

Effectiveness of Preventive Stress Interventions

A quantitative and qualitative review of effect and its predictors



Jaap Denissen

9622144

Supervisors: Drs. Eva Jané-Llopis

Prof. Dr. Clemens M.H. Hosman

Dept. of Clinical Psychology and Personality

Prevention Research Centre

University of Nijmegen

October 2000

Für Claudia

For the successful completion of my essay, I wish to thank everybody that has helped and supported me during the process of thinking, writing and revising. First of all, these are my girlfriend Claudia, and my family and friends. Also, I thank Eva Llopis for her patience and for being a guide in my development as a scholar.

Essen 26 October,
Jaap Denissen

Table of contents

ABSTRACT.....	1
INTRODUCTION.....	2
CHAPTER 1: THE STRESS PROCESS.....	6
1.1 Introduction.....	6
1.2 Strain reactions.....	7
1.2.1 Physical reactions.....	7
1.2.2 Behavioural reactions.....	8
1.2.3 Psychological reactions.....	8
1.2.4 Societal impact of stress.....	10
1.3 Stressors.....	11
1.3.1 Types of stressors.....	11
1.3.2 Is the presence of stressors on the work floor increasing?.....	12
1.4 Mediator and moderator variables.....	14
1.5 An integrative model of the stress process.....	15
1.5.1 Introduction.....	15
1.5.2 Description of the model.....	15
CHAPTER 2: STRESS INTERVENTIONS.....	20
2.1 Introduction.....	20
2.2 Characteristics of prevention programs.....	20
2.2.1 Actor/program provider.....	21
2.2.2 Target group.....	21
2.2.3 Types of goals.....	21
2.2.4 Risk and protective factors.....	22
2.2.5 Instruments.....	23
2.2.6 System level.....	23
2.2.7 Time aspects.....	23
2.2.8 Setting of the intervention.....	24
2.3 Methods for stress reduction.....	24
2.3.1 Individual methods.....	25
2.3.2 Organisational methods.....	26
2.4 Implementing the program.....	27
CHAPTER 3: METHOD.....	28
3.1 Introduction: Why perform a meta-analysis?.....	28
3.2 Description of the analysis.....	28
3.2.1 Research questions.....	29

3.2.2 Literature search and selection of studies.....	30
3.2.3 Coding system and procedure	33
3.2.4 Calculations of effect size	35
3.2.5 Analysing, reviewing and interpretation of the results.....	36
CHAPTER 4: RESULTS.....	38
4.1 Research question A: Magnitude of effect size.....	38
4.2 Research question B: Homogeneity analysis.....	40
4.2 Research question C: Decisions on research hypotheses.....	41
4.2.1 Quantitative analysis	41
4.2.2 Qualitative analysis	49
CHAPTER 5: DISCUSSION.....	54
5.1 Overview of results	54
5.1.1 Research question A: Magnitude of effect size	54
5.1.2 Research question B: Homogeneity analysis	54
5.1.3 Research question C: Decisions on research hypotheses.....	54
5.2 Implications for prevention theory.....	55
5.2.1 Earlier effect reviews	55
5.2.2 Implications for effect hypotheses	56
5.3 Limitations of the study.....	59
5.3.1 Ambiguities in the coding procedure.....	59
5.3.2 Small sample	60
5.3.3 Keeping the outliers in the quantitative analysis	60
5.3.4 Validity of the qualitative analysis	60
5.4 Recommendations for future theory and practice.....	61
5.4.1 Implications for future studies	61
5.4.2 Implications for stress and intervention theory	61
5.4.3 Implications for future practice.....	62
REFERENCES.....	63
LIST OF INCLUDED STUDIES.....	65
APPENDIX 1: LIST OF USED PSYCLIT SEARCH TERMS.....	67
APPENDIX 2: SHORTENED VERSION OF THE CODING SHEET.....	70
APPENDIX 3: OPERATIONALISATION OF “DESIGN” ITEMS	71
APPENDIX 4: DESCRIPTION OF OUTLIERS	72

Abstract

The present study is a review of the effectiveness of preventive stress interventions. First of all, it was tried to determine the effect size of a sample of such interventions. Secondly, it was assessed whether this sample was homogeneous. As a third objective, it was tried to identify some of the working factors, or effect predictors, of the stress interventions. It was hypothesised that the following characteristics would be positively related to effect: carrying out the intervention after working hours, a high-risk target group, an on site intervention location, a longer intervention duration and the use of multiple methods. Furthermore, a differential effectiveness on the basis of intervention method was expected. Organisational interventions were expected to be equally effective as worker-oriented ones. At follow-up, organisational interventions were expected to increase in effect relative to worker-oriented ones. Finally, it was expected that the quality of the intervention design would be effect-related. As an additional factor, the effect of publication year was assessed. In order to address the research questions, a meta-analysis was carried out. Through an extensive literature search, 47 evaluation studies were retrieved. These articles were coded using a specialised coding system. On the basis of their outcomes, a standardised measure of effect size was calculated. Results showed a medium intervention effect on psychological outcomes. Small effects were found for behavioural outcomes, whereas the effect for cognitive, physiological and organisational measures was very small to zero effects. Due to 5 outlier studies, the sample was heterogeneous. As expected, after work hours programs were (marginally) more effective. Moreover, the quality of the interventions design was positively related to effect. Contrary to the expectations, results indicated that individual-oriented and off site programs were more effective. Finally, publication year was negatively related to effect. The other results were non-significant. The statistical conclusions were supplemented by a qualitative analysis. This largely confirmed the quantitative findings and suggested a number of additional factors that appear effect-related. At the end of the study, suggestions for further research are given.

Introduction

Stress is a very complex phenomenon. Take for the example the number of disciplines that occupy themselves with the subject. Medicine is contributing with findings of biochemical correlates of stress and has demonstrated that it can have an effect on the number of immune cells in the human body. Economists are focusing on the effects that changes in production methods have on stress and how the costs of this development can be prevented or reduced. There are also scholars who have pointed at the ideological background of the stress concept (or “discourse”, in Foucauldian terms). And there is of course psychology, a heterogeneous discipline in itself, where schools such as humanistic psychology, cognitive-behavioural psychology and labour psychology all have different things to say about the topic. Maybe this diversity is one of the reasons that there is not yet a widely accepted definition and theory on stress (Gaillard, 1996; Kahn and Byosiere, 1992). Despite of this, academic interest is vastly increasing. Critics have remarked that much of the attention seems to be due to social and scientific trends (Newton, 1995). However, there appear to be very objective and consequential reasons too. As can be seen in later in this essay (chapter 1.2), one of the most important causes for this vast attention lies in the outcomes of stress, which are well described as “real, painful and costly” (Kahn and Byosiere, 1992).

As shall be described in the following chapter, these aversive consequences of stress are large in number and serious in nature. All of the above mentioned disciplines encounter such stress related problems. For the medical profession, stress mainly appears as biological reactions. Such physiological changes belong to the first documented results of stress. Hans Selye for example, one of the founding fathers of stress, studied the responses of rats in reaction to aversive environmental stimuli. The changes he reported included enlargements of the adrenal gland cortex, involution of thymus and lymphatic structures, and the development of ulcers (Murphy, 1996; Newton, 1995). For people, stress can have very serious physiological consequences on cardiac, biochemical and gastrointestinal systems (Kahn and Byosiere, 1992). Moreover, a long-term consequence of stress is the “wear and tear” of the body’s immune system. One reason for this is that cortisol, an important stress hormone, suppresses the body’s immune cells. Another class of responses to stress pertains to behaviour. Stress can cause multiple aversive changes that can negatively impact both the person himself and broader systems, such as his organisation or family. Examples of this include alcoholism, absenteeism and role decrements.

Most important for the preventive psychologist, however, are outcomes relating to psychological health and ill health. Does stress affect people’s minds just as it does their bodies and behaviour? Almost any stress theory would answer this question with yes: there are numerous well-established psychological changes that have been linked to variations in stress levels. In one review by Kahn and Byosiere (1992), as many as 43 possible outcomes are mentioned. Among these outcomes are organisation-related factors as job satisfaction and commitment; emotional states like boredom, confusion and frustration; and self-concept variables such as self-esteem and self-confidence. Moreover, stress has been linked to psychopathological symptomatology such as depression, anxiety and irritability (Murphy, 1996). This connection is supported by a number of sources, which will be elaborated in more detail in the first chapter. Here, research shall be presented that found significant correlations between stress on one side and depression and anxiety on the other. Moreover, the risk factors for depression share a great deal of face-similarity with the stress construct. For example, what to think of “a high level of aversive events”, a risk factor reviewed by Veltman and Hosman (1996), which seems to overlap very clearly with a stressing environment.

Besides reviewing the most important outcomes of stress, an overview of stressors and so-called mediator/moderator variables shall also be given. To begin with the first, a stressor could be defined as a threatening factor in or outside the individual (Lazarus, 1995). One area that has been reported as potentially very stressful is the workplace. The essay shall give a brief overview of the most important ones, such as role ambiguity and job uncertainty. A “stressor” does not always produce the same fixed outcome, however. Its impact is greatly influenced by other important factors, such as the relative gravity of the stressor. Getting new neighbours is probably not going to produce such dramatic changes as the death of a loved one. Factors that have a direct influence on the chain of effects are called mediator

variables¹. Crucial in these mediator variables is the way a person copes (deals with) a stressor. However, choice of an appropriate coping mechanism is in turn influenced by so-called moderator factors². This choice depends on the person of the worker and his and social support (Gaillard, 1996). The most common types of personality and social support factors shall be reviewed in chapter 1.

When the role of stressors, stress reactions and mediator/moderator factors will have been clarified, the essay shall provide a model developed by Hosman (1998) to summarise this information. This Integrative Stress-theoretical (IS) model is an effort to integrate research findings into one coherent whole. It consists of multiple steps such as the existence of a stressor, its perception and appraisal by the individual, and the preparation and execution of the effort to cope with the stressor. The model, which is presented in chapter 1.4, shall include most of the evidence that is reviewed in chapter 1 and shall thus be a very comprehensive and useful theoretical background to understand the stress process. With the help of this framework, it is also easier to understand the rationale behind the numerous efforts to intervene in the stress process. Such a deeper understanding is critical for the future design of more effective and successful programs.

In stress interventions, a number of distinctions can be made. What is the target group for a certain program? Does the program explicitly state its goals or remain vague about them? Was the program carried out with people that were already at risk for stress complaints or was the intervention implemented in a group that did not yet show signs of developing problems? The number of such distinctions is in principle endless. They might sound arbitrary too, if it wasn't for the fact that many of these factors actually seem to make a difference. For some categorising variables, at least, it can be proven that they are linked with eventual program effectiveness. Variables for which this is the case are called effect predictors, because their presence in the program characteristics predicts more successful results. The present essay shall analyse the data of a number of evaluation studies in order to reach some conclusions on the determinants of such success. For this purpose, a number of effect hypotheses were generated that, when confirmed, result in evidence for the existence of effect predictors.

Determining such effect predictors would be highly useful. For a long time, there did hardly exist effect studies, so the value of preventive efforts was not without doubts (Kahn and Byosiere, 1992). More recent studies did research the effectiveness of stress interventions, though, and this has led to the conclusion that they can indeed contribute to reductions in stress (Murphy, 1996). What remains unclear, however, are the mechanisms that contribute to this change. Moreover, little is known about the so-called efficacy of preventive stress programs. Whereas effectiveness refers to the success of a program within a given context, efficacy pertains to the question whether the program *as such* would also be effective in other contexts and with other providers. When drawing this distinction to the stress field, it is striking to see that almost all studies are developed and implemented by the researchers themselves. When such programs are shown to be effective, it is a reasonable question whether this has more to do with the

¹ For a definition see Bunce (1997): "A mediator is "the generative mechanism through which the focal independent variable is able to influence the dependent variable of interest."

² A moderator is a "variable that affects the direction and/or strength of the relation between independent or predictor variable and a dependent or criterion variable." (Bunce, 1997)

particular characteristics of the program's providers or implementation context than with the features of the program itself.

A potentially very powerful effect predictor is the method that is used to reduce stress. In selecting a methodology, one can choose between a variety of approaches. In the worker-oriented category, these include cognitive-behavioural skills training, which targets cognitive structures and coping skills of the client. Moreover, worker-oriented interventions include muscle relaxation, exercise and biofeedback. These interventions all involve physical changes that are either beneficial in reducing already built up tension or make the body more "fit" to handle new stressors. Finally, there is meditation, which is based on invoking a state of mental calmness. In the category of environmental interventions, different methods can be most easily characterised according to their target. Here, there are three classes of interventions. Firstly, it is possible that the program targets the amount of social support the worker experiences. Secondly, the degree of control, or decision latitude, can be enhanced. Finally, it is possible to target stress in a very direct way by taking away or reducing the stressor itself (see chapter 2 for a more detailed description of organisational programs).

As mentioned above, the techniques of biofeedback, exercise and relaxation all focus on physical changes. Other methods, such as cognitive-behavioural training, target the psyche of the worker. One hypothesis is that these differences in focus will be reflected in different patterns of effectiveness. Such a prediction is supported by Murphy's (1996) review on worker-oriented interventions, in which very outcome-specific effects were found (e.g. muscle relaxation and biofeedback were only effective on physiological outcomes). It is hypothesised that this will again be the case: **methods are mainly effective on the type of processes at which they aim**. Another hypothesis concerns the distinction between worker- and environment-oriented interventions. Here, it is predicted that **organisational measures are at least as effective as individual ones**. This would be in line with numerous authors (e.g. Daniels, 1996), who state that the two approaches are equally important in reducing stress. A third hypothesis concerning the used method is that **the longer the follow-up, the more effective organisational interventions become in relation to individual measures**. One of the reasons for this effect could be that environmental programs take away the stressor, whereas worker-oriented ones focus on individual changes. To maintain their results, the latter would require booster sessions but, as these are seldom provided, their effectiveness is expected to diminish.

Often, programs do not stick to a single method, though. It is quite common that an intervention uses more than one method to accomplish a reduction in stress. Suppose for example that highly tensed personnel is first being massaged in order to relieve the worst tension. In a second phase they could then receive cognitive-behavioural skills training to prevent such a build-up in the future. A hypothesis is that such **multi-method programs are more effective**. The background for this hypothesis roots in the so-called span of change that such interventions are able of. Suppose that, say, counterproductive cognitions account for 10% of the variance in depressed mood and stressing interpersonal interactions account for another 10%. Programs that have both cognition and interpersonal components thus have a span of 20% of possible change, which is much higher than each separate component would have. Span of change could also be a rationale for another effect hypothesis, namely that **programs aimed at a relatively high-risk group are more effective than programmes that do not**. High-risk workers probably have a larger number or more powerful risk factors for stress problems, which makes their span of change considerably higher.

In two hypotheses, learning principles play an important role. The first prediction involves the effect of the timing of a program. Participants of programs that are run during the working day are probably still stressed. After all, work is not yet done and demands are still waiting once the intervention has ended. Subjects of after working hours programs are thought to be less stressed, as for them, the working day is over. For these after-hours participants, learning is thought to be easier. This prediction is backed up by the so-called Yerkes-Dodson law of learning, which states that high levels of stress interfere with the learning of new skills (Liebermann, 1999). Therefore, **after working hours programs are hypothesised to be more effective**. Another hypothesis that roots in a learning principle is the prediction that **on site interventions have greater success**. The principle that plays a role here is that learning generalises more easily to situations that are similar to the environment where the new knowledge or skill has been

acquired. Because of this, stress management skills that are learned at the workplace are expected to generalise more easily to practical work situations. Thus, more stress reduction can take place.

A final hypothesis has to do with the quality of the methodological design of the intervention study. This is different across studies because researchers use reliable methodological standards, such as control groups and randomisation, whereas other studies are more loosely set up. In this essay, a hypothesis is that **the quality of the research design is linked with program effect**. Such a link with effectiveness was also suggested by Wolf in his 1986 review of the meta-analysis technique. There are a number of possible reasons for this effect. When the study procedure does not use an equivalent comparison group, for example, it is possible that this leads to higher reported effect. Maybe the experimental group was more motivated to change and therefore more eager to comply with the program's goals. In this case, methodological design would be negatively related to effect.

In total, there are thus 9 different effect predictors that shall be addressed. Reaching a conclusion on these hypotheses is not as easy as it might seem however. How can it ever become clear that factor X predicts effect in program Y? For this problem, there seem to be several alternative solutions. One alternative is to set up an experiment and see if an experimental group receiving a program with factor X is better of than a control group with the same program minus that factor. This is relatively troublesome, however, because it is very hard to draw hard scientific conclusion on the basis of one study. After all, success can be due to other factors than the program itself, such as the motivation of the provider. Another possibility is the exploration of subject expert opinions. However, these have the disadvantage of being influenced by all kinds of human biases. Thus, valid conclusions regarding the research hypotheses are both threatened by lack of reliability and a high degree of subjectivity. In this essay, it was therefore chosen to perform a meta-analysis. This is a technique that makes it possible to link certain previously coded variables to a measure of effect size. This method has the great advantage that it is not limited to a single research because a greater amount of studies is used. Moreover, it is relatively objective because it uses statistics instead of human opinions (Wolf, 1986).

The present meta-analysis will be performed as follows. First of all, studies that provided evaluations of relevant stress interventions were retrieved. This was done using an extensive search procedure in scientific databases such as PsycLit and references in relevant articles. When this was done, they were coded using a specialised coding system that was designed for classifying prevention programs. The result of this step is a set of quantified variables that can be statistically elaborated. The most important variable in this process is the measure of effect size called Cohen's *d*. This is a statistic that is calculated on the basis of the effect of a study compared to its Standard Deviated, which makes possible a standardised comparison across studies. In the statistical elaboration of the data, the different coding variables were then linked to this effect size measure. This will make it possible to see whether one of these coding characteristics is related to study effect. In this case, strong evidence for the existence of an effect predictor shall have been found.

To summarise the structure of the essay, the first chapter shall present some evidence on types of stress reactions, stressors and mediator/moderator variables. Also, a model that tries to integrate these factors (the IS-model by Hosman, 1998) shall be introduced here. After this, programs that aim to prevent such stress-related problems will be described in chapter 2. This chapter shall also include a listing of the most important dimensions of these programs. Moreover, some of the reasons to prevent stress shall be covered, as well as a review of common types of programs. These programs can be described in multiple characterising variables. When such a variable is linked with program effectiveness, it is called an effect predictor (chapter 2). This study is concerned with the identification of some possible effect predictors. The method that shall be used for this purpose is a meta-analysis, of which a full description shall be given in chapter 3. An important part of this procedure is the coding of study characteristics into research variables on the basis of which statistical inferences can be drawn. It is hoped that this makes possible a conclusion on the above-mentioned hypotheses is possible (for a more detailed description of the hypotheses, see chapter 2; for the results of the analysis, see chapter 4). Finally, in the last chapter (chapter 5), the outcomes of the analysis shall be discussed in terms of their implications for the current state of the field and its direction for the future.

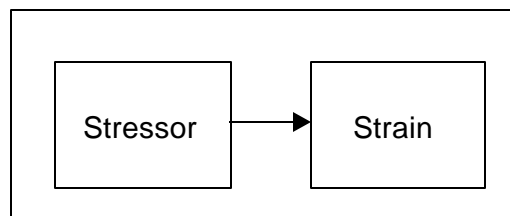
Chapter 1: The stress process

1.1 Introduction

When tracing back the history of stress research, the first great hallmarks originate in the early 20th century. There, scientists like Hans Selye and Walter Cannon came with the first academic reports. They found out that certain environmental conditions (such as dangerous animals or chemical toxicants) produce a universal pattern of reactions in organisms. For example, Selye injected laboratory rats with various toxic substances and exposed them to extreme temperatures. He observed that this led to a universal pattern of physiological changes, which included an enlargement of the adrenal gland cortex, involution of thymus and lymphatic structures, and the development of ulcers. He termed this universal pattern the general adaptation syndrome (GAS), which is a precursor of the stress concept. Another "founding father" of stress was Walter Cannon. He also related to the stress concept when he described a general human reaction to dangerous situations. This reaction, clearly rooted in evolutionary history, is either to fight the attacker or to run away from it. To prepare these reaction modes, the subject is physiologically aroused through bodily reactions such as elevated heart rate and blood pressure, increased blood flow to so-called sympathetic regions (such as major muscle groups or the brain) and a decrease in vegetative functions. It were findings such as these that first indicated that man's reactions to threatening environmental stimuli can influence the body and, as would become clear soon after, also the mind. With these findings, the position of stress as a construct of interest was established. In earlier days, the word stress only referred to some kind of external pressure on an object. Since Selye and Cannon, it became known in the modern meaning of "the non-specific response of the body to any demand" (Murphy's definition, 1986).

This above-outlined process by which a strain reaction results from a threatening stimulus is called the stressor-strain process. In this terminology, stressor stands for some threatening factor inside or outside the individual, whereas strain pertains to the reaction to this (Lazarus, 1995). An example is the development of ulcers as a result of chronic work overload. Here, work overload is the stressor, whereas the ulcers can be thought of as the resulting strain reaction. In a model, this process can be outlined as follows.

Figure 1.1
General stressor-strain model



The process that is outlined by the model makes clear that there are always at least two aspects involved in stress: a stressor and a stress reaction. As a basis for further elaboration, this definition is very useful. However, many important aspects remain quite general and vague. When a stimulus can be treated as a stressor is not stated, for example. Moreover, how a stressor results in strain is not made explicit. Therefore, this chapter shall further elaborate both sides of the stressor-strain relationship. In the first section (1.2), the strain side shall be discussed. The most important types of strain reactions, which include reactions of physical, behavioural and psychological nature, shall be described. Special attention shall be given to psychological consequences of stress, on which the analysis shall focus. After this, in section 1.3, the most common work-related stressors shall be covered. The final category of stress variables is a number of factors that influence the relationship between stressor and stress. These mediator and moderator variables shall be introduced in the third section of this chapter. When all this is done, a model shall be presented that tries to integrate these interrelated variables into a structural whole. With the help of this so-called Integrative Stress-theoretical model, a more complete picture of the stress process will be achieved. This can serve as a background for reading and understanding the rest of the chapters.

1.2 Strain reactions

As stated in the introduction, defining stress is not an easy task. Especially, it is often difficult to specify the precise role that each of its different determinants has in relation to the others. This can for example be seen in the definition of Kahn and Byosiere (1992), which, focused at defining a stressor, is able to elaborate strain as well.

“A stimulus is defined as stressful (i.e., a stressor) because its effect is assumed, hypothesised and (cumulatively) demonstrated to have certain undesirable effects -adverse physiological changes of the kind described by Selye [...], decrements in role performance, emotional tensions, onset of physical symptoms such as sleep disorders and gastric disturbances and so forth- in the population about which we prognose to generalise.” (p. 575)

In this section, these reactions shall further be elaborated. For this purpose, responses to stressors shall be divided into three broad categories: physiological, behavioural and psychological reactions. Because the present study was written in the framework of clinical psychology and the prevention of mental illness, the accent shall lay on psychological effects. After the review of these three classes of responses, some broader consequences of stress shall be described. As shall be outlined here, stress not only results in biological, behavioural and psychological reactions, but also in very significant medical, societal and economic costs.

1.2.1 Physical reactions

As can be read back, the first “discoverers” of stress focused mainly on bodily reactions to external threats. After them, much additional research has followed and this has led to the establishment of a great deal of other outcomes. Following a distinction by Kahn and Byosiere (1992), three categories of bodily reactions can be distinguished. First, there are cardiovascular symptoms, such as blood pressure, cholesterol level, cardiac activation, etc. Secondly, there are biochemical outcomes, which include alterations in catecholamines (such as epinephrine, norepinephrine and dopamine), corticosteroids (cortisol) and uric acid. Thirdly, there are gastrointestinal symptoms, which can for example exist of peptic ulcers. Although much of these outcomes are relatively short-term, there also exist serious long-term consequences. After a longer period of arousal, the body can become exhausted and show long-term physiological responses. This process is also known as wear and tear. In this process, stress can lead to chronic health effects. Although much of the mechanisms in this process are still unclear, the following factors have been shown to play a role (all from Gaillard, 1996).

-Overactivity

People under stress react more intense than necessary to external demands (Gaillard, 1996). Many researchers (such as Karasek and Theorell (1990) in Gaillard, 1996) see in this reaction the main cause of pathophysiological processes.

-Insufficient recovery

Central in this process is the fact that stress reactions leave behind a residue. When these residues are not removed, serious consequences can eventually accumulate. Insufficient recovery is often stated in combination with overactivity, which is because overactivity results in greater residues. The result of this is an accelerated accumulation.

-Higher resting values

Stressed individuals are known to have higher physiological activation while at rest (e.g. during sleep). These higher values have shown to be a good predictor of health damage, especially in the domain of cardiovascular activity.

-Sustained activation

Normally, acute reactions to stressors (e.g. adrenaline) are reduced to normal values when the threat is over. However, higher order mechanisms in the central nervous system, which are being controlled by psychological factors (e.g. anxiety), can block these deactivation systems. In this way, stress can lead to an enduring arousal of the nervous system.

-Disrupted set points

For certain kinds of physiological activation processes, the body has set points (normal values of activation). These set points are part of certain feedback mechanisms that regulate bodily activation. When these mechanisms are disrupted, this means that the psychobiological equilibrium is thrown off balance and physiological functions do not return to their starting values.

-Suppression of the immune system

It is known that sustained sympathetic activation and elevated cortisol levels lead to a decrease in immune system functioning. For example, people under stress produce less so-called killer-T cells. These cells are responsible for warding off alien cells, thereby protecting the organism from disease. This causes stressed individuals to be more susceptible to infections or viral threats.

1.2.2 Behavioural reactions

In the above section (1.2.1), bodily reactions to stress were described. Not only the body reacts to a stressor, however. In a second category of reactions, stress can also penetrate behaviour. This might not be surprising, because tension in the body produces parallel psychological stress. Sometimes, it stays there (as will be elaborated further when discussing the third class of psychological stress responses) and sometimes it is "acted out" in behaviour. There are many instances of such stress reactive behaviours. Because the current analysis shall also focus on behavioural reactions to worksite stressors, especially in the context of the organisation, it is interesting to review some of the most important organisational effects. They shall be listed according to a division made by Kahn and Byosiere (1992), who grouped stressor-evoked behaviours along five broad categories. Although most of these categories pertain to labour/organisational life, they can be easily expanded to other life settings. What is important to realise is that stress can disrupt important life patterns or roles. For example, the relational equivalent of flight from job could be the termination of the relationship. Aggressive behaviours could focus on child abuse instead of on stealing on the job. In principle, stress can affect many realms of behaviour, but as stated above, the essay shall predominately focus on organisational outcomes. Below, the most important of these reactions are listed:

-Degradation/disruption of the work role itself

This consists of decrements in job performance, accidents and errors, and alcohol and drug use at work. These are very costly behaviours for employers because they lower productivity and effectiveness.

-Aggressive behaviour at work

This includes counterproductive acts, such as stealing, purposeful damage, spreading rumours, etc.

-Flight from job

When work becomes too stressful, workers often try to "escape" these aversive environments by absenteeism and turnover (leaving the organisation for other employment). Other escapist behaviours include early retirement and even strikes (Kahn and Byosiere, 1992).

-Degradation/disruption of other life roles

In this category, stressful work spills over into private and social life. Life roles that are shown to suffer from stress include the role of spouse (e.g. spouse abuse), friend and citizen.

-Self-damaging behaviours

These behaviours can cause high medical costs. Excessive smoking, for example, can lead to lung cancer, a disease with often-fatal consequences and costly treatments. Other examples of self-damaging behaviours include alcohol and drug use, caffeine use and accidents.

1.2.3 Psychological reactions

The third broad category of reactions to stressors consists of psychological reactions. Such psychological factors are highly important in the experience of stress. In modern industrial life, work is

increasingly dominated by cognitive demands and it is therefore not surprising that the problems associated with adverse working conditions are psychological as well (Gaillard, 1996). Psychological effects have been reasonably well studied and are considered as painful and costly (Kahn and Byosiere, 1992). There exist many types of these responses. These include organisational/job outcomes such as job satisfaction, organisational commitment and turnover intent. Psychological reactions also include emotional states, such as boredom, confusion, frustration, irritation, etc., and self-concept variables like depersonalisation, self-confidence and self-esteem. Moreover, a common psychological strain reaction to these demands is burnout, a stress syndrome that is characterised by three symptoms: (1) the feeling of being extremely tired and empty ("burned out"), (2) a decreased involvement in others, especially in patients, students or clients, (3) and a lowered job competency (Gaillard, 1996). As a final type of responses, there is even evidence that stress is linked with psychopathological outcomes such as depression and anxiety.

Because it shall constitute the core of the present essay and its analysis, the stress-psychopathology link needs to be elaborated further. For this purpose, numerous sources were used. Firstly, some individual correlational studies shall be explored that researched the relationship between stressors and multiple outcomes of psychiatric distress. Secondly, a review of risk factors for depression by Veltman and Hosman (1996) shall be presented. In the following section, both of these sources shall be described in more detail.

Evidence from correlational research

There are a substantial number of correlational studies that support the conclusion that stress can be linked with pathology. Caplan and Jones (1975), for example, investigated the effect of a computer shutdown on psychological stress. 73 male users of a university computer system were used as subjects. Because these people were highly dependent of the university computer (note that in 1975 people did not have personal computers yet), an increase in stress level was expected. It was indeed found that workload and role ambiguity, two common measures of stress, were higher during the shutdown than in a control period. Moreover, workload correlated significant with anxiety, whereas role ambiguity was related to both anxiety and depression. Thus, although the study lacked a control group, it strongly indicated that stress can be linked to psychopathology.

In another study, Beehr (1976) held a large survey among 651 respondents. These included employees from a printing company, a research and development company, two automotive supply companies and four service departments of a hospital. The survey measured the relationship between role ambiguity and a large number of strain variables. All the included psychological variables (i.e. job satisfaction, life satisfaction, low self-esteem and depressed mood) correlated significantly with the role strain measure. Similarly, Ganster, Fusilier and Mayes (1986), collected data from 326 contracting firm employees. Measures included 6 common stressors (role, conflict, role ambiguity, work underload, lack of variability, underutilisation and responsibility) and 4 outcome variables (life dissatisfaction, job dissatisfaction, somatic complaints and depression). Of these, life satisfaction was associated with role ambiguity and underutilisation. Somatic complaints was linked to role conflict, role ambiguity and underutilisation. For depression and job dissatisfaction, connections were even more powerful. Depression correlated significantly with all stressors but lack of variability and job satisfaction with all stressors except responsibility.

Evidence from epidemiological research

Other evidence that points to a connection between stressor and mental illness comes from research that addresses the risk factors for mental illness. In the section below, the example of depression shall be elaborated in detail. For this purpose, an overview shall be presented of the variables that are associated with the onset of a depressive disorder. When inspecting these variables, some clear points of overlap with the stress concept arise. The risk factors that are reviewed are based on a project designing a prevention program for depression among adolescents (Veltman, Ruiters and Hosman, 1996).

-A first risk factor that was found is negative selfvaluation. Because one of the psychological outcomes of stress is lowered self-esteem and self-confidence (Kahn and Byosiere, 1992), stress could lead to a greater risk for depression.

-Veltman and Hosman (1996) also found a high level of aversive events to be related with the onset of depression. Because stress can be thought of as aversive environmental pressure, this risk factor has considerable overlap with the concept of stress itself. Again, this provides a link by which stress could be a cause of psychopathology.

-Internal, stable and global attributions in uncontrollable, negative situations were another risk factor. Because stressful situations are often uncontrollable (e.g. a company reorganisation or chronic poverty), stressed people with these thinking patterns have a considerable risk of developing depressive patterns.

-A fourth risk factor was having interpersonal problems. Because social interactions at work can be a considerable source of stress, a problematic and stressful social climate (e.g. conflicts with boss or quarrels with spouse) can be a risk factor for depression.

-Finally, life events and chronic daily difficulties in combination with vulnerability factors and a lack of social support were also related with depression. This calls to mind the fact that chronic stress can have a more serious influence than do life events (Kessler, Price and Wortman, 1985). Work related stress seems to fall more in the category of chronic stress and has therefore a high potential of triggering depressive patterns.

Summary

To summarise the chapter so far, some possible outcomes of the stressor-strain process were covered. These outcomes could be divided into three broad categories, namely physiological, psychological and behavioural consequences of stress. Physiological outcomes included aversive changes in cardiovascular symptoms, biochemical outcomes and gastrointestinal symptoms. It was shown that stress can result in chronic conditions, and a number of mechanisms in this “wear and tear” process were reviewed. The second category of reactions was made up by behavioural changes. These included the following 5 groups of problems: (1) degradation/disruption of the work role itself, (2) aggressive behaviour at work, (3) flight from job, (4) degradation/disruption of other life roles and (5) self-damaging behaviours. Examples in this category include alcoholism and absenteeism, and can result in very costly consequences. The third main group of stress outcomes was constituted by psychological variables. Stress at work can result in organisational/job outcomes such as job satisfaction; emotional states, such as confusion and frustration; and self-concept variables like depersonalisation and self-esteem. Finally, stress is linked with outcomes of psychopathology such as anxiety, irritability and depression. In the case of depression, an analysis of risk factors was presented that indicated an important role that stress can play in the development of this mental illness.

1.2.4 Societal impact of stress

When considering the outcomes of the stress process, it is clear that stress can give rise to very serious problems in the individual. These outcomes have been reviewed in the previous sections 1.2.1 – 1.2.3). Undesirable changes in the spheres of biology, psychology and behaviour do not exist in a vacuum, however. Not only stressors are located in an environment; so are stress outcomes. Individual outcomes, such as discussed above, have broader societal-economical consequences. Below, the most important ones are reviewed.

Societal costs

In the form of societal costs, stress can have a very destructive impact on the larger framework of society. When people become depressed, for example, they can end up in clinics, or even commit suicide. Also, when people have to quit their job (because they are unable to perform any longer), they are robbed of a very important source of meaning and satisfaction. These people lack the access to many resources and do not enjoy the socialising experience that work can be. For the working population that remains, jobs are becoming increasingly stressed (for reasons for this development, see 1.3.2). Time constraints seriously cut back some very important possibilities to share experiences and activities with relatives and friends. “Jobs”, such as doing the garden, preparing food, or fixing the car are more and more changing from important ties to “simple life” to frustrating obstacles in a tight time-management scheme. The family is one area that especially suffers from these developments. Here, a so-called spill over effect of work stress is possible. In this case, a worker “brings home” the problems he

experiences in his work. When both partners vent out their problems on each other, serious problems for the relationship can result (Dipboye, Smith and Howell, 1994). This process is especially striking in the case of dual career stress. As more and more couples both have ambitious career plans, it is becoming incredibly difficult to harmonise work and private life. When one partner is being promoted to another city, it is not that obvious anymore that the rest of the family will simply join the “breadwinner”. The full impact of the “modern dilemma” is not yet clear, but it could well be that we are facing one of the biggest challenges of the new millennium (Cooper, 1983).

Economic costs

In terms of money, behavioural, physiological and psychological outcomes of stress can become very costly. In the case of behavioural problems, stress can result in drops in sales because of strikes, damaged company property, etc. Moreover, stress-related illness results in costs for sickness leave and medical treatment. When stress results in psychopathological outcomes, such as depression, he will have to rely on a disability pension. Already, 34% of the current influx in the Netherlands is being caused by working stress (LISV (2000), telephonical account). This picture is basically the same for all industrialised countries. Macro- economically it is estimated that in the UK the costs of working stress and mental disorders are 15 billion a year. In the US the total amount of money spent on stress-related disorders is estimated on 100-150 billion US\$, or even 300 billion when costs for sickness leave are included (Matteson and Ivancevich, 1987). This includes the costs for a so-called worker compensation system (see also 2.1), which amounts to 50-60 billion US\$ per year. On the company level, the money that is spent on disability claims can range from 3 to 6% of the payroll, although in some industries, extremes of 40% exist (Schwarz, Watson, Galvin and Lipoff (1989) in Kohler Moran, Wolff and Green, 1995).

1.3 Stressors

1.3.1 Types of stressors

Now that the consequences of stress have been reviewed (1.2), the focus shall shift to the causes of stress problems. When looking again at the definition of Kahn and B., a stressor is defined as follows:

“A stimulus is defined as stressful (i.e., a stressor) because its effect is assumed, hypothesised and (cumulatively) demonstrated to have certain undesirable effects [...]”
(p. 575)

As noted in the definition, a stressor is something that leads to very serious biological, psychological, behavioural and societal outcomes. Unanswered however, remains the question into the origins of stress. Already briefly mentioned as such factors are stressful life events: events with a very far-reaching impact for the person involved. However, recent research has shown that chronic stressors -and workplace stress is often chronic- can have a much higher impact on individual functioning than episodic “event stressors” (Hosman, 1998; Kessler, Price and Wortman, 1985). Work can be a very important facet of people’s life. It can be a major source of satisfaction, social interaction, confirmation and resources. By the same token, however, stressing work can pose a threat to these valuable experiences. Moreover, work can possess many stressful circumstances. Because people spend a high proportion of their lifetime working, these stressors are extremely influential. In the following section, the most important sources of workplace stress shall be presented. They shall be grouped along categories used by Cary Cooper (1983) in her review of stressors at work.

Factors intrinsic to the job

To begin with, there seems to be very large differences in the amount of stress experienced by different professions. It has been found that occupations with a very high workload but relatively low decision latitude are especially troublesome (Gaillard, 1996). Jobs in the helping professions, such as police officers, nurses and social workers, are also considered as extra stressful. One of the reasons lies in the emotional nature of these roles: sometimes the emotional pressure is so high that the worker threatens to become exhausted or “burned out” (Gaillard, 1996). Besides high work load/low control and emotional exhaustion, six other job stressors can be distinguished (Cooper, 1983):

-Poor physical conditions, such as noise, fume and heat (Dipboye et al., 1994). These are mainly found in more traditional blue-collar jobs.

-Shift work. Working on irregular times is more tiring, whereas sleeping irregularly reduces sleep quality. Moreover, there often has to be worked under chronic sleep deficiency (Cooper, 1983), which is physically very demanding. Another common complaint from shift workers is that working and resting at unusual times produces the feeling of being “excluded from society”.

-Job overload. This happens when there is too much work to do, or when the job is too difficult. In any case, the worker feels overtaxed (Gaillard, 1996; Dipboye et al., 1994).

-Job underload, which occurs when the job is too repetitive or boring. Sometimes, as in the case of surveillance personnel, periods of boredom are suddenly disrupted by brief moments of fierce activity. These extreme contrasts are reported to be highly stressing.

-In some occupations, the risk of physical danger is especially high (Cooper, 1983). This is for example the case with fire fighters and police personnel, or with nurses working in a HIV clinic.

-A bad person/environment fit, which occurs when work does not match the personality and expectations of the worker.

Organisational role

A second class of stressors is related to a person’s role in the organisation. Here, the two stressors that have been most frequently studied are role conflict and role ambiguity. Role conflict happens when somebody has two or more roles that conflict with each other (e.g. mother and career woman). Role ambiguity occurs when role demands are unclear or unknown. Another role characteristic that can be sometimes perceived as stressful is responsibility (Cooper, 1983; Dipboye et al., 1994). A possible explanation is that responsibility is often paralleled by risks of failure.

Career development

Career development can serve as a stressor in the following instances: overpromotion, underpromotion, status incongruence (the impossibility of job advancement), lack of job security and thwarted ambition. Some of these career blockages are especially common among women managers, who often face a so called “glass ceiling”, the phenomenon that it is very difficult for women to reach jobs beyond certain levels in the organisational hierarchy, despite a “publicly” communicated message of equality (Brehm and Kassin, 1996).

Relationships at work

These can include problems with supervisors and co-workers. Besides being a stressor in itself, interpersonal problems also lead to ineffective communication, which further increases role ambiguity and other problems (Dipboye et al., 1994).

Organisational climate

An example of a stressful organisational climate is personnel policy (Kahn et Byosiere, 1992). Are job promotions guaranteed or do they have to be “conquered” through fierce internal competition? Similarly, inadequate minority practices can have a very stressful impact for women and people of other colour and religion. Another organisational feature that can produce strain is lack of participation in decision-making (Cooper, 1983; Jackson, 1983). In this case, the worker has too little influence over his job.

1.3.2 Is the presence of stressors on the work flour increasing?

From the previous section, it can be concluded that the work flour can give rise to a great variety of stressors. In recent times, there are indications that the presence of these stressors is increasing. Stress seems to be on the rise and it does not seem that it will disappear from economic life that easily. “Stress is here to stay”, as one author (Murphy, 1996) puts it. As shall be outlined in this section, increases in stressfulness primarily root in macro-economical developments. For a very long time, economic life was marked by the Industrial Revolution, which brought machines and factories. This development signalled an era of centralisation, specialisation and standardisation of capital and labour.

Out of the first little entrepreneurs, huge corporations emerged. In the first phase of this development, there were many competitors fighting for a piece of the market. This resulted in the necessity to produce more effectively and profitably, and stimulated an immense growth of the production powers; not only by scientific innovations, but also by a different deployment of labour forces itself. Replacing handicraft, mass production reached a dominant position, and manufacturing became highly standardised and centralised. The joint result of these developments was that the emphasis came to rest on routine jobs, i.e. jobs that required relatively little initiative or education from the worker. As a result of this, work became increasingly predictable. When could be spoken of stress, it was predominately physical, which is still the case in more underdeveloped regions of the world (De Gier, 1995).

On the verge of the new millennium, western economies changed dramatically. In his article in *Job Stress Interventions* E. de Gier (1995) describes how industry, due to the rise of new information-technologies such as the mobile phone and the Internet, increased its pace. The concept of mass production, which had reigned for almost a century, was no longer competitive on the world market. The higher pace of changes enhanced the need for so-called flexible specialisation, i.e. "a strategy of permanent innovation: accommodation to ceaseless change rather than an effort to control it (De Gier, 1995). While in the past, companies had often tried to spread risk and build up reserves, dynamics and challenge were once again being embraced. These changing demands asked for a "structural adjustment" of western economy, or a reappraisal of free market principles in the form of flexibilisation, deregulation, privatisation and adjustment of economic structures. (De Gier, 1995).

This structural adjustment had tremendous consequences for the quality of work. Following the ongoing flexibilisation process, the number of routine jobs went down in favour of person-to-person and symbolic-analytic services (Reich (1992) in de Gier, 1995) Moreover, a transformation of big companies into smaller networks of core organisations and suppliers took place. In these dynamic corporations, new production concepts arose. At least in some industries, modes like "just-in-time" and "lean" production more and more replaced the dominant position of traditional mass production. Of course this had a revolutionary impact on the labour process. Instead of the traditional assembly line, where every segment of the production process was chronologically done by individual workers, so-called autonomous working groups appeared. These working groups demanded new skills of workers, not only because they combined formerly fragmented tasks, but also while a substantial amount of social and communication skills is needed to perform effectively. All these economic changes had one other consequence, however. Accelerated work pace, increased responsibility and social interaction contributed to heightened unpredictability and thereby to a very serious by-product: stress.

The assertion that stress is increasing is backed up by numerous sources. In an American example, the percentage of stress-related cases from the Northwestern Mutual Life Insurance Company more than doubled from 1982 to 1988 (Murphy, 1996). Other evidence for the increase in stressfulness in the US comes from the skyrocket increases in the amount of money that is paid to workers that sue their companies for stress complaints. From 1985 to 1990, claim costs have more than doubled, placing an increasing burden on many companies (Kohler Moran et al., 1995). Another source are the recent figures from the Netherlands. Here, it is estimated that 34,23% of the new influx comes into the occupational disability pension from psychological grounds. The total amount of occupationally disabled, however, is 31,17% (LISV Amsterdam (2000), telephonical account). Because people that become occupationally disabled on psychological grounds still comprise a larger percentage of the influx than of the total amount of cases, this is indicative of an increase in stress, since a high proportion of these cases is stress related. Figures from the German Federal Republic point in the same direction. Here, the total number of occupational sickness cases dropped with 1.2 percent from 1998 to 1991. In the same period, the percentages of cases due to psychiatric conditions rose with 0.7 percent, however. When the number of sickness day is considered, the contrast becomes even more striking. Whereas the total number of sickness days went down with 2.3 percent, missed days because of a psychiatric disease increased 2.0% (Zoike, 1999). Figures such as these are clearly supportive of the conclusion that stress in the modern economy is indeed on the increase.

1.4 Mediator and moderator variables

The first epidemiological studies into the relationship between stress and psychopathology were focused on so-called life events. Initially, very high correlations with psychopathology were found. However, although the influence of stressful life events is still being estimated on 9%, later research failed to confirm these high correlations. Apparently, the link between stressful life events and illness is not a mono-causal one as researchers had originally assumed. Whether someone develops an illness as a result of a life event turned out to be linked to mediator and moderating variables. And indeed: when social support and coping skills are added as moderating respectively mediating factor, the predictive power of life events increased dramatically (Kessler, Price and Wortman, 1985.) The same is true for working stress: the impact of a stressor cannot be explained without accounting for intervening constructs and variables. In the following section, the most important of these factors shall be elaborated.

Coping

The central role in these intermediate processes is played by the coping construct. A key figure in the development of this construct is Richard Lazarus, who outlined the process of cognitive appraisal, an intra-psychoic process that translates objective events into perceptual experiences. Lazarus (1995) divides cognitive appraisal into two steps. The taxation of the amount of threat in a situation is labelled primary appraisal. The evaluation of one's response possibilities as either adequate or insufficient is termed secondary appraisal. After the appraisal phase, individual will develop some way to deal with the stressor. This way of reacting to these external threats Lazarus calls coping, which can be distinguished along two dimensions: problem focused vs. emotion focused and active vs. passive coping. Combined, these dimensions give rise to the following four categories: **active problem coping** (e.g. trying to alter the stressor), **passive problem coping** (e.g. trying to ignore the problem), **active emotion coping** (e.g. trying to talk with friends about difficulties) and **passive emotion coping** (e.g. taking drugs to "escape" from misery). Certainly, each form of coping can have some adaptive value in certain situations, and how much success a certain way of coping has is dependent on the nature of the problem. Nevertheless, when faced with solvable problems, it is in general thought better to employ active and problem-oriented coping styles.

Which coping mechanism a person chooses is not solely dependent on the amount of success this way of reacting would have, however. The choice of a coping strategy is also determined by the personality and social support of the worker (Gaillard 1996). The influence of a person's social network on his reaction to stress has been the topic of much research. Here, research has conformed that this is a highly powerful determinant (e.g. Kessler, Price and Wortman, 1985; Kahn and Byosiere, 1992; Ogden, 1996). Apparently, having friends or colleagues, who can act as role models and who are available for advice and discussing problems, is very important.

Personality factors

The second factor that influences a person's cope style is formed by personality factors. Although the list of personality factors that have an influence on stress processes is in principle endless, the following ones are widely accepted and have been the topic of considerable research (Kahn and Byosiere, 1992).

Locus of control

This refers to the extent in which a person places the impetus for events within himself or onto external influences. Whether a person is internally or externally oriented thus influences the appraisal process, where people have to judge whether an event is uncontrollable or to be managed.

-Type A personality

People with this personality structure are characterised by excessive competitiveness, impatience, hostility and vigorous speech (Ogden, 1996). Almost by nature, these people will opt for active coping styles. Although such an active style is usually adaptive, there are important exceptions. It is for example not very useful to get overly excited when stuck in a traffic jam: it won't change anything. This is one of the reasons why type-A is an important risk factor for heart problems (Ogden, 1996).

Hardiness (Ogden, 1996)

Individuals with this personality structure share personal characteristics of a) personal feelings of control, b) a desire to accept challenges and c) commitment. Hardiness could be involved in the primary appraisal phase of Lazarus coping model. Here, “hardy” people are more apt to see stressing circumstances as challenges and under their control. This can lead to less stress reactions.

Summary

In sum, the relationship between stressors and strains has been shown to be a complex process, where the presence of stressor A cannot always predict the development of symptom B. Whether this actually happens depends on how people appraise the stressor and how they eventually cope with it. The outcome of the appraisal process and the choice of a coping strategy are in turn dependent on a person’s personality and social support. People with a good social network have been shown to deal with stress better. With regards to personality, there is evidence that the factors of locus of control, type A personality structure and hardiness have an influence on the stressor-strain relationship.

1.5 An integrative model of the stress process

1.5.1 Introduction

As was stated above, the aim of this last section is to integrate the previously covered research into a unifying model. Because stress is a very complex phenomenon, a model is needed that captures this complexity, as well as clarifies the variables and the relationships of the depicted process. The choice of such a model is not an easy task, however. There exist quite a variety of approaches and some of them fit large aspects of the data quite well. However, most of these models seem to neglect important aspects. For example, the so-called Michigan model (Gaillard, 1986) highlights the objective and the subjective environment, as well as outlining the role of personality and social support. However, it is not made very clear by what processes the objective environment leads to a subjective representation. This process is more clearly conceptualised by Lazarus (Gaillard, 1996), who distinguishes between primary and secondary appraisal and different ways of coping. However, in the latter’s model, the role of social support and personality factors are not clearly distinguished.

Because, until now, single models that depict the stress process in all its important aspects are rarely to never found, an integrative approach was used. Such integration has the important advantage that it reduces the amount of “blind spots” that each theoretical model is doomed to have. The model that was chosen for his purpose is Hosman’s IS-model and is presented in the following section. The IS-model integrates numerous other theoretical orientations, but its cornerstone is Lazarus’ model of stress and coping, which emphasises the role of the subjective processing of objective environmental characteristics. As shall be seen below, the model goes beyond this, however. It specifies constructs more clearly, builds in feedback loops and accounts for multiple system levels (Hosman, 1998). To cover the model in this chapter, a description shall first be given. Within this description, links shall be made to the evidence that was reviewed in the previous chapter. This description is best read with the model itself at hand (figure 1.2), since at first sight, it has a fairly complex structure. At the end of the chapter, when the model has been presented and explained, it is hoped that the reader has a clearer understanding of the role of each of the reviewed aspects of the stress process. Also, with a comprehension of the structure and the relationships within the stress process, it shall be easier to understand the rationale behind the study’s hypotheses (which will be presented in chapter 2).

1.5.2 Description of the model

The integrative stress theoretical model describes a stepwise process through which stressors in the environment eventually result in adverse or pathological outcomes. In this model, many of the influences are reciprocal, in that one factor exerts an influence on the other and vice versa. For example, goals and needs have an influence on the perception of a stressor, but this goes the other way around too. Furthermore, the IS-model accounts for multiple system levels. This is depicted by the “onion shells” of the meso and macro environment. Thus, it should be kept in mind that the stress process, with all its phases and elements, is also influenced by a person’s different contexts and vice versa. In the section below, the IS-model shall be explained in detail by reviewing each of its individual steps: (1) appearance

of stressors/compensating factors, (2) perception, interpretation and valuation of the stressor, (3) problem experience (occurrence of crisis), (4) anticipation of problem reaction, (5) performance of problem reaction and (6) consequences of this reaction (feedback loop).

Appearance of a stressors and protecting factors

The process depicted in the model begins with the existence of certain stressors in the environment. Indeed, this is the same basic notion as in the general "stressor-strain" concept that was reviewed previously. The most important kinds of stressors, such as understimulation and threat of unemployment, were already covered in section 1.3. The influence of stressors, however, is buffered by the availability of protective factors. These are resources in the environment that are beneficial to people's health and functioning. On this topic, not much research has yet appeared. For this reason, only an example of a protective factor can be listed, without wanting to claim that there are no others. This example is physical work, which was mentioned by Kahn and Byosiere (1992). The two factors of risk and protective factors together constitute the environmental input of the IS-model. After the occurrence of these risk and protective factors, the model continues with the subject that reacts to these influences.

Perception of the environment

The core process of this step is that a person perceives the environmental factors from step 1 and judges whether they form a threat to important resources. This process is basically similar to Lazarus'

IS MODEL

conception of primary appraisal that was covered in the previous section (1.4). In the model, it is dependent on a number of factors whether someone perceives his environment as threatening or not. Firstly, this phase is influenced by a person's goals and needs. Thus, an event such as a major company reorganisation will be perceived as stressful because it threatens to take away highly valued resources of money, prestige and satisfaction. This process is reciprocal, because the threat of losing important resources also triggers the goals and needs that are at stake, making them more salient (Brehm and Kassin, 1996). Secondly, the amount of threat an individual perceives will depend on his appraisal of how difficult he expects it to be to overcome the environmental threat. This will vary according to his problem solving skills because if the person possesses a very powerful and broad range of coping skills, fewer situations will be perceived as stressful. This description is very similar to Lazarus' secondary appraisal construct, in which a person judges to what degree his coping resources are sufficient to solve a problem. Again, this process goes the other way around too because perception has an influence on a person's perceived coping skills. For example, if the situation is highly dangerous and overwhelming, the person might lose his confidence in his available coping resources. A third moderator of the appraisal and coping process is the expected social support. If this is available, stressors become less threatening because the social system can help a person cope with it. Fourthly and finally, the process of perception is influenced by the subsequent phase of crisis perception. If the individual already sees the situation as highly threatening, that will lead to a focusing on the stressors and other kinds of perceptual biases. This will further strengthen the appraisal of the environment as stressing.

Crisis experience

In the case of a sufficiently serious stressor, the result of the previous step is an attribution of the situation as dangerous and a threat to important resources such as money. This leads to an emotional and physical "alarm" (Hosman, 1998) that arouses the person to invest great energy in solving the problem. Such an alarm state bears close resemblance to Cannon's "fight/flight" concept, which was introduced at the beginning of the first chapter. Also, this agitated condition calls to mind Selye's general adaptation syndrome, which included arousal symptoms such as an enlargement of the adrenal (arousal hormone) gland cortex. In case of a crisis, this emotional reaction can be so fierce that the problem overwhelms the person. This process is influenced by a number of factors. The influences of this crisis phase follow more or less the same lines as the previous step, as crisis experience is influenced by problem solving skills and social support. Furthermore, there exist a two-way relationship with the preparatory reaction. If the person prepares a reaction of which he anticipates an insufficient result, the experience of crisis will only be sharpened.

Preparatory reaction

In this process, the individual selects and anticipates a way of responding to the environmental threat. In other words, the person prepares the way by which he is going to cope with the problem. Here, he can choose between different coping types. From different possibilities, the person selects the way he is going to react to the experienced crisis. This resulting process is influenced by the problem solving skills of the person. There might exist a preference for certain styles, or some forms might be more sophisticated and trained. This process is reciprocal, however, because preparing and using a coping skill will also affect competency with this skill. The preparatory reaction is further influenced by the available social support because certain modes of coping can only be successfully deployed if a stimulating environment supports them.

Stress response

In this phase, the person performs the selected response from step 4. As stated above, there are many different ways with which a person can cope with a stressful situation. Coping is essentially an effort to adapt to the stressing environment. The result of this adaptation determines whether an individual succeeds in preventing future problems or becomes involved in a feedback loop of distress and adversity (see below). The exact success of the way a person copes with his environment depends on the particular characteristics of the situation. There are, however, a number of factors that play an important role. In the model, the way this response is carried out is influenced by the person's social support. People can attract advice and resources from their social environment, which will make efforts to cope more effective. Also, the eventual effect of a response depends on how competent the person is in performing this response. This is an important reason why many stress interventions include some kind of (coping) skills training.

Consequences of the reaction

In its final phase, the IS-model contains a feedback mechanism that loops back the results of the stress response to the person himself and the other factors in the model (depicted by the block arrows). In this chapter (1.2), many types of such consequences have been reviewed. One of these categories is constituted by physiological health problems, which can take the form of cardiovascular symptoms, biochemical outcomes and gastrointestinal problems. Another type of outcomes, which is of particular interest to the current writing, is mental health problems. When a person fails to cope with a threatening stressor, psychological illness such as depression and anxiety might be the result. Finally, besides accounting for such individual consequences, the model feeds back to the environment as well. Effects can include financial costs to companies and governments and social costs to families and society as a whole (see also 1.2.4). In the figure, these outcomes are specified by special boxes with the header "CONSEQUENCES". These boxes should be regarded as illustrative and are not meant as specific elements or phases of the model.

Other factors that receive feedback from the model are goals and needs, problem solving skills, the social support system and the existing risk and protective factors. For example, the stress response of alcoholism can alienate a person from his environment so that his social support system weakens. Furthermore, his coping skills will also be reduced because alcoholism makes the use of certain ways of coping (such as active problem solving coping) less probable. A coping response can also result in actually altering the stressing environment. Through coping, a person can decrease stressors and increase the presence of protecting factors. A person might, say, complain about certain harmful aspects of his working environment. When successful, such actions lead to a reduction in the amount of stress.

Chapter 2: Stress Interventions

2.1 Introduction

In the previous chapter, it was pointed out that stress can lead to numerous adverse consequences that manifest themselves on the level of physiology, psychology and behaviour. Moreover, these individual outcomes affect broader system levels such as society, family life and economy. As noted in the introduction, there are a lot of disciplines that are confronted with these stress-associated problems. They are also faced with the problem of reducing these problems. "Doing something about stress" is not as simple as it seems, though. What exactly needs to be done? On first thought, the answer seems simple: when someone is experiencing high blood pressure as a result of stress, try to lower the blood pressure (e.g. by means of anti-hypertensive medication). This strategy evaluates a problem in terms of its symptoms and sees its solution in reducing these symptoms. An important argument for this approach is that often the underlying causes of a problem are not known, so reducing symptoms is the only possibility. Moreover, some underlying causes, such as genetic dispositions, are simply not (yet) amenable to change. Finally, in some instances where they could be changed, this is not wanted because underlying mechanisms serve highly valued other purposes. It might for example be very stress reducing to abolish the use of deadlines in a newspaper agency, but when this would result in a reduction of quality, it is not likely to be implemented.

Besides this symptom approach, there is another possibility. This other approach tends to see symptoms as a part of a deeper underlying problem. In this view, symptoms are caused by pathogenic factors, so in order to reduce the symptoms one should eliminate these factors. Although the value of the arguments for the symptom-oriented side are recognised, this essay starts from the second perspective. It does so for a number of reasons. Firstly, focusing on symptoms cannot completely eliminate a disease. Because the underlying causes are still present, the curing of symptoms has to be repeated over and over again, whereas taking away the causes would solve the problem once and for all. This is the more attractive because, as noted in chapter 1, curative care is socially and economically very costly. A second reason for the use of preventive interventions is that it is unethical to let people suffer the symptoms of a condition when intervening in its causes could have prevented this. Thirdly, a symptom-oriented perspective