. One extra working day in a certain quarter, for example, may lead to a higher growth, while the structural trend is declining.

Figure 1.



B. Changes on preceding quarter

A second method to present quarterly growth figures is the change in GDP in a certain quarter compared with the preceding quarter. European countries using the year-on-year method described above often use this method to provide supplementary information. The resulting figure is called the q.o.q. (quarter-on-quarter) or s.a. (seasonally adjusted) rate.
This method requires a separate correction for seasonal effects to take account of economic differences between quarters. This correction renders two successive quarters comparable with each other, and a growth percentage can then be calculated.

One disadvantage of seasonal adjustment in general is that the results depend on the procedure chosen and on the exact way it is applied. Section 3 gives an example. The consequence of this is that seasonally adjusted figures are less objective than the unadjusted year-on-year growth rates. The lack of a standard international correction method also means that the seasonally adjusted figures cannot be fairly compared between countries.

As an indicator for the ultimate annual growth the method has the drawback that the quarter-on-quarter growth rate is of a different order of magnitude than the 'real' annual rate. Growth calculated according to this method, for example, may be 0.6% in the first quarter while the annual growth rate is 2.8%.
Depending on the method chosen for seasonal adjustment, a correction can also be made for differences in working days and even for a number of clear and quantifiable incidental factors, such as extremely cold or extremely mild winter weather. This may help to give a clearer picture of the underlying trend. However, as the incidental factors are taken on board in the calculation of annual growth, adjusting for them in the quarterly figures will give a less accurate interpretation of the ultimate annual growth.

This method is relatively quick to identify turning points and changing growth rates. As we can see in Figure 1, it indicates a decrease in the value of a variable quite soon after the actual turning point. If calendar effects and a number of incidental factors are also taken into account in the seasonal adjustment, this method may also give a more accurate reflection of the underlying structural growth. In that case this method also gives a better insight into actual turning points or changing growth rates.

C. Quarter-on quarter growth converted to annual rate

A third method consists in extrapolating the seasonally adjusted change compared with the previous quarter to a figure for the whole year. The quarterly growth is multiplied by four, which gives a same order of magnitude as a 'real' annual growth rate. This method implicitly assumes that this growth will continue for four quarters. In fact the seasonally adjusted quarterly growth is amplified, so that the difference between two successive figures becomes more clearly visible. This method is used by among others the US and Japan. The rate is also called the annual rate or s.a.a.r. (seasonally adjusted annual rate).
As an indicator for the ultimate annual growth, the advantage is that the quarterly figure is in the same order of magnitude as the 'real' annual growth rate.
The method has the same pros and cons with respect to the correction for incidental factors as the quarter on quarter seasonally adjusted figures. Depending on the method chosen for seasonal adjustment therefore, differences in the number of working days and quantifiable incidental factors can be explicitly taken into account. However, the remaining incidental influences are amplified, increasing the risk of differences with the ultimate annual growth, in which while the incidental factors are incorporated, they will have less impact for the whole year than in one quarter.

The method is just as quick to spot turning points and rate changes as the seasonally adjusted rate. An added advantage it that is easier to spot when the growth rate accelerates or slows down, as just a small difference will be magnified. On the other hand, a drawback of this magnification is that it will tend to exaggerate or underestimate the extent of a turnaround or rate change. Another disadvantage is that the effect of remaining incidental factors is magnified, which may obscure the turning point.

Which method?
The decision on which of the three methods to use will depend on the purpose for which the figures are intended. Economic growth rates are often used to compare national economies, to give an indication of annual growth and to determine moments of change.

For purposes of international comparison, the year-on-year method is in many respects the best, mainly because it is objective and transparent. Because the other two methods require seasonal adjustment, it is not completely clear what has and what has not been corrected for. Without supplementary information, it is difficult to see whether adjustments have been made for, for example, more or fewer working days in a quarter, for abnormally warm or cold weather, etc. Moreover, the results partly depend on choices that are made in determining the methods of seasonal correction.
A slight demerit compared with the other two methods is that incidental factors are not corrected for. This may make interpretation more difficult in the case of factors which are specific to a certain country. Additional information can be given on what the effect is of quantifiable incidental factors: for example, GDP per working day.

For an indication of the ultimate annual growth, the method of quarter on quarter seasonally adjusted growth rates has the disadvantage that the results are of a different order of magnitude. The other two methods are not really suitable either, as economic growth is not necessarily spread evenly across the year. The year-on-year method has an option to obviate this by calculating the cumulative growth rate as the year progresses: growth in the first quarter, then the first six months and lastly the first three quarters are compared with the corresponding period in the preceding year. The resulting growth rates will come closer and closer to the ultimate annual rate. The quarter on quarter methods do not offer this possibility.
One drawback could be that in the annual rates the effects of a number of incidental factors are removed, whereas they do have effects in the annual growth, although this obviously depends on the exact method of seasonal adjustment. The annual rates have the added disadvantage that they start out from an assumption which is almost never the case: that the current quarterly growth will continue for four quarters in a row.

To determine turning points and changes in the rate of growth the year-on-year changes have the disadvantage of slowness and the unadjusted effect of incidental factors. The other two methods do partly adjust for incidental factors. Section 3 shows, however, that in practice the seasonally adjusted series still have substantial fluctuations which indicate a considerable influence of the remaining incidental factors.
Of the other two methods, the annual rate has the advantage that turning points become visible sooner, but the disadvantage that the remaining incidental effects are magnified. The seasonally adjusted rate would therefore seem the best option for this purpose, but it is relatively the least informative for outsiders. In information for a wider audience the latter might be a good argument to opt for an annual rate.

Another factor to be taken into consideration is that in practice, quarterly rates undergo a number of revisions. The first provisional estimate is released quite soon after the end of the quarter, based on relatively little statistical information. After a time, founded on more information, the second estimate is published. The size of the revision is relatively speaking smallest with the year-on-year method, especially compared with the annual rates method. In the Dutch quarterly accounts the revision of economic growth in the second estimate is in practice half a percent point at the most compared with the first estimate. With a growth rate of 3% this is a maximum correction of 17%. The revision of the accompanying seasonally corrected quarter-on-quarter growth is in practice regularly half a percent point in the Netherlands. At a quarterly growth rate of 0.6%, this gives a revisions of nearly 100%. For the annual rate based on this, the adjustment will be in the same order of magnitude. This means that for the successive estimates for the same period, quarter-on-quarter growth rates give the most stable picture of the economic cycle.
The methods with an explicit seasonal adjustment contain the risk that a small adjustment in the estimate may lead to a sizeable

seasonal correction, depending on how the outlying values are taken into account (see also Section 3).
A last argument that may be relevant in the choice of presentation method is the nature of the statistical production process. If an organisation uses figures based on the year-on-year method, it can also calculate the quarter-on-quarter increase. However, in some countries the central statistical authority receives only seasonally adjusted quarter-on-quarter changes as basic material, and so does not really have a choice. This is the case in the United States, for example.

3. Seasonal adjustment in practice

As Section 2 has shown, seasonal adjustment is an essential factor in the presentation of quarterly growth rates. Therefore we now turn to the methods to correct for seasonal effects and a number of other factors which may affect the picture of economic growth in a given quarter. In addition to a general explanation, the method currently used for the Dutch quarterly accounts will be discussed and compared with the method used previously in the Netherlands. This will show clearly that seasonal adjustment is not an objective process: the results depend on a number of choices a statistician makes.

Seasonal and calendar effects

To be able to compare quarterly growth rates with the each other we have to take account of differences between the quarters that may have an effect on the growth. In practice, seasonal and calendar effects are taken into account.
First of all, differences between quarters may be caused by seasonal effects. In the summer conditions for economic growth are different than in the winter. The effect can be corrected for by taking account of the influence of an average summer, or winter etc. This is actually a structural and predictable effect. In addition to the influence of an average season, abnormally cold, warm, wet or dry spells also affect economic growth. These are incidental unpredictable effects. They can be quantified to a certain extent, for example with information on temperature, but the actual economic influence can only be determined with the aid of statistical estimation methods.

In the second place, quarterly economic growth may be affected by 'calendar effects': differences in the number of working days, holidays, a leap day, etc. The payment of holiday and end-of-year bonuses in a certain quarter also leads to economic effects between quarters, in part these are incidental, but they are predictable and quantifiable. Here, too, the actual economic influence can only be determined with the aid of statistical estimation methods.
Then there are other – mainly incidental – factors that may have an effect in comparing the economic growth between two quarters. Widespread industrial action, livestock disease, changes in legislation or tax regulations, trade boycotts etc. all have their effects on economic growth. These are not usually corrected for in the seasonal adjustment of quarter-on quarter or annual rates methods.

Methods of correction

The basic principle of seasonal correction is estimating certain effects with the aid of statistical techniques on the basis of times series analysis. For this reason seasonal corrections can only be calculated if sufficiently long time series are available, usually at least 28 quarters.
To estimate seasonal or calendar effects the progressive average is often used: the average value of a variable in a given number of past quarters. The exact length of the period taken into account differs per effect and is ultimately determined by the statistician. The period is also called the 'filter'. Generally speaking, the longer the filter, the more stable the adjusted series and the smaller the effect of incidental factors will be.
In the Netherlands, for example, the winters of 1987 and 1998 were extremely cold. To establish the effect of an average winter, it makes a difference whether these winters are in the filter or not.

The occurrence of incidental factors may make it difficult to calculate progressive averages. Some are quantifiable however: calendar effects and deviations from average temperature. Therefore some methods of seasonal correction have a sort of pre-correction for the quantifiable incidental effects.
Other – non-quantifiable – incidental factors may lead to the economy performing particularly well or particularly badly in a certain quarter, what we call an 'outlier': a value that deviates strongly from the preceding and subsequent values, and thus distorts the progressive average. For this reason these values are sometimes left out of account in the calculation of the moving average.
The seasonal adjustment method is not only used to correct for seasonal and calendar effects but also for the effects of incidental factors. To this end a long period is chosen for the progressive average, so that a trend cycle can be calculated.

Seasonal and calendar corrections are not calculated manually but with the aid of specially developed computer programs. The most often used are REGARIMA X-11 and REGARIMA X-12 of the United States Bureau of the Census and TRAMO/SEATS of Eurostat, the statistical office of the European Union.
X-11 is a program that calculates only seasonal corrections and offers little opportunities for pre-correction. X-12 and TRAMO/SEATS can incorporate pre-corrections. Moreover these programs provide the possibilities of setting more options separately.

Two different applications of REGARIMA X-12

To illustrate how the seasonal corrections are made, we shall describe two applications of REGARIMA X-12 as used in the Netherlands. The first (method 1) is the application used at the moment in the Dutch quarterly national accounts. The second (method 2) is the predecessor of the present method. In both cases the option for preliminary corrections are used. The example shows that the results are sensitive for choices made by the statistician. In the calculation of the annual rates the differences are amplified. In the example this leads to differences of as much as over one percent point in the growth rates.

Table 1 gives an overview of the calculations on the basis of figures for 1998, 1999 and the first quarter of 2000. In the first quarter use is made of the first estimate of the quarterly accounts, not the second on which the other tables in this publication are based. This is because in the first estimate the growth was lower than in the fourth quarter of 1999, so that the last quarter was considered as an outlier. In the second estimate, growth was adjusted upwards so that it was no longer an outlier.

Preliminary corrections

The first preliminary correction applied is for the winter effect, and is applied in both methods. A severe winter has a number of important consequences for production, especially in the mineral extraction (natural gas) and the construction industries. In a severe winter more natural gas is produced and more value added is generated in this sector. There is a high correlation between value added and the number of degree-days (an indicator for the deviation from the average daily temperature). In the construction industry, on the other hand, production is low in a severe winter.
This winter correction is complemented with a delayed winter option in method 2, as it turns out that following a severe winter in the first

quarter, the construction industry tends to catch up on at least part of its backlog in the second quarter.

A second preliminary correction is applied for calendar effects. These too are applied in both methods but the way in which this is done differs. Calendar effects are caused by differences in the numbers and types of days between quarters. The number of calendar days can differ between quarters, sometimes even between corresponding quarters (leap days) and according to a regular pattern between successive quarters. In addition, there is often a difference in the number of working days.

The consequences of one working day more or fewer are quite complicated. One working day more often implies one non-working day fewer (in the comparison of corresponding quarters). But in some sectors work does go on even on non-working days – police and public transport, for example, and in others production is even highest on non-working days: amusement parks, for example. The total effect of one working day more or fewer is therefore the result of a long sum of pluses and minuses.

One important problem is that in most countries there is no information on the actual number of working days in a quarter. Holiday leave is not taken evenly throughout the year, and the holiday pattern may change from year to year. In the Netherlands, for example, there is the compulsory staggering of school holidays. Moreover in weeks with one or two national holidays, many people take three or four days leave to have the whole week off. It makes an essential difference, for example, whether Christmas and the New Year are weekdays or fall in the weekend.
Method 1 standardises to a certain number of working days per quarter. If there are more working days in the quarter concerned, the preliminary corrections are applied and the value added is diminished accordingly (and vice versa). The total effect averaged over a relatively long period is then around 0.2 of a percent point per working day.

In method 2 the average effect of an extra calendar day is calculated on the level of the value added in a quarter. Then the working day pattern is examined, as not all working days are equally productive.

A third preliminary correction concerns the treatment of the outliers. Both methods incorporate the removal of strongly deviating values, but method 2 is stricter in defining and detecting outliers than method 1. This is partly caused by the slightly more detailed preliminary corrections for the effects of winter, Easter and the calendar than in method 2.

Seasonal correction

After these preliminary corrections, the actual seasonal correction is applied. The calculation of this correction is the same for both methods. The results differ because of the differences in the preliminary corrections.

The seasonally adjusted GDP

Table 1 gives the results of the two methods for the last nine quarters mentioned above. Probably the first thing we notice in the two seasonally adjusted series are the relatively large fluctuations (see Figure 2). Relatively high growth rates in the first quarter followed by much lower rates in the subsequent quarter etc. This indicates that incidental factors may also have a considerable effect in the seasonally adjusted series.

Table 1
Two methods to calculate seasonally corrected growth rates

	Unit	1998				1999				2000
		I	II	III	IV	I	II	III	IV	I [1)]
Method 1 (current method)										
Gross domestic product (GDP), 1995 prices	*bln gld*	180.0	189.0	178.7	190.7	185.4	195.0	185.0	199.4	193.1
Growth on one year previously	*%*					3.0	3.1	3.6	4.6	4.2
Corrections for:										
Winter effects	*bln gld*	0.6	0.0	0.0	0.5	0.5	0.0	0.0	0.0	0.5
Calendar effects	*bln gld*	0.7	0.0	–0.3	0.0	0.7	0.0	–0.3	–0.4	0.0
'Outliers'	*bln gld*	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Season	*bln gld*	1.4	–5.1	6.6	–4.3	1.7	–5.0	7.2	–4.4	2.0
Seasonally corrected GDP	*bln gld*	182.7	183.9	184.9	186.9	188.3	190.0	191.9	194.6	195.6
Growth on previous quarter	*%*		0.7	0.5	1.1	0.8	0.9	1.0	1.4	0.6
Ditto, extrapolated to annual rate	*%*		2.7	2.2	4.3	3.2	3.6	4.0	5.7	2.2
Trend cycle	*bln gld*	183.1	184.2	185.4	187.0	188.5	190.1	192.0	193.9	195.7
Growth on previous quarter, annual basis	*%*		2.5	2.7	3.3	3.4	3.3	4.1	4.1	3.7
Method 2 (previous method)										
Gross domestic product (GDP), 1995 prices	*bln gld*	180.0	189.0	178.7	190.7	185.4	195.0	185.0	199.4	193.1
Growth on one year previously	*%*					3.0	3.1	3.6	4.6	4.2
Corrections for:										
Winter effects	*bln gld*	0.6	0.1	0.0	0.4	0.5	0.1	0.0	0.0	0.5
Calendar effects	*bln gld*	0.2	0.1	–0.3	–0.6	0.5	0.1	–0.6	0.5	–0.1
'Outliers'	*bln gld*	0.0	0.0	0.0	0.0	0.0	0.0	0.0	–3.1	0.0
Season	*bln gld*	2.0	–5.4	6.3	–3.4	2.5	–4.8	7.4	–3.0	2.8
Seasonally corrected GDP	*bln gld*	182.7	183.8	184.7	187.2	188.8	190.3	191.8	193.9	196.2
Growth on previous quarter	*%*		0.6	0.5	1.3	0.9	0.8	0.8	1.1	1.2
Ditto, extrapolated to annual rate	*%*		2.5	1.9	5.5	3.5	3.3	3.2	4.3	4.9
Trend cycle	*bln gld*	182.8	184.1	185.4	187.0	188.5	189.8	191.3	193.2	194.8
Growth on previous quarter, annual basis	*%*		2.7	2.9	3.5	3.2	2.8	3.3	3.9	3.6

[1)] This example is based on the figures of the first estimate for the first quarter. Method 2 views the fourth quarter of 1999 as an outlier. Because of the upwards adjustment of the growth rate for the first quarter of 2000, in the new figures the fourth quarter of 1999 is no longer seen as an outlier.

The next to catch our eye are the differences between the two series. For most quarters, the seasonally adjusted GDP according to method 1 differs slightly from that according to method 2. The quarter-on- quarter growth rates therefore also deviate from each other. In most cases the differences are 0.1 to 0.2 of a percent point. In the conversion to annual rates this difference is magnified to 0.4 to 0.8 of a percent point.
The differences between the two methods are much greater, however, in the fourth quarter of 1999 and the first quarter of 2000. This is because method 2 sees the increase in the fourth quarter of 1999 as an outlier and applies an outlier correction of no less than 3.1 billion guilders downwards. Method 1 does not do this and therefore has a larger seasonal adjustment, but in spite of this gives a much larger adjusted GDP. This shows a change of 1.4% on 1999-III, while method 1 gives a change of only 1.1%. On an annual basis the difference is even larger: nearly 1.5 percent points.
The difference is larger still for the first quarter of 2000. Method 1 gives a change on the previous quarter of 0.6%, method 2 results in 1.2%. Naturally the difference is magnified substantially when the annual rate is calculated, 2.2% versus 4.9%.

Figure 2. Growth on previous quarter extrapolated to annual rate

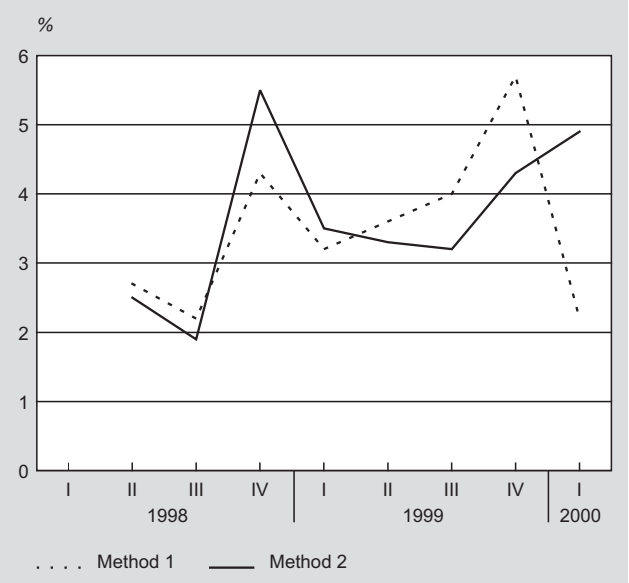
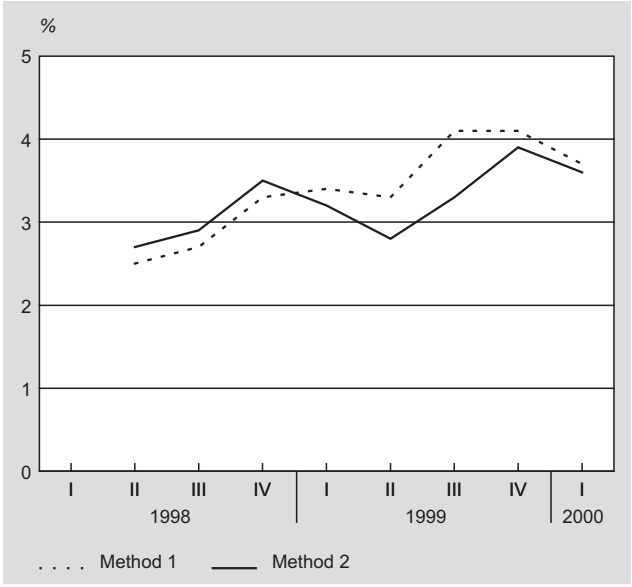


Figure 2 illustrates the relatively sizeable differences in seasonally adjusted annual rates between the two methods. One and the same adjustment program can lead to differences of more than one percent point in annual rates with two sets of starting values.
The difference in the growth rates is not all that important to indicate turning points or changes in the rate of growth. Instead it is important that both methods show the same pattern of increasing or declining growth rates. In this respect the choice of starting values of X-12 is very important. Method 1 gives a slightly increasing growth rate for the first three quarters of 1999, while according to method 2 the rate slows down. In the first quarter of 2000 method 1 shows a strong decrease in the growth rate, while method 2 shows a further acceleration.

The trend cycle

With REGARIMA X-12 it is also possible to calculate a trend cycle in which the effects of incidental factors is nearly completely removed. The differences between the two methods in the seasonally adjusted annual rate are then considerably smaller, as can be seen in Figure 3. The course of the trend cycle is also clearly more stable. Turning points or changes in rates of economic growth are therefore less easy to detect. If the effect of incidental factors is only small the annual trend cycle may therefore be a reasonable indicator for the ultimate annual growth.

Figure 3. Trend cycle quarterly growth on annual basis



4. Conclusions

This article describes the three most frequently used methods for presenting quarterly economic growth: the year-on-year or not seasonally adjusted method, the quarter-on-quarter or seasonally adjusted method and the annual rate or seasonally adjusted annual rate. The three methods turn out to be very different by nature. The choice of which method to use will depend on what the figure is to be used for.

For international comparison, the year-on-year method seems to be best. It is immediately clear what has been calculated and what is included and excluded. The other methods are not so clear on which adjustments have been made for calendar effects and incidental factors. Section 3 shows clearly that differences in the starting values may lead to very considerable differences in the annual rates even when using one and the same seasonal adjustment program.

None of the three methods are really suitable to indicate for the ultimate annual growth. The best is the cumulative calculation of the growth on the same period in the previous year: in succession the first quarter, the first six months, and the first three quarters. Section 3 shows that the annual trend cycle is also a good alternative.
The seasonally adjusted quarter-on-quarter figures and the annual rates based on these growth rates are preferable to indicate turning points and changes in the rate of growth. However, Section 3 shows that both methods are sensitive to accurate settings of the seasonal adjustment program and do not give complete certainty on turning points.

An extra consideration in choosing a presentation method is the size of adjustments of successive estimates in a certain quarter under review. The adjustments are relatively smallest in the year-on-year method and therefore this presentation gives a more stable picture of the economic development than the other two methods.

For this reason, Statistics Netherlands will continue to use its present method of presentation. The quarterly economic growth is expressed as the change compared with the same quarter in the previous year. Any effects of incidental factors and working day effects are explained explicitly in the press releases and in the quarterly publications on the Dutch economy. To get an indication of the ultimate growth in a calendar year cumulative changes are calculated for several quarters compared with the same quarters in the previous year. To identify turning points or slower or faster growth rates seasonally adjusted quarter-on-quarter changes are calculated and presented as supplementary information.

To prevent misunderstanding, the differences in the various methods of presentation will from now on be included in an explanatory note to the press releases.

For further information, please contact Gert Buiten: gbtn@cbs.nl or Ronald Janssen: rjnn@cbs.nl.

Baseline: the basis for process transformation on the input side [1)]

Jean Ritzen

Abstract

The issues of reducing response burden and improving efficiency are both high on the agenda of Statistics Netherlands. More and better use of business data available in registrations will contribute to both. In order to make the information held in registrations accessible in accordance with the desired statistical concepts and harmonised within the office, a first version of the Central Input database Baseline was completed in June 2000. This article describes the concept of Baseline and discusses some items regarding technology and methods of development.

Key words: Baseline, EDI, response burden, statistical use of registrations, statistical co-ordination, matching

1. Introduction

On of the spearheads in Statistics Netherlands' policy is making use of data from registrations to compile statistics. Traditionally, statistics based on data not originally collected for statistical purposes, such as registrations of other institutions, are called secondary statistics. Similarly, when data from registrations are made available electronically, we speak of secondary EDI (Electronic Data Interchange). The application of secondary EDI makes it possible to compile statistics more efficiently: not only does it cost less, it also poses less of a burden for respondents, who would otherwise have to complete a questionnaire.

Most registrations are wide-ranging, that is, they include a variety of elements. For instance, the VAT registration from the fiscal authorities has data on all enterprises that pay VAT, irrespective of their main activity or size. When Statistics Netherlands sees one characteristic of an enterprise as typical for the organisation of the statistical processes within an agency, it means that the registration can be important for various departments at Statistics Netherlands. This is the first kind of multiple use: more organisational units make use of it. The second kind of multiple use is when certain data in a registration (or other data set) are used for more than one statistical process, for example both for employment statistics and for production statistics. Such kinds of multiple use of a registration or of registration data make it efficient to centralise the storage and dissemination within Statistics Netherlands.

On the same principle it may also be useful to make data received directly from enterprises (primary observation) and suitable for multiple use centrally available. It may be useful to set this as a goal if the infrastructure can incorporate it.
Statistics Netherlands has recently developed a central database for storing input micro-data, called *Baseline*. The purpose of Baseline is to co-ordinate the dissemination of available enterprise data for the production of statistics and to co-ordinate the availability of these data. The respondents as well as the statistical process, and as a result the users of statistics, will benefit from Baseline.
The Baseline concept has been developed to accommodate the centralisation of enterprise data as described above. The first version of the database was completed in June 2000.

2. Definition and aims

Baseline can be defined as the database infrastructure in which enterprise data gathered from different sources can be stored in a combined and related form for the purpose of being used in the statistical process.

This definition and the principles underlying the concept of Baseline require some explanation. First there is the concept *enterprise*. Two statistical units are relevant in the compilation of most economic statistics: the 'group of enterprises' and the 'enterprise'.

The group of enterprises is the unit used for the statistical description of the financial process and the distribution of income, whereas the enterprise is the basic co-ordinated statistical unit used in the description of the production process of businesses. One of the starting points for the development of Baseline is that it should in no way undo what had been achieved up to then in the area of conceptualisation and statistical co-ordination, including the use of statistical units [2)] as registered in the General Business Register (GBR).

In the GBR the enterprise is assigned various characteristics, including an activity and a size class code. At the start of the statistical process this is the unit used to classify populations in the GBR and to assign them to production statistics categories. In direct observations, the enterprise is also the reporting unit, unless the respondents desire otherwise. So in the observation stage the data are almost always gathered at the enterprise level. Therefore, but also and especially for co-ordination purposes in the case of multiple use, the Baseline concept takes the enterprise as the reference unit, i.e. the unit to which the input data refer.

If the original input data need to be converted to obtain the enterprise level, then this conversion need take place only once and in one place if they are fit for multiple use. The various users can use the same converted data. The central Baseline concept prevents a situation in which different users of the same data convert these differently, and thus eliminates an important cause of conflicting statistical figures.

In practice and by definition, data from registrations refer to different units, or rather different types of units, than data collected especially for statistical purposes. A case in point is the data in the GBR, which are supplied by the Dutch Chambers of Commerce. The information refers to legal units, and the differences this causes are shown by the large number of legal units that combine into enterprises.

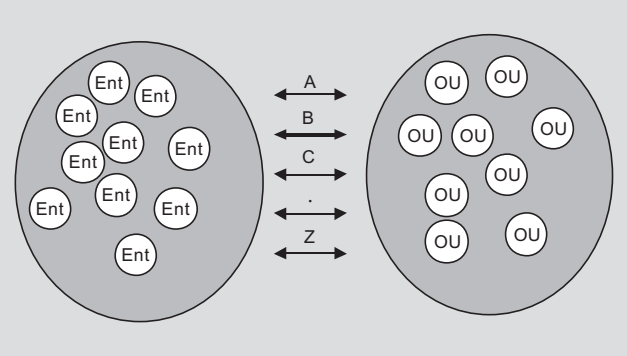
Likewise, each registration has its own specific units to which the registered data refer. Industrial insurance boards, for instance, register units with a payroll, the tax authorities supply data on units liable to certain taxes (wage tax, VAT, corporation tax, income tax). Fortunately, the tax authorities themselves have a central register in which the legal unit and the taxable unit are linked in a way similar to the way Statistics Netherlands links legal units and enterprises.

Because the units in registrations have different definitions, we call these units observation units, even if there is often a one to one relationship in all registrations. This means we need some form of conversion tool to turn these observation units into enterprises. For this purpose Statistics Netherlands has developed the Central Observation Units Register (CWR) that enables the conversion from registration units to enterprises. This register is so close to the GBR that can be considered as being part of the GBR system. The key problem is that there are no common numeric identifiers (ID numbers) of basic units in use in the Netherlands. This means that linking and matching have to be based on other characteristics. The development of a Basic Single Business Register (BBR) in the Netherlands is intended to lead to major improvements in this area. This BBR will be the source of basic information (name, address,

activity code, etc.) on basic units for several large and important business registers kept by different agencies in the Netherlands, including the Chambers of Commerce, the tax department, social security administrations and Statistics Netherlands. Where this BBR will be set up and maintained has still to be decided, but it will be outside Statistics Netherlands because of its administrative function. However, what is clear is that the creation of this BBR will greatly improve matching results of the data from different administrations.

A good central observation units register (CWR) is the key to accessing register data. The conversion at the unit level is shown in the figure below, where the observation units (OU) from the fiscal administration (taxable unit) are shown in relation to the population of enterprises (ENT) according to the CWR.

Figure 1.



Information on the quality of matching is important for the statistical use of the observations. For instance, is the observation unit exactly the same as the enterprise or are consolidations or exceptions involved? Is the relationship fully known? The access classification (A, B, …, Z) per enterprise and per registration can be used for this. It indicates the quality of the matching of observation units from the administrations with the statistical units. The possible relationships between observation units and enterprises are apparent from the meaning of the most important codes.

A A full one to one relationship between the enterprise and the reporting observation unit
B No full enterprise coverage by the reporting observation unit; the observation unit only contributes to this enterprise.
C A full coverage of the enterprise by n reporting observation units which only contribute to this enterprise.
D No full enterprise coverage by n reporting observation units; these observation units only contribute to this enterprise.
V The enterprise is part of too complex a structure for linkage with observation units from administrative files.
W The enterprise is linked to a reporting observation unit which is also linked to another enterprise in the statistical reference period
X The enterprise is matched with only non-reporting observation units (e.g. exemptions).
Y The enterprise is matched with partly non-reporting observation units (e.g. exemptions).
Z No observation units found to be linked with the enterprise.

If the units in the registrations differ from enterprises, they require a conversion at the data level, through consolidation or deconsolidation. This is done in the process of filling Baseline, but without changing the data as such in a conceptual sense. A specific problem here is the phenomenon of internal deliveries between parts of one enterprise, for example between legal units that are combined into one enterprise. Initially consolidation is carried out by adding information if there is no other option. Users of Baseline data are informed about this. This method is preferred to the introduction of missing values in this case.

The important principle is that Baseline must incorporate input data in their original state, as they are available from the registrations. No statistical processing, such as imputations, are done. Furthermore, Baseline does not select variables from more than one registration. Similar variables, employment data for instance, may be included in data files of more than one registration. Such variables are stored several times over, because choosing between options (if there is a choice) is part of statistical analysis. However, there are quality checks such as technical and comprehensiveness checks in the process of filling Baseline.

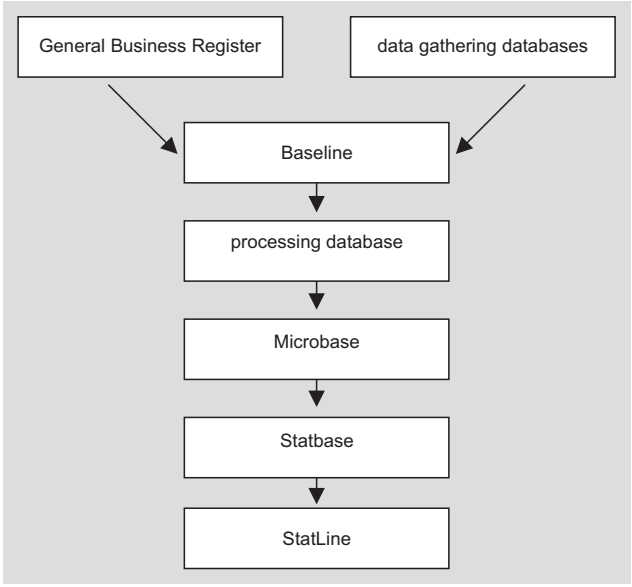
3. Baseline in the statistical process

As explained above, Baseline encompasses various types of data sets. To begin with there is the GBR and, related to it, the CWR. Then there are the individual datasets with registration and survey data, which need to be entered individually. At present, we feel that these data need to stay available for a given period, so that information can be reproduced if the need arises. Just how long this period should be has still to be decided.

Data in Baseline and the processes leading to registration need to be well-documented, both at the level of variables and at the unit level, either in terms of definition or derivation rule. This will take place in a meta-system that is being developed simultaneously and in parallel.

The development of the input microdatabase Baseline ties in with the restructuring programme for economic statistics currently being developed at Statistics Netherlands. Clearly, Baseline and the systems directly related to it (including the meta-system) are situated at the beginning of the overall statistical process. To a large extent, the quality here – at the basis – determines the quality of the resulting statistics. Baseline's position in relation to the overall system can be represented as follows:

Figure 2.



The data gathering databases are copies of the original files as made available by the sources or as collected in statistical surveys. In Baseline these data are brought together and combined in accordance with the statistical units as defined in the GBR. In the next stage of the statistical process input data are edited into statistical variables belonging to statistical units. The editing process is done in intermediate processing databases.

The statistical microbase is the resulting output microdatabase in which statistical data are assigned to the individual statistical units (enterprises). This is mostly done by editing the data available from Baseline, a stage performed in the processing databases. Statbase, still to be developed, contains all statistical information available for publication and dissemination. This information is stored in an aggregated way, but is as detailed as possible. Lastly, StatLine, which is operational, is the public electronic data warehouse that can be consulted by users. It can be accessed online via Statistics Netherlands website (www.cbs.nl) and is also available on CD-ROM.

A very carefully designed system of meta-information has been set up to make this base-driven approach successful. The figure may give the impression that the data are frequently transferred and transformed as they are moved from one database to the next. This is indeed the case, and there are very valid reasons for this in an organisation with a process-oriented division of tasks, i.e. a breakdown into input, statistical processing and integration, and dissemination. However, registering things in different places involves a certain amount of risk, and we shall have to solve this problem. The danger lies in inconsistencies between the various databases in such a dynamic process in which the aspects of time and periods play a key role. The solution will have to be found in computer-aided throughput, transfer and transformation procedures, in combination with careful management of the time dimension. The incorporation of adequate meta-information systems is essential for this.

4. Data and databases stored in Baseline

The ultimate aim is to make Baseline the central input micro-database for business data at Statistics Netherlands. This needs to be done at the concept and at the operational level, so in principle no prior restrictions are set. Of course priorities have to be set in developing Baseline to ensure that the best possible use is made of registration and survey data, and that users' wishes are fulfilled as much as possible.

The priority of the restructuring programme for economic statistics at Statistics Netherlands (IMPECT: IMPlementation of EConomic statistics Transformation process) lies in making the processes more efficient. One factor in this is making fiscal data accessible and available (from the VAT and corporate tax registrations). Among other things this requires a stable and flexible infrastructure for the storage and dissemination of these data, for which the necessary functions must also be made available.

Once these data are accessible and can be disseminated through Baseline, the same model can be used to add other registrations, for which in turn priorities have to be set. It is already clear that there is much demand for access to data on the self-employed, which can be found in the income information system (IIS). The tax forms submitted by the self-employed for income tax purposes should ideally be added to Baseline; this would lead to a considerable reduction in the response burden for small enterprises, which is high on the political agenda in the Netherlands. However, fiscal registrations are not the only focus: in principle all registrations can be included in or added to Baseline, as long as they contain business information and as long as these data can be converted to enterprise types through the Central Observation Unit Registration. One example of an interesting non-fiscal registration is the social security administration (VZA).

One precondition for inclusion is that the registrations are available electronically, which is not yet always the case yet for the desired data. Achieving this will require substantial negotiations with the sources and a fair amount of co-ordination. Furthermore, the survey data generated by Statistics Netherlands itself are also candidates for inclusion in and dissemination through Baseline. For instance data from the production statistics, observations through EDISENT, data from the central observation of large businesses, business surveys, labour data, and so on. There should be no prior restrictions when such information can help create the harmonised and efficient management of input.

In conclusion, more and more sources will be made accessible through Baseline as the work progresses, and thus more enterprise data will be made available. This also means that when the foreign trade statistics are linked to the General Business Register, they will use the enterprise as their institutional framework so they too can obtain input via Baseline. This will certainly be possible for input from VAT registrations. Another conclusion is that statistics based on units other than the enterprise will not be able to obtain data directly through Baseline in the short term: for instance, statistics describing financial processes, which are based on enterprise groups. If there is a link, however tentative, with units from registrations, such as the fiscal unit from the corporate tax registration, as an approximation of groups of enterprises, then the input information for the financial statistics can be taken directly from that registration. Unfortunately, Baseline cannot fulfil the co-ordination requirements in such cases.

5. Baseline, from *prototype version 0.1* to *production system version 1.0*

Baseline *production version 1.0* was preceded by *prototype version 0.1*. This included the annual figures from the enterprises derived from a small number of sources, namely corporate tax and the VAT registers for a limited number of years. To this we added data available from surveys for the production statistics from 1995 onwards. Data are supplied on request and under strict conditions, so that confidentiality and security can be carefully maintained. The professional versions (after 1.0) will include more sources and cover longer periods. The data refer not only to years but also to months and quarters. Moreover the time dimension has now been implemented, which is important for selection purposes.

In line with the restructuring programme for economic statistics at Statistics Netherlands, we are looking at the possibility of setting up a related system of co-ordinated population estimates. In combination with the GBR Baseline will contribute to conditions for this, leading to an improvement in consistency and coherence within the framework of economic statistics.

Baseline version 1.0 and subsequent versions is set up within certain pre-conditions according to the need for use. Metaphorically speaking:
- in the form of a mail-order company (service on demand);
- in the form of a window shop (consulting individual business data);
- in the form of a vending machine (you select, you take);
- as an instrument for data mining.

In other words Baseline is trying to live up to the concept of data warehousing. The level of ambition is high, but it cannot fully be realised from the beginning.

Of course the structure and use of the database have to meet security criteria, most importantly they have to based on informed consent and privacy regulations. This aspect requires further study; so far, we understand that regulations pertaining to individual personal data cannot be transferred on a one to one basis to the Baseline environment, because the minimum level cannot then be met: the Baseline concept means that all use focuses on the individual enterprise level. Clear regulations and authorisation procedures are needed to provide a solution for this, where details can be changed according to specific circumstances. Statistics

Netherlands can adopt a more open procedure for internal access for general qualitative data, for example, than for sensitive financial data. Flexibility is essential for the optimal use of such data; for instance, an authorisation at the variable level may be necessary for some users, while for others authorisation might need to be given at the SIC level.

The first production version of Baseline (version 1.0) became operational at the end of June 2000. In this version very strict authorisation procedures are applied and the regulations on content and responsibilities required legally were approved by the Executive Board of Statistics Netherlands.

6. Some technical aspects of the process of development and contents

The first basic version of Baseline became operational at the end of June of 2000. In order to develop this database important parts of it and the related procedures were worked out in four coherent projects by closely co-operating project groups:

a. a project to develop import procedures including checks on errors and consistency and procedures for data consolidation or data deconsolidation related to the required statistical units;
b. a project to create facilities for translation from reporting units into statistical units, the “bridge” between the administrative and statistical units (see above on CWR);
c. a project to develop tools and procedures related to the necessary meta-information;
d. a project to develop the Baseline database, its contents, functionality, authorisation procedures for access, data protection procedures, etc.

The development of Baseline is a joint effort by subject matter departments, methodologists and the ICT department, in very close contact with data users, especially groups within Statistics Netherlands involved in the restructuring of economic statistics.
Several project groups collaborated on the modelling, writing specifications, building and testing. The whole programme and its progress were discussed weekly by the team of project managers and the programme manager.

Visual Basics was used to develop the software, in an SQL-server environment. These new tools were chosen as the standards for the database management systems (DMBS) at the beginning of 1999, i.e. at the start of the project. It took quite a long time to become acquainted with these tools and to train staff in their use. The development of Baseline was the first important project for which these instruments were used at Statistics Netherlands. This was quite a risk, as there were some uncertainties pertaining to the millennium bug issues, which by definition had a higher priority in 1999.

Because of the use of new software tools it was decided to take a parallel approach in the initial stage. The software programmers made proposals in a prototype of the system and the subject-matter experts wrote the basic specifications. In September 1999 the prototype was confronted with the specification, and from then on the development consisted of an iterative process of additional specifications, software building and testing.

Hardware calculations were made for the required capacity. To meet the short-term demand a machine with a hard disk capacity of 80 Gb would be sufficient to manage a undivided database of about 60 Gb. Experiences from back-up and recovery tests had shown the need for substantial research on handling very large databases. For the moment 80 Gb must be seen as the upper limit. Technically the 80 Gb are available from five hard disks of 16 Gb each. Somewhat to our surprise, the 80 Gb capacity turned out to be insufficient to meet the needs specified beforehand, and therefore some ad hoc measures were necessary during the building and filling of the database to free disk space. Redundant indices and “redundant” basic information had to be removed.

The Baseline database should be seen as a set of relatively small tables and one large table. Each record in the large table represents an observed value of the variables that become available from the registrations. So far the large table contains nearly 400 million records. These are available for all possible selections, which can be made in a very flexible way. The record structure is defined by the entity model, in which the entities and the relations between them are defined. Meta-information is stored in a relational environment: in separate tables and in the so-called classification server. All monthly, quarterly and yearly VAT information is stored from 1998 onwards; corporate tax information from1997 onwards.
The manpower involved in the development of Baseline totalled 20,000 hours. Half of this can be assigned to subject matter experts, including activities for testing.

7. Further developments

The principle underlying Baseline is that data from registrations are brought together as much as possible in a related and harmonised way for statistical use. Therefore more and more sources should be added, as well as data on other periods. Survey data can also be added, so that it can be linked to information from registrations. In the future this will lead to implementation of the policy to conduct surveys only if no information is available from registrations.

A number of infrastructure problems must be solved before Baseline can function as we want it to. Some of these concern the handling and managing of very large databases, including related technical security guarantees (backup and restore).

Nevertheless, Baseline is the starting point for a database-driven statistical process.

For further information, please contact Jean Ritzen: jrtn@cbs.nl.

Notes

[1] This article is based on a paper presented at the 14th International Roundtable on Business Survey Frames, Auckland, 2000.
[2] The definitions of statistical units are in accordance with European regulations.

Dissertation on internationalisation and competitiveness: the case of the Netherlands

On 2 November 2000 Robert Goedegebuure of Statistics Netherlands was awarded a PhD by the Erasmus University Rotterdam for his thesis *Internationalisation and Competitiveness: Seeing Through the Netherlands.*

The thesis describes the results of Statistics Netherlands' project charting the internationalisation and competitiveness of the Dutch economy. The project was based not only on a vast range of 'traditional' macro-economic and meso-economic indicators, but also on a unique set of economic variables on individual enterprises. In comparison with other studies on internationalisation, the project made use of a very large set of enterprises, and covered a longer period of time.

The study focused on three themes: (a) the relevance of locationally bounded comparative advantages as contrasted with the broader concept of competitive advantages, (b) the analysis of micro-level data on internationalisation strategies, and their impact on corporate performance, and (c) the implications of these empirical findings for national industrial policy. Some conclusions of the studies are:

(a) The relevance of comparative advantages

– It is pointed out that traditional indices of 'revealed comparative advantage', like the Balassa index, provide an incomplete picture of competitiveness. For identifying strong and weak sectors, a portfolio box is proposed that, in addition to the Balassa index, identifies competitive strengths as revealed by the country's growth in market share in particular sectors, the attractiveness of the market in terms of growth, and the degree of concentration within industries.
– More than any other country the Netherlands realises export growth by increasing market shares. Germany, in contrast, is active in fast growing industries, but saw its market shares decrease. The portfolio approach reveals that international trade performance is, to a large extent, the result of dynamic competitive processes rather than stable comparative advantages.

The main tenet of the traditional trade theory is that countries will specialise in those products in which they enjoy a comparative advantage. Division of labour and immobility of production factors imply that if trade is indeed caused by the existence of comparative advantages, then further analyses of the Balassa index outcomes would reveal that (a) countries will tend to have relatively low imports for industries in which they are strong, and (b) countries will tend to export products from industries in which they are strong, to partner countries that have weaknesses in these particular industries. Whether or not it is true that exports indeed flow from strong to weak countries can be tested by replicating the Balassa index for import flows, analogously to the Balassa index for export flows. The assumption of comparative advantages, that would lead to the expectation that the two indices are negatively correlated, is strongly contradicted by the facts. Although there are several partial explanations for this finding (re-exports and transit trade; differences in local demand), the most likely explanation is the existence of *international clusters*, involving two-way intra-industry trade flows in raw materials and intermediate products, in addition to manufactures-for-manufactures that are rolled out internationally through the distribution networks of multinational enterprises.

Formal tests confirm the significance of what we have called 'international clustering', in addition to traditional trade explanations. Trade flows are more common if (a) the exporting country is strong and the partner country weak (comparative advantage), or (b) both countries are strong (international clusters).

(b) Internationalisation strategies

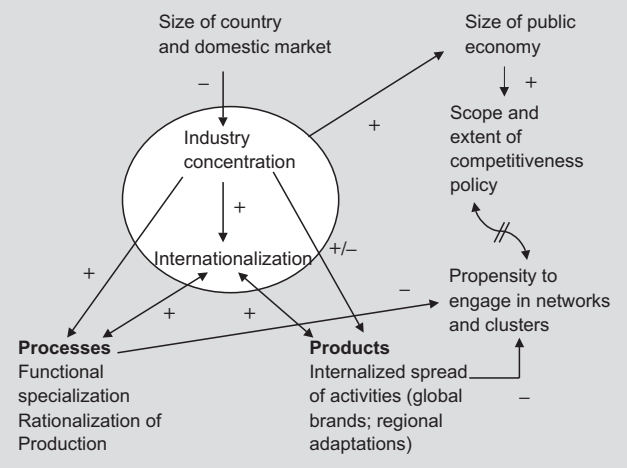
– The conclusion drawn from the analysis of macro-level data that international clustering is relevant in understanding international operations, is confirmed by micro-level analysis. For all major trading partners, a small group of home and host enterprises accounts for a large share of outward investments, imports and exports. The majority of import and export flows with these trading partners is accounted for by enterprises that both import from and export to these countries. Therefore, it is not only the relation between exports and outward FDI that is complementary in nature, but in addition imports are found to fit in this pattern of complementarity.
– Using the geographical spread of imports, exports and foreign investments, enterprises have been categorised in ten groups that represent internationalisation strategies. Of all enterprises, 19% have a national orientation, 61% a European orientation, and 20% a global orientation. Of the latter group, one in ten enterprises can be regarded as a global enterprise, characterised by a high geographical dispersion of imports and exports, and of foreign investments, both world-wide and within Europe. In addition, their share of mutual trade (imports and exports with the same partner countries) is very high. The small group of global enterprises accounts for 38% of total imports, 40% of total exports (over 50% of imports and exports refer to the same partner country), and 82% of outward foreign investments. Of all their exports, 59% are intrafirm.

(c) Policy implications

– Following the well-known work of Michael Porter, the concept of clusters (or networks) of independent enterprises has become popular in the formulation of industrial policy. This is also true for the Netherlands. Although the cluster project in the Netherlands reveals a lot of interesting data about the 'real economy', it fails to point out how the various policy measures contribute to the competitiveness of clusters. More generally, the project fails to formulate a clear set of indicators for international competitiveness. In addition to informal links within clusters, formal links between enterprises in the Netherlands, and between Dutch enterprises and their foreign affiliates as well, in our view provide a much broader basis for transferring knowledge.
– Studies on international supply chains can be seen as complementary to studies on domestic clusters. It has been found that in the Netherlands both Dutch (home) and foreign controlled (host) enterprises in the manufacturing industry tend to downgrade towards distribution centres, while especially host enterprises in the wholesale sector tend to upgrade towards manufacturing plants with secondary or even primary production. There is strong evidence that this form of upgrading positively influences employment. Concentration on core activities by producers enables service providers (transport,

logistical services) to evolve into distribution centres with secondary production.
– A comprehensive framework has been proposed that encompasses the key items of the study. It is argued that there are close links between country size, internationalisation, industrial concentration and innovation.

Figure 1. Framework for studies on competitiveness



Both exports and the penetration of host enterprises are higher in concentrated industries. Home enterprises in these sectors are relatively innovation intensive, but their participation in international networks lowers their propensity to engage in domestic clusters. In general, their strategy focuses on international division of labour and international roll-outs of international brands, as evidenced by (1) the shift of employment from the Netherlands to overseas, (2) functional specialisation (de-industrialisation of the manufacturing industry and industrialisation of service industries), and (3) the close complementary relations between foreign investment, export and import patterns.
– Applying the framework to the Netherlands, it was concluded that, as foreign investments do not generally lead to superior profit performance, the main aim of industrial policy should be to identify sectors with negatively motivated foreign investments (foreign investments motivated by the assumed lack of domestic investment opportunities), and to divert these investments to fast growing enterprises, preferably within the same cluster. The data used in the study, and qualitative information from the cluster monitor project, indicate that domestic growth in employment is mainly achieved by small and medium- sized enterprises, rather than very large enterprises. It is however the latter group that has the means and the facilities at their disposal.

For further information, please contact Robert Goedegebuure: rgde@cbs.nl.

Index Netherlands Official Statistics Volumes 11–15, 1996–2000

Policy and organisation

Business statistics

Employee training and productivity in Dutch manufacturing firms *Martin Boon and Ben van der Eijken*	1998 spring
Atlas of plant communities in the Netherlands *Lodewijk van Duuren, Joop Schaminée and Eddy Weeda*	1998 winter/ 1999 summer
Economic accounts for agriculture *Peter Pauli*	1998 winter
Business cycle surveys for the manufacturing industry *Ad Abrahamse and Thom Werkhoven*	1999 summer
Land use statistics in the Netherlands *Wim Lengkeek*	1999 autumn
Structural business statistics: a key role in the system of economic statistics, the Dutch case *Symon Algera*	1999 winter
Bilateral differences in Intrastat *Robert V. Goedegebuure and Piet J.M. Koken*	1997 spring
Increased competition in Europe: what European enterprises think of the impact of the Single Market programme *Anne Katrien Amse, Raoul Depoutot, Ben van der Eijken and Henry Nieuwenhuijsen*	1997 summer
CBS-IRIS: an EDC tool for international trade statistics *Vick H.M. Smeets*	1997 autumn
Electronic commerce and EDI: a European perspective *Philippe Lebaube and Maarten Boon*	1997 autumn
New methodology for traffic accident registration: just how safe are Dutch roads? *Jacco C. Provoost*	1997 winter
Redesign of the Dutch Travel Survey *Ger Moritz and Werner Brög*	1999 autumn
Software estimates in the Netherlands *Kees van den Ende and Piet Verbiest*	2000 spring
One figure for the supply and demand of services *Bart F.M. Bakker and Johan van Rooijen*	2000 summer

Social and spatial statistics

Marital status and cause of death, 1950-1990 *Inez M.N. Joung, Jacobus J. Glerum, Frans W.A. van Poppel and Jan P.W.F. Kardaun*	1996 spring
Deflating health expenditures *Abraham J. de Boo and Jan M Smit*	1996 spring
The causes of moving out of poverty *Eric Schulte Nordholt*	1996 summer
Effects of firm performance and technology on wages: evidence from cross-sectional matched worker-firm data *Martin Boon*	1996 autumn
Social indicators for Europe *Ko Oudhof and Pieter C.J. Everaers*	1996 winter
Health interview surveys: towards international harmonisation of methods and instruments *Agnes de Bruin*	1996 winter
Increasing the relevance of income statistics : experiences and plans in the Netherlands *Paul van der Laan and Henk K. van Tuinen*	1997 spring
Long-term population scenarios for the European Economic Area *Joop A.A. de Beer*	1997 spring
Household actual final consumption in the Netherlands *Mies Bernelot Moens*	1997 spring

Explaining gender wage differences 1999 winter
Bart .F.M. Bakker, Kea G. Tijdens and Jeroen W. Winkels

Labour accounts principles and practice: experiences in Denmark, the Netherlands and Switzerland 2000 spring
Brigitte Buhmann, Wim Leunis, Alain Vuille and Kirsten Wimar

Social statistics on the electronic highway; co-ordinated statistics in StatLine 2000 autumn
Anne Katrien Amse

Determinants of days lost from work because of illness 2000 autumn
Martin Boon

Comparing national housing situations within Europe 2000 autumn
Jeanne Roijen

Volume measurement of education 2000 winter
Paul Konijn and Foske Kleima

Macro-economic statistics

Estimating consumer price indices for small reference populations 1996 spring
Martin Boon

Calendar effects on quarterly GDP growth rates 1996 winter
Ted Reininga and Brugt Kazemier

The regional allocation of the construction industry: discussion on the methodology for the regional accounts, location of the construction company versus location of the building site 1996 winter
Henk J. Nijmeijer and Bas de Vet

Less government, more market 1997 spring
Fred G.J. Arkesteijn

The knowledge-based economy in the Netherlands 1997 summer
Norbert H.L. van den Hove

A new approach to seasonalising Gross Domestic Product 1997 summer
Anil Mancham and Peter C. Oomens

Balance sheets for the Netherlands 1997 summer
Marcel Pommée

EFLO: Electronic exchange of local government financial data
Fred G.J. Arkesteijn 1997 autumn

Water in the Dutch national accounts: a NAMWA for 1991 1998 autumn
Mark de Haan

Illegal production in the national accounts 1998 autumn
Ron van der Werf

Measurement of capital stock and consumption of fixed capital in the Netherlands 1998 autumn
Piet Verbiest and Peter van de Ven

A new method for compiling time series of national accounts data 1998 winter
Gert P. den Bakker and Robert A.M. van Rooijen

Revision of the Dutch National Accounts; first results and backgrounds 2000 spring
Gert Buiten, Jacqueline van den Hof and Peter van de Ven

A simple model linking social protection statistics and the national accounts 2000 spring
Frits Bos

Knowledge: expenditures, output and growth; provisional results of a knowledge module supplementing the Dutch national accounts 2000 autumn
André van den Berg

Presenting quarterly growth rates 2000 winter
Gert Buiten and Ronald Janssen

Technology, data collection, survey techniques