Chapter 3
NOTATION OF FINDINGS

3/1 Standard excerpts of research reports
- 3/1.1 Structure of the excerpts
- 3/1.2 Place in the database
- 3/1.3 Example of an excerpt

3/2 Why excerpts are made
- 3/2.1 Getting an overview
- 3/2.2 Better comparability

3/3 Standard information in excerpts
- 3/3.1 About the investigation
- 3/3.2 About measured happiness
- 3/3.3 About correlational findings
  - Measured correlate
  - Observed relation with happiness

3/4 Technical terms used in the excerpts
- 3/4.1 Words for study design
  - /4.1.1 Population
  - /4.1.2 Sampling
    - Probability sampling
    - Non probability sampling
    - Response
  - /4.1.3 Assessment
    - Strategy
    - Measurement
- 3/4.2 Words for measured happiness
- 3/4.3 Words for measured correlate
- 3/4.4 Words for observed relationship

3/5 How the excerpts are made
- 3/5.1 Procedure
- 3/5.2 Tools

3/6 Summary

Appendices
- A Notation form on paper
- B Electronic notation form

Intro

Filling a bookcase with acceptable studies is just the first step. The next is to take out the relevant findings and present these findings conveniently. For that purpose standard-excerps of each research-report are made. These excerpts consist of three parts, which are stored in a database system and which can be retrieved independently.

This chapter starts with a short description of this approach and an example in

section 3/1. Then section 3/2 enumerates the reasons for this time consuming technique. Next section 3/3 describes what information is taken in the excerpts are made and presents illustrative examples. Since there is no common research terminology in the social sciences, section 3/4 defines the technical terms used in the excerpts. Finally, the practice of excerpting is shortly described in section 3/5. The next chapter (chapter 4) describes the statistics used in the excerpts and provides further notation examples

3/1 STANDARD EXERPTS OF RESEARCH REPORTS

3/1.1 Structure of the excerpts

3/1.2 Place in the database

3/1.3 Example of an excerpt

This catalog presents correlational research findings in standard abstracts. These abstracts are part of a focused excerpt of the research report. These excerpts are stored and sorted in this finding-browser.

3/1.1 Structure of the excerpts

The standard excerpts involve three parts:

Part 1 **Study**
- Bibliographics of the research report
- Design of the empirical investigation

This part reports never more than one study

Part 2 **Measured happiness**
- The query used for assessing happiness
- The distribution of happiness observed

When the study involved more than one indicator of happiness, this part reports these separately.

Part 3 **Correlational findings** (standard-abstracts)
- Summary of study and happiness query
- Measurement of correlated variable
- Observed statistical relationship with happiness

When a study related happiness to more than one variable, this part presents more than one mini-abstract.

3/1.2 Place in the database

The structure of the excerpts corresponds with three sections in the database system.

The first section lists the investigations that have considered correlates of happiness. It can be browsed on characteristics of the report (such as the author's name) and on characteristics of the investigation (such as the kind of people). This section is linked to the 'Bibliography of happiness'.

The second section is about measurement of happiness. This section enumerates the techniques (queries) used for assessing happiness and the observed distributions. This section is linked to the ' Catalog of Happiness Queries'.

The third section of the database comprises the abstracts of correlational research finding. This section is the heart of this 'Catalog of Correlational Findings'. It can be

browsed on subject.

These excerpts are not exactly 'summaries'. They do not aim to cover all the issues the author has raised in a report, but focus exclusively on his empirical findings on happiness. *All* the findings on happiness are excerpted, not only the findings stressed by the author or the ones that seem most relevant. The excerpts are not restricted to significant correlations; non-correlations are noted as well.

An example of an excerpt is presented in exhibit 3/1.3 .

3/1.3 Example of an excerpt

World Database of Happiness
Correlational Findings

Excerpt of study HAY 1979

Excerpt of Study: HAY 1979 Among: Adult, general public, non-institutionalized, Great Britain, 1976 N= 2000

Reported in:

Author: Hay, D.
Title: The spiritual experiences of the British

Bibliography: New Society, 12/4/1979, p.72-74

Excerpted by:

First input: Piet Ouweneel 26-4-91
Last update: Leo Herwig 20-10-99

Design of the investigation:

Population: Adult, general public, non-institutionalized, Great Britain, 1976

Time: 1976

Sampling: Stratified sample

Non response: ?

N: 2000

Observation: Partly highly structured, partly focused open-ended interview

Remarks: Data also reported in: Hay & Monsy (1978) "Ecstatic, Paranormal or Religious Experience in Britain and USA: a comparison of trends". In: Journal for the Scientific Study of Religion, 1978.

Excerpt of Study: HAY 1979 Among: Adult, general public, non-institutionalized, Great Britain, 1976 N=2000

Measured Happiness

Page:

Query:

Focus:	Affect: Balance (Bradburn's	*Code:*	A-BB
Time-frame:	last month, last few weeks	*Code:*	cm
Observation:	multiple questions	*Code:*	mq
Rating type:	verbal scale	*Code:*	v
Rating range:	-5-5	*Code:*	2
Variant:		*Code:*	a

Full Text: Selfreport on 10 questions:

"During the past few weeks, did you ever feel ....?" (yes/no)
A Particularly exited or interested in something?
B So restless that you couldn't sit long in a chair?
C Proud because someone complimented you on something you had done?
D Very lonely or remote from other people?
E Pleased about having accomplished something?
F Bored?
G On top of the world?
H Depressed or very unhappy?
I That things were going your way?
J Upset because someone criticized you?

Answer options and scoring:
yes = 1
no = 0
Possible range: -5 to +5

Name: Bradburn's 'Affect Balance Scale' (standard version)

Author's label: Psychological well-being

Observed responses:

Frequencies:	*%1*	*%2*	*%3*	*%4*	*%5*	*%6*	*%7*	*%8*	*%9*	*%10*	*%11*

	On original -5-5 scale	*Transformed on 0-10*	*%DKNA*
Mean:			
Standard Deviation:			
Error estimates:			
Remarks:			

Excerpt of Study: HAY 1979 Among: Adult, general public, non-institutionalized, Great Britain, 1976 N= 2000

Finding

Page: 72

Measured Correlate

Classification: R 1.5 Current religious experiences

Author Concept: Spiritual experiences (1)

Measurement:
```
0 No spiritual experience
1 Reports to have had religious or paranormal experience
```

Measured Values: 1: about 30%

Error-estimates:

Remarks:

Observed association with happiness:

Happiness Query	*Statistics*		*Elaboration/Remarks*
A-BB/cm/mq/v/2/a	DM=+	p<. s	Ss reporting religious experiences were significantly more likely to be balanced and personally happy than other people.

End of Report

3/2 WHY EXCERPTS ARE MADE

3/2.1 Getting an overview

3/2.2 Improving comparability

3/2.1 Getting an overview

One reason for making standard excerpts is that the number of findings is getting too big to overview. When I took interest in this matter in the late 1960s, the number of empirical studies on happiness was still small and review articles could summarize the findings fairly well. That is not possible anymore. Even an experienced scientist who is willing to spend a year reading will not manage to overview all the available research, in particular not unusual findings. The more the field develops, the more difficult it becomes to oversee all its variations.

This problem is greatly reduced when the findings are entered in a relational database, from which they can be retrieved in several ways. This enables an easy overview of findings on a specific subject (f.e. all the findings on the relation between happiness and intelligence) or in a particular public (f.e. all the studies about happiness among intellectually gifted people). Entering in a database requires that the findings be described in a uniform format. That is what the excerpts do.

3/2.2 Improving comparability

As already noted in chapter 1, the available findings are difficult to compare, and therefore lend themselves not easily for synthetic analysis. Excerpting the available studies in a uniform way can help to overcome that problem.

Different organization of reports

Research-reports differ in the way they present data about happiness. Some report relevant information in footnotes and appendices, while others present information in separate (and not easily accessible) statistical supplements. A lot of research-reports are simply chaotic. Uniform excerpts help to get an overview.

Different labeling of variables

Not all investigators use the same word to depict what is called 'happiness' here. Current labels for the same phenomenon are 'morale', 'general satisfaction' and 'personal adjustment'. The same problem appears in labeling variables to which happiness is related; essentially similar factors are adorned with quite different names. Answers to questions about 'self esteem' for instance are labeled with terms as 'mental health', 'role-adjustment' and 'identity'. On the other hand, the same term is sometimes used to cover different concepts. The term 'health' for example refers sometimes to 'absence of apparent disease', sometimes to 'frequency of complaints' and in another instance to 'longevity'.

This problem is solved by forgetting the theoretical label investigators attach to their observations and by focusing on what they have actually measured. All variables are rubricated one well-defined conceptual classification. Delineation of the happiness

variable has been reported in chapter 2. Re-definition of correlated variables takes place in the subject-classification in chapter 5.

Different technical vocabulary

A similar problem is that the studies do not use the same technical terms. The term 'reliability' for example is used sometimes to refer to similarity in response to the same question asked twice and sometimes to the association between answers on different questions believed to represent the same variable. Likewise, terms such as 'scale', 'non-response' and 'sample' carry different meanings. This confusion of meanings is nicely illustrated by the 'Thesaurus of Social Research Terminology '(VanderMerwe 1974), a voluminous book, the purpose of which is to list current technical jargon. Obviously, this situation can easily lead to misunderstanding. Therefore a standard technical vocabulary was defined. Terms used in the research-reports were translated into that terminology. This vocabulary is presented in section 3/4.

Incomparable statistics

Several investigators report their results in frequency tables. These tables are included in the mini-abstracts of correlational research findings and can therefore be easily compared. Comparison is facilitated by presenting the tables in the same format, that is, in percentages and with happiness vertically and the correlated variable horizontally.

On the basis of these frequency distributions one can compute summary statistics, such as 'eta' or 'tau-c'. Such statistics are also reported in the mini-abstract. A routine for their computation is in development. Chapter 4 explains what the best comparable statistics are at different levels of measurement.

Quite often research findings are only reported in summary statistics. These statistics are based on slightly different assumptions about the mathematical qualities of the data and linearity of the relationship. For instance, the most commonly statistic 'r' assumes that both variables are measured at the metric level. Results expressed in different statistics are not quite comparable. Therefore the abstracts mention as many statistics as possible, the greater the number of statistics, the better the chance of getting sufficient findings expressed in the same statistic. When possible, additional statistics are computed by the excerpter. This matter is discussed in more detail in chapter 4.

3/3 STANDARD INFORMATION IN EXCERPTS

3/3.1 Investigation

3/3.2 Measured happiness

3/3.3 Finding abstract(s)
- Measured correlate
- Observed relation with happiness

The excerpts do not summarize the research reports, but extract only information about findings on happiness. Since it is *standard* excerpts, the same information is sought for all investigations. These pieces of information are listed below. They are all entries in the excerpt. See the example on exhibit 3/1. Below, the term on the left are the keywords in the excerpts. The right column describes the meaning and notation.

3/3.1 Investigation

The first page of the excerpts mention the source of information and characterize the investigation.

Study	Empirical investigation on happiness
Study code	Short name, first five characters of the first author's name and year of publication. 1, 2,... added if the publication reports more than one investigation. A, B, ...added if the name code year occurs in a year.
Report	Report of an empirical investigation
Author	Writer of that report
Title	Name of that report
Bibliographics	Where and when the report was published
Excerpist	The person who extracted the information on happiness from the report and entered it in this database
In	date of first entry
Up	date of latest correction
Design	What people were investigated and how
Population	The people investigated
Time	When the investigation took place, in what year
Place	Where the investigation took place, in which nation
Public	What kind of people were investigated
Sampling	How subjects were selected
Non-response	% non-participation of selected subjects
N	Number of subjects investigated

3/3.2 Measured happiness

The second page of the standard excerpts summarizes how happiness was measured in the study concerned. The following term are used:

Happiness query	How happiness was assessed, precise questions used to elicit self-reports or other ratings.
Focus	Manifestation of happiness measured
• Overall	statements about comprehensive enjoyment of life
• Affect	statements about how well one feels
• Contentment	statements about how reality meets wants
• Mixed	statements which combine the above
Timeframe	Period referred to
Mode	Kind of query used
• Self report	Introspective account of one's own happiness
• Rating by others	Estimate of happiness (affect) of another person (based on observation of expressive behavior)
Scale type	Device for rating degree of happiness
• Graphical	Degree of happiness denoted by line representation
• Numerical	Degree of happiness denoted by digits
• Verbal	Degree of happiness denoted by words
Range	Number of degrees distinguished
Sub-variant	Minor differences between measures that are Identical in focus, timeframe, query and rating-scale
Question-Code	Short name for unique kind of happiness question
Measured happiness	Applied measure and observed responses
Authors name	How happiness was labeled in the study report
Frequencies	Observed distribution on rating scale
Mean	Average
CI [..]	95% confidence interval of the sample mean
SD	Standard deviation
Original scale	Score based on scale range of question used
Transformed	Converted score on range 0-10 (either by Thurstone procedure or linear stretching)
Error estimates	Indications of measurement error in observation, such as repeat-reliability or Conbach's alpha.

3/3.3 Correlational findings (finding-abstracts)

In the example on exhibit 3/1 the last page presents an abstract of a findings. That abstract involves four pieces of information:

The investigation (summary 3/3.1)

Measurement of happiness (summary 3/3.2)

Correlated variable and its measurement.

Observed statistical association between happiness and the correlate.

These latter two elements are described as follows:

Correlate	Variable of which relation with happiness is investigated (co-variate)
Correlate classification	How correlates are grouped by subject
Correlate code	Character label for class of correlates
Standard name	Verbal label for class of correlates
Authors name	How the variable was originally labeled in the study report
Measured correlate	Applied measure and observed distribution
Measure of correlate	How the correlate was quantified
Frequencies	Observed distribution of ratings
Error estimates	Indications of measurement error in observation, such as repeat-reliability or Conbach's alpha.
Observed association	Relationship between happiness and correlate
Statistics	Numbers for summarizing statistical patterns
• distribution	measures of modal tendency and dispersion
• association	measures for degree of correspondence
• difference	measures for testing hypotheses, typically by assessing chance that an observed value in a sample exists in the population investigated.
Elaboration	Specification of observed relationships

These finding-abstracts are the substance of this Catalog of Correlational Findings. The catalog is in fact an organized collection of such abstracts. In chapter 5 we will see how that collection is organized and on what features it can be can be searched.

3/4 TECHNICAL TERMS USED IN EXCERPTS

3/4.1 Words for study design
/4.1.1 Population
/4.1.2 Sampling
/4.1.3 Assessment

3/4.2 Words for measured happiness
/4.2.1 Query
/4.2.2 Observed distribution
/4.2.3 Error estimates

3/4.3 Words for measured correlate

3/4.4 Words for observed relationship

Not all authors use the same technical terms in describing their investigations: what is called a 'pseudo-random sample' by one author, is for instance labeled as an 'ordinal sample' by another. Likewise, some refer to a 'scale' as a series of questions on one subject-matter, while others use the same term to denote an answer-device for a multiple choice question. Similarly, there are great differences in the meaning attached to common words, such as 'reliability' and 'validity'. Comparison between the various research-findings reported in this catalog requires that the design of the investigations be described in one same language. Therefore all technical terms used in the excerpts are enumerated and explained below. They are ordered in sequence of their appearance in the standard-excerpt.

3/4.1 Words for study design

Empirical studies on happiness are characterized by 1) the kind of people aimed at (population), 2) the selection of participants (sampling) and 3) the method of assessment. Below is an overview of the technical terms used for that purpose.

3/4.1.1 Population

Time, place and public define the population of a study. That is, the year of investigation, the nation and the kind of people involved.

Time

In most cases the year of investigation suffices. In case of follow-up studies the years of the waves must be noted. The first assessment is commonly denoted as T1, the second as T2, etc. So T1: 1965, T2: 1975 means that the first assessment took place in 1965 and the second ten years later.

Place

Since the present day world is neatly split up in nations, all populations studied can be characterized by the country of investigation. In case of an *international* study, more

than one nation is involved. Investigations in a particular area within a country are denoted as *regional* studies.

Public

Many studies aim at assessing happiness of the average citizen in a country, in other words on the *general public***.** Other studies rather aim at people who share some special characteristics, such as age, income, occupation, etc. We then speak of a *special public*.

The distinction is used in the ordering of findings in this catalog of correlates. In each subject-category we first present the findings yielded general public samples and next findings among special publics.

3/4.1.2 Sampling

Sampling is the process by which inference is made to the whole by examining only a part. Usually a limited amount of the subjects is selected out of a wider population. When the process of sampling is spread over a longer period we speak of *time sampling or multi-phase sampling*.

Two types of sampling methods can be distinguished, 'probability' sampling and 'non-probability' sampling. *Probability sampling* implies selection of subjects at random and allows generalization over the population in which selection took place. *Non-probability* methods on the other hand use implicit or explicit criteria for selection. Therefore these methods provide no basis for generalization. Non-probability methods are often used in exploratory research, where the focus is on generating new ideas. Probability methods fit better with the objectives of describing and testing studies.

Probability sampling

The are several modes of probability sampling.

Simple random sampling

This is selecting of respondents without any system or criterion. For example haphazardly picking names out of parish registers. The only system allowed in this method is preventing that the same person is invited more than once to participate in the study.

Systematic probability sample

Mostly there is some system in choosing the subjects, as long as this system does not interfere with the aim of the study. For example drawing every 10th name out of a register or inviting the head of the household of house nr. 5 of every street in town. This procedure is also known as 'pseudo-random' sampling and 'ordinal' sampling. A variant of this method is 'random start' sampling. Here some files of a register are chosen at simple random and next used in a systematic way. The procedure is also known as the method of 'inter-penetrating' sub-samples.

Stratified sample More criteria are introduced with *stratified sampling*. Here the distribution of special characteristics in the sample is manipulated. This procedure is also known as 'controlled' sampling, 'over' sampling, or 'optimal' sampling. For example the age of the subjects may be controlled. If the investigator wants the distribution of his sample to be equal to the distribution of age in his population, he draws a sample 'proportionally stratified by age'. If he wants to be sure that there will be enough 90-year-old subjects in his sample he will be inclined to draw a 'disproportionate' sample

with relatively more old people. Stratification can take place at the moment of sampling itself. We then speak of 'stratified' or 'balanced' sampling. Another possibility is random discarding of some of the subjects with characteristics that are sufficiently represented in the sample. We speak here of *post-stratification*. Another way of adjusting the sample afterwards is adding cases in under-represented categories. For instance, if there are too few males in a sample, one can raise the number of male respondents by entering male cases twice. This referred to as *weighing*.

Cluster sample Sometimes the population is too great to take a simple random sample. For instance, the inhabitants of a nation are not all represented in one great file. So for convenience's sake samples are sometimes drawn from *clusters*, such as towns, companies, households, schools, etc. Usually they are selected by simple or stratified random sampling methods. When the cluster is geographically defined, we speak of *area sampling*. The method of first selecting a number of areas, next a number of households within these areas and finally the subjects within the households is called *multi-stage sampling*.

Non-probability sampling

The following non-probability methods of sampling can be distinguished:

Accidental sample

The most widely used method of non-probability sampling is accidental sampling. One simply reaches out and takes the cases that fall to hand, continuing the process until the sample reaches a designed size. For instance, recruiting respondents among one's acquaintances, or calling for volunteers in an advertisement. This way of sampling may involve unintended selectiveness.

Chunk sample

Another method is selecting a specific group of people, for example, a class of students, members of a club, or employees of a firm. This method is called *chunk sampling*. This way of recruiting respondents is also quite selective, but in this case the kind of selectiveness is clearer.

Purposive sample

This is selecting subjects on basis of criteria deemed relevant for the study. In this case selectiveness is intended, and part of the research-design. When criteria are vague and complex, selection often takes place by *experts*. For example in a study of drug addiction a police officer can select high addiction districts in the town or he can bring the investigator into contact with some addicts who on their turn can supply more subjects. That latter variant is called *snowball sampling*. Expert choice is also used when the composition of *contrast-samples* is necessary: for example selection of healthy and unhealthy employees by the factory-doctor.

Quota sample

Another form of non-probability purposive sampling is quota sampling, also called *stratified non-random* sampling and *interviewer-selected* sampling. Here the interviewer chooses the subjects by him self on the basis of instructions. For example, he is instructed to find subjects of certain age, sex and educational level.

Response

'N' symbolizes the number of subjects actually participating in the investigation. Usually the number is lower than the number selected for the study. This may be due to several reasons, e.g.: *Unattainable*: contact could not be established with the subject for reasons of illness, wrong address, change of address, etc. *Refusal*: subject refuses to participate. *Incomplete*: the subject participates, but due to misunderstanding or incomplete responses his data have to be omitted.

The subjects who failed to participate in the study for these or other reasons are known as the *non-response* category. Their number is usually expressed as a percentage of the original sample. A high non-response can interfere with the representatives of the sample. When there is a high non-response due to illness among elderly people, the sample is no longer representative for age and all conclusions may be severely distorted.

To cope with this problem a mini-study of the non-response group is sometimes made in order to establish the degree of deviance on a limited amount of variables; usually some variables which are easily measured such as sex, age and income. If, for instance, it turns out that the percentage of females in the non-response group is significantly higher than in the response group, females are said to be *under-represented* in the sample. If relatively few females did not respond, females are *over-represented.*

3/4.1.3 Assessment

The appraisal of the selected subjects can be denoted by 1) study strategy and 2) measurement method.

Strategy

Assessment strategies differ in at least three respects, in whether subjects are followed over time or not, in whether any change is induced and whether the aim is to compare within a culture or across cultures.

Longitudinal vs. snapshot

A study is called *longitudinal* when it follows people through time. Other words for the same are *panel-study* and *follow-up study*. Most studies limit to one observation only. We refer to these as a *snapshot* study, which is synonymous to *synchronic* or *cross-sectional*. Distinction between data based on longitudinal studies and snapshots is relevant for identification of causal effects. Therefore, findings from longitudinal and snapshot studies are presented separately. In the enumeration of findings in the next volumes, each subject-category is sub-divided in a 'career' section and a 'current' section. Findings based on longitudinal studies are presented in the career-section, findings on antecedents of happiness in the 'earlier' sub-section, and findings on subsequent of happiness in the 'later' sub-section. Most of the findings based on snapshots are presented in the current-section.

Experimental vs. non-experimental

A study is considered *experimental* when during the course of the investigation a change is induced by the investigator and when the effects of that change on the dependent variable are assessed, eventually by comparing with a control group. In all other

cases studies are considered ' *non-experimental*'. Experimental data are of value for identification of causality as well. Findings based on experimental studies are also presented in the 'career' sections within subject-categories, in this case in the sub-section 'change'.

Cross-cultural vs. uni-cultural
Almost all studies involve comparison but some studies are especially designed for particular comparisons. These investigations are called *comparative* studies. When the aim is to contrast different cultures, they are denoted as 'cross-cultural', when nations are compared as 'cross-national'. In this catalog, the term comparative is not used for studies that use contrast groups within nations, not even when purposive samples are compared, such as bright and dull students.

Most findings in this catalog are neither longitudinal, nor experimental, nor comparative. When otherwise, that is indicated with the findings concerned, so that these valuable pins can be found back in the haystack.

Measurement
Information about the subject can be gathered in several ways.

Observation
Anthropologists favor observation of the subject's behavior in their normal daily routine and setting *(field observation or naturalistic observation)*. Psychologists rather tend to observation in a laboratory situation where the subject is confronted with controlled stimuli (*laboratory observation*). The investigator may make his observations hidden behind a one-way mirror *(disguised observation)* or he may decide for *open observation*. If observation involves the sharing of daily activities with the subject, we speak about *participant observation*. Observation may be structured *(systematic or controlled)*, using detailed observation schedules or *unstructured (natural, simple or qualitative observation)*.

Interrogation
A more common way of data gathering in happiness-research is by posing questions. Questions may vary in their degree of structure. *Open-ended* questions, 'free-answer' questions or 'unrestricted-answer' questions leave the subject free to formulate his answer. This is not the case with *closed questions*, also referred to as 'cafeteria' questions, 'multiple-choice' questions or 'fixed-answer' questions. Here, the respondent is asked to choose from a list of assorted words or statements one or more that best represents his view. The latter kind of interrogation is most current in this collection, both in the assessment of happiness and in the measurement of correlates.

When using *direct questions* the interviewer asks directly what he wants to know. Question can also be posed in a more hidden way. In that case the interviewer infers the things he wants to know (e.g. anxiety) from responses to questions about something else (Fe. perceived prevalence of crime, anxiety of average citizen). We refer to that kind of interrogation as *indirect questioning*. Sometimes this latter method involves projective techniques (see below).

In the above discussion of happiness queries (section 2.2) we have seen that happiness

cannot be measured adequately by indirect questioning. Hence, this catalog does not include studies that measure happiness by indirect questioning. Several correlates are measured by indirect questioning however; for instance the number of pleasant reminiscences of the liking of odor measures 'positivism'. See subject-category 'positive personality' (P 4.85). Questions may be presented in written questionnaires. *Structured* questionnaires use specific questions, often with closed answer format. This method is often used in mailed questionnaires and large-scale survey research. *Half-structured* questionnaires use less specific questions, for example, a request to write a story on a certain subject or to give one's opinion on a handful of topics.

Questions can also be posed verbally in a face-to-face interview. In a *structured interview* the interviewer fills out a questionnaire on the basis of the subject's responses on standard questions. This technique is often used in telephonic interviews and in large-scale surveys. It is often referred to as 'standardized personal interview'. The *half-structured*, 'open' or 'qualitative' interview does not use identical questions for all subjects. The interviewer is more sensitive to directions in the conversation opened by the subject himself. Other names for this kind of data gathering are 'non-directive', 'non-schedule', or 'exploratory' interview. When the interviewer concentrates his questions on a limited number of issues, we speak about *focused interview*. If the term *clinical interview* is used, we aim at an interview of this kind that focuses on assessing the psychological condition of the subject: usually in a therapeutic setting. The term *depth interview* denotes an even more specific interview of this kind, were the aim is to tap unconscious motivations and ideas, using techniques of free association, indirect questions and projection.

That latter approach is systematized in schemes for rating the subject's verbal and expressive behavior when being confronted with ambiguous stimuli. This technique requires a subsequent content analysis of the answers (see below). *Pictorial* projective techniques use pictures as projective material; for instance, the inkblot in the Rorschach test or the series of pictures in the Thematic Apperception Test. *Verbal* techniques use sentence-completion tests, word-association test and indirect open-ended questions. Role-playing is sometimes used to evoke projections as well. As such it figures in techniques like psychodrama, socio-drama and doll playing. In these techniques, interrogation merges with observation.

Interviews are usually conducted at the house of the subject, but also in a laboratory setting, in a clinical setting, at the place of work, or by telephone.

The various interrogation techniques differ in vulnerability for measurement bias of different kinds. Differential bias may complicate comparison of findings. Therefore, the excerpt notes systematically which method is used, both for happiness and the correlated variable.

Content analysis

Another way of data gathering is analyzing written documents, such as dairies, essays, correspondence, etc. In case this involves objective, systematic and qualitative description of the manifest content of the documents, the term *content analysis* is used. This requires the development of a scheme of analysis according to which information can be selected and categorized. The term does not concern loose attempts to 'feel one's way into the matter'. Content analysis is also used for coding transcriptions of open. interviews

3/4.2 Words for measured happiness

Under the heading of 'label' in the notation sheet we note the name the original investigator gave to what we call 'happiness'. As mentioned earlier, different labels are often used for similar phenomena. Most investigators use labels like 'morale', 'general satisfaction', 'elation', etc. Sometimes, however, different names are used, indicating quite another interpretation of the observations: for example: 'psychological health' and 'adjustment'. Differences in labeling sometimes go together with differences in conceptualization, but not always.

3/4.2.1 Query on happiness

With the term *query* we aim at the 'indicator' or the 'measurement-instrument' used, the concrete and specific definition of the variables in terms of the operations by which observations are to be categorized. A classification of happiness queries was presented shortly in chapter 2 of this introductory text. The matter is discussed in more detail in the introductory text to the 'Catalog of Happiness Queries'.

Observation schemes

When happiness is assessed in an indirect way by behavioral observation or by projective techniques, the measurement follows usually explicit instructions for observation and scoring. We call this a *scheme of analysis*.

Questions

When happiness is measured by direct questioning the term 'query' refers to the *question* used and its answer-categories. Responses to closed questions are usually recorded on a *rating scale*. For example: 'Do you feel: happy 1----2----3----4----5----6 unhappy'

When such a rating scale has six answer categories it is said to be a *six-point scale*. These items can be points on a linear scale or separate multiple-choice statements, such as 'very happy', 'happy', 'not too happy', or 'unhappy'. When *graphic scales* are used the subject indicates his rating by simply placing a check at the appropriate point on a line that runs from one extreme of the attribute in question to the other.

Often several questions are used to assess one variable and the scores for these questions are added up. In that case we speak of an *index.* For example, popularity can be measured by perceived esteem of one's, boss, one's friend and one's spouse. When the answers on three of these questions are summed up into one score we speak of a *three-item index*. When the investigator assigns equal weight to the items, we speak of a *simple index*, if not, of a *weighted index.*

Sometimes, series of questions are first tested for 'scalability'. A crosscheck is made as to whether other people also consider the question to be indicators of the same variable. Moreover one often tries to select questions in such a way that the answers offer a more accurate picture of the continuum on which the variable may vary. This is commonly called a 'scale'. The word scale here has another meaning than that of the 'rating scale' mentioned above. We therefore stick to the term 'index'. Several types of indexes can be constructed: among other things: 'cumulative' indexes, in which the

items are supposed to represent an increasing monotonous function of the variable. Variants of this type index are e.g. the so-called Thurstone-scales, Likert-scales and Guttman-scales.

3/4.2.2 Distribution

The pattern of responses to a happiness query in a sample is referred to as the *frequency distribution* or shortly 'distribution'. For example, in 1991 Americans responded as follows on a single question on happiness type O-HL/u/sq/v/4/a: very happy 37%, quite happy 54 %, not very happy 7 % and unhappy 1 %, while 1% did not answer. (Data World value Survey 2, which is included in the catalog of happiness in nations).

Central tendency statistics are derived from this frequency distribution, such as the *mean* (average) and the *modus* (most often chosen category). Measures of dispersion can also be computed, for instance the standard deviation. These descriptive statistics are discussed in more detail in chapter 4.

3/4.2.3 Error estimates

Measurement involves inevitably some error and self-reports of happiness seem pretty vulnerable for that. Error can be either systematic (e.g. general overstatement of happiness) or random (e.g. mistakes in answering). If the size of this error can be estimated, one can correct the measured values. For instance, if we know that people tend to overrate their happiness by some 10% we can detract one point of scores on a ten-point scale. Likewise we can correct correlations for random error by 'dis-attenuating' observed values (cf. Headey & Wearing 19??). Measurement error \is estimated by test for 'reliability' and 'validity'. Advanced MTMM techniques allow simultaneous estimates of both (e.g. Saris ??).

Reliability

The term *reliability* refers to the consistency of data yielded by an indicator, irrespective of what it may measure. For instance, if people give different answers to the same question on happiness on Sundays than on working days, that question does not provide consistent information about the (presumed) stable state of mind. Reliability is in fact absence of measurement error'. Prevalence of error in response can be estimated in the following ways:

One way is assessing change in response, in situations and over periods that real change in happiness is unlikely to occur. This is done by repeating the same questions at different times in the course of the interview (*repeat reliability*), or by re-interviewer the same subject again some days or weeks later *(retest reliability)*. Reliability is commonly expressed in the correlation of responses in both instances.

Reliability of measures should not be confused with long term 'stability' of happiness itself. Reliability coefficients of happiness items tend to be low (about +. 50), but stability over the years high. This manifests in a very gradual lowering of retest correlations over time.

Low reliability of an indicator tends to depress correlations with other factors; at least when measurement error is random, and when measures of the correlated factor are not distorted in the same way (correlated error). The muted correlations can be restored to their presumed 'true' level by 'dis-attenuation'. The method for upgrading used here is dividing the observed correlation by the square root of both variables. See

below under 'association'.

Sometimes, the term reliability is also used to denote the degree to which different items of an indicator measure the same phenomenon. This is called testing for *equivalence*. We consider such procedures as validity testing. See below under 'structural validity'.

Validity

An indicator is considered *valid* if it measures what it is supposed to measure: if it is free of bias, systematic or non-sampling errors. This is one of the greatest problems in social research and in happiness research in particular. Does a set of questions on happiness tap the evaluation of life of the subject or does it reflect a value-orientation, a defensive self-image, a social norm, etc.? The validity of a happiness indicator can be assessed in two ways: by the logical consistency of its items and by its correspondence with other indicators of the same or related matters.

Internal validity

In the first case we speak of *internal validity*. This kind of validity is assessed by checking whether the questions or other observational devices we want to pose all represent the same meaning: its *substantive validity*. This can be done by carefully inspecting the matter (*face-validity testing*) or by an inter-subjective procedure of *content-validity testing*, often using judges. The substantive validity of many happiness indicators is discussed in much detail in chapter 4 of the earlier mentioned "Conditions of Happiness" (Veenhoven 1984a). Selection of studies for inclusion in this catalog is based on that validity check.

A final check on this substantive testing is the inter-correlation of responses on items. We call this a testing for *structural validity*, 'equivalence', 'consistency' or 'congruence'. High correlation suggests that items measure the same phenomenon indeed. Current measures of equivalence are *alpha* (Cronbach, 1951) and *omega* (Heise & Bornstedt, 1970), both ranging from zero to one. Still another method is computing *split-half* correlations, which involves comparison of scores on the indicator in two random halves of the sample.

External validity

A second method of validity testing is assessing correspondence with other indicators. We speak here of *external validity*, 'practical' or 'empirical' validity. Two variants can be discerned:

Firstly, assessing the association with other indicators of the same variable can make estimates of validity. For example, a happiness question can be validated on other happiness questions, facial expression, expert ratings, peer report, etc. We speak here of *congruent validity*. This catalog involves many findings on inter-correlation of happiness-indicators. See subject-categories 'Current Happiness' (H 6*) and 'Life-appraisals' (L 4*)

Sometimes happiness indicators are validated on essentially other phenomena that are related to happiness, such as mental health, social adjustment or social participation. We then speak of *concurrent validity*. This method is not very easy to apply in happiness research because the relations of these factors are neither complete nor constant. Still, the interested reader can find a lot of relevant findings in this catalog. See cross-reference category 'Quality of Life'.

When the association with later events assesses congruent or concurrent validity, we speak of *predictive validity*. When phenomena of the past are taken as a point of reference, we speak of *retrodictive validity*.

3/4.3 Words for measured correlate

Correlates of happiness are phenomena that appear to be statistically related with happiness. For example, when the happy appear to be healthier than the unhappy, health is said to be a correlate of happiness.

The excerpts describe correlational findings in separate mini-abstracts. On the next page is an example of such an abstract. On this form we meet with the following terms:

Correlate	Variable of which relation with happiness is investigated (co-variate)
Correlate classification	How correlates are grouped by subject
Correlate code	Character label for class of correlates
Standard name	Verbal label for class of correlates
Authors name	How the variable was originally labeled in the study report
Measured correlate	Applied measure and observed distribution
Measure of correlate	How the correlate was quantified
Frequencies	Observed distribution of ratings
Error estimates	Indications of measurement error in observation, such as repeat-reliability or Conbach's alpha.

3/4.4 Words for observed relationship

Two variables are *related, associated or correlated* when changes in one variable are systematically accompanied by changes in the other. For example, happiness is said to be associated with health, when people who feel happy report less health-complaints than unhappy people do.

3/4.4.1 Statistical association

Association has aspects of 'direction' and 'strength'. Happiness and health are said to be positively associated when happiness goes hand in hand with good health and negatively when it is accompanied by bad health. This *direction* is indicated by + and - signs in the column 'value of association' in the abstracts.

Some practical problems may arise in indicating the direction of the relationship. When we say there is an association of G = +. 50 between happiness and education, the direction is clear, the more education, the more happiness and vice versa. Difficulties arise however, when a variable is labeled so that the direction of the relation is less clear. For example, when an investigator reports an association of G = +. 50 between happiness and marital status. Now it is not clear whether being married is positively associated with happiness or negatively. In the column 'operationalization' (column 2) we therefore indicate less and more of the variable concerned. Usually this is done with cipher, for instance '0' for unmarried- and '1' for married state. An earlier way of notation is 'unmarried vs. married', where the sequence is from less to more

Association reaches its highest possible *value* when all happy people are healthy and all unhappy people are unhealthy. Association is zero when as many people are

healthy as unhappy people and negative when happy people turn out to be less healthy than unhappy people. Values of association are usually expressed in a number between one and zero. For instance: r = -. 25 indicates a modest negative statistical relationship. Still, the meaning of values yielded by different statistics is not always identical. A same relationship may be values -. 25 when expressed in r, and -. 30 when expressed in Gamma. This limits possibilities for comparison of findings expressed in different statistics. Different statistics can to some extend be converted to a same 'effect size'. See Rosenthal (1984). However, the available information does not always allow transformation, and conversion involves often loss of information.

As noted in the above discussion on 'reliability' of happiness indicators, measurement-error reduces observed correlation. If the size of measurement-error is known (e.g. by the above reliability-coefficients), 'true' correlation can be estimated by 'dis-attenuation' of observed ones. Dis-attenuated correlation is computed by dividing observed correlation by the square root of the reliabilities of both variables.

There are many ways of expressing the strength of associations. These *measures of association* use different assumptions and statistical techniques. It would lead us too far to discuss the strong and weak point of the various statistics. We may do with an outline in section 3.3.1.

3/4.4.2 Statistical significance

The term *significance* refers to the likelihood that an observed empirical relationship results from sampling error. A relationship is said to be significant at the .05 level (p< .05) if the likelihood of its being only a function of sampling error is no greater than 5 percent.

Assessing whether chances of sampling error are sufficiently small is called *testing for significance*. Such procedures make sense only when representatives can be assumed. Like in the case of association there are several statistical methods for this purpose. The symbols are explained in section 3.3.2. Often, investigators do not report what kind of test-statistic they used. In such cases column 7 in the excerpt remains blank.

The character 'p' denotes the probability that in spite of the association found in the sample the actual association in the population is zero. For example, when in a sample of the Dutch population we find a correlation of G = +. 40 between health and happiness, which is significant at the 99% level (p <. 01), this means that there is a chance of less than 1% that in the Dutch population as a whole health and happiness are unrelated. Sometimes it is not the chance of a zero association that is computed, but is another point of reference chosen. For example, when the association between health and happiness is +. 30 for males and +. 50 for females, it is possible to compute the likelihood that this difference is due to error. *Unless stated otherwise, significance levels mentioned in the excerpts refer to the probability that there actually is no association at all in the population the sample was drawn from*.

Sometimes the investigator does not report probabilities, but suffices to note that the correlation is 'significant'. In these cases we note *s*. If the investigator considers his results non-significant we note '*ns'*. Usually, p> 05 is considered non-significant.

3/4.4.3 Elaboration of the relationship

If happiness is found statistically related to health, these phenomena are not always equally linked in all parts of that population. Happiness may be strongly related to health among young, adults but less so among the elderly, because older persons expect decline of health suffer fewer side effects in love and work. Inspection for such differences is called *specification*.

If in the example at hand, happiness is indeed more strongly related to health among the young than among the old, the association is said to be *stronger among* the young. If there is no correlation at all in this part of the population, we note *not among* the young. Sometimes the direction of the association is different in a part of the population. For instance, young people whom dislike working can be happier when disabled. We then write *reversed among* the work-shy.

A statistical relationship between happiness and education does not necessarily mean that happiness and education are causally related (either that good education makes happier or that happiness adds to the chance that one does well at school). The variables may in fact be unrelated causally, and the statistical relationship an artifact. Good health and high intelligence could for instance be responsible for a *spurious correlation* because they both add to educational success and to a positive appreciation of life. It is also possible that education as such does not add to happiness, but that it favors indirectly a positive appreciation: for example because it opens doors to good paying jobs.

One same factor can in fact be both a spurious factor and a link in a causal chain between education and happiness. In the above example: good health may inflate the correlation between education and happiness, because it affects both in the same way (which produces spurious correlation). At the same time it can also be responsible for a reality link: happiness fostering good health, which on its turn adds to educational success.

Such effect can be demonstrated by specification procedures as mentioned above. They can also be checked by computing *partial correlations,* that is, assessing the correlation that remains when effects of one or two further variables are checked. Partial correlation coefficients are symbolized with ' rpc '; partial (or standardized) gamma's with 'Gs'. The results of such procedures are noted as follows:

If the correlation between happiness and education appears to be mediated by income, we write that it *disappears when controlled for income*. If at least part of the common variance remains, we note *lower when controlled for income*.

Controls can also demonstrate an actually reversed relationship. Happiness could for instance be positively related to education, while education is in fact detrimental to it. That could be so, when negative effects were masked by the fact that highly educated people are typically born in the higher social ranks and for that reason enjoy greater self-respect and a better financial start. In that case we note that the correlation is *reversed when controlled for* social milieu of origin.

Likewise it is possible that a zero correlation masks a reality link. A positive effect of education on happiness could for instance be disguised by the fact that members of minority groups be disguised by the fact that members of minority groups are over-represented in the higher educational levels, education being the only chance for mobility. These people may in fact take more enjoyment in life because of their educational achievement, but discrimination may still prevent their happiness to be

above average. If such an effect is demonstrated by the existence of positive correlations in both the minority and the majority or when a positive partial correlation appears we say that a correlation did *appear when controlled for* minority status.

Quite often such procedures reveal that suspicions about spuriousness of hypotheses about mediating variables are false. We then note that the correlation appeared *unaffected by* the control variable(s).

These elaborations involve often the comparison of association values: comparison of associations found in sub-populations (mutually and with the association found in the general population) or between controlled and uncontrolled associations. It is often not clear whether the differences that appear are really worthwhile. The original reports do not always tests for significance of differences in correlation. It is possible to compute significance of differences afterwards, provided that one knows the distribution of cases in categories concerned. However, that information is often not available in the reports. Therefore, we applied the rule-of-thumb that differences greater than .10 on range 0 to 1 are worth mentioning, whatever the measure of association. When differences smaller than that appeared, we noted *unaffected by*. In that differences were checked for significance, the results of that check are mentioned in the excerpts.

Measures of association inform us about the strength of association. They do not provide information about the pattern of association, but more or less suggest a linear pattern. However, associations may follow non-linear patterns as well. For example a positive correlation between happiness and education may cover a great variety of patterns, some of which are shown in exhibit 7. Uncovering such patterns is called *elaboration for shape*. Exhibit 8 enumerates the terms used in the excerpts to denote the pattern on the scattergram.

When possible, the excerpts present the % happy in all categories of the correlated variable. This provides no full view on the shape of the relationship, but at least an indication.

Most research-reports do not dwell on the shape of the relationships observed. It is often implicitly assumed that the observed relationships are linear. As a result, the excerpts cannot include that

3/5 HOW EXCERPTS ARE MADE

3/5.1 Procedure

3/5.2 Tools
- Database
- Forms
- Computation program

3/5.1 Procedure

Research reports are read and then excerpted. This is a laborious job, which takes one full workday per report on an average. As indicated above in section 3/1 excerpts involve three parts. First the design of the investigation is characterized, second the happiness queries are described in full detail, and third the findings are entered piece by piece. Excerpts are entered in a database, initially marked as provisional.

The provisional excerpt is printed on paper, and then checked by another excerpter. Mostly the corrected version passes another check. The final version is send to the original investigator, if available. Investigators are asked to check whether the excerpt reflects their findings correctly. They are also invited to add findings that were not in the report.

When all the available information is gathered the project leader decides to release the excerpt. Only then it is really added to the database and published on the web.

Excerpts are made by social scientists, mostly students and retired academics.

3/5.2 Tools

Database

The excerpts are entered in a database system. This is an MS-Access application, specially designed for this purpose by Henk DeHeer. It is called a 'Finding Browser'. The software is available on request.

Notation forms

Because many studies are too complex to be entered directly on the PC, there is a form on paper on which one can make preliminary notes. This form is presented on appendix A.

The database system requires that the data be entered on one computer. Still part of the work can be done elsewhere. There is also an electronic form on which one can enter most of the information. From that form, text block can be copied to the appropriate fields of the database. The electronic form is presented on appendix B.

Computation program

Excerpting requires sometimes that statistics are computed, or transformed. For that purpose there is a program in MS-Excel, which provides routines for the following calculations:

- Computing means and standard deviations from a frequency table
- Transforming means and standard deviations to range 0-10
- Computing 95% confidence intervals around these statistics of central tendency
- Computing some statistics from a frequency table

3/6 SUMMARY

Standard-excerpts of each research-report are made. These excerpts consist of three parts, which can be retrieved independently.

Part 1 **Study**
- Bibliographics of the research report
- Design of the empirical investigation

This part reports never more than one study

Part 2 **Measured happiness**
- The query used for assessing happiness
- The distribution of happiness observed

When the study involved more than one indicator of happiness, this part reports these separately.

Part 3 **Correlational findings** (finding-abstracts)
- Summary of study and happiness query (part1 and 2)
- Measurement of correlated variable
- Observed statistical relationship with happiness

When a study related happiness to more than one variable, this part presents more than one mini-abstract.

This chapter describes what information is comprised in each of these parts and defines the technical term used.

Appendix A
Entry preparation form, paper version

World Database of Happiness, Correlational Findings,

N O T A T I O N F O R M

STUDY Excerpt by: dd:

Report ----------

Code:

Authors:

Title:

Bibliographics:

Design ----------

Population:

Public:

Place:

Time:

Sample:

Non-response:

N:

Data gathering:

Remarks:

MEASURE OF HAPPINESS used in study

Fill out separate form for each measure of happiness
Number _____ of _____

Study code: Page:

Authors label:

Text:*note keywords to check precise item in database*

Measure type:*if unclassificable, enter X and fill out mutation form*

Meaning: Scope: Time-frame:

Method: Observation: Rating: Variant:

Measured values: Range: from____ to_____ Mean: SD:

% responses a: b: c: d: e: f: g: h: i: j: k:

Error estimates:

Remarks:

FINDING

Fill out separate form for each measure of happiness

Study
Code
Page: Number ___of_____

Measured correlate
Authors label Our code

Measurement *End each line with return*	Measured values
	Error estimates
	Remarks

Observed statistical relation with happiness

Measures of Happiness	Statistics	Elaboration/remarks *End each line with return*

Appendix B

Entry-preparation form, electronic version

REFERENCES

Headey, B. & Wearing, A (1992)
Understanding happiness. A theory of subjective wellbeing
Longman Cheshire, Melbourne, Australia

Saris, W.E., VamWijk, T., & Scherpenzeel, A. (1998)
Validity and Reliability of Subjective Indicators
Social Indicators Research, vol. 45, pp.173-199

VanderMerwe, C. (1974)
Thesaurus of Social Research Terminology
Universitaire Pers Rotterdam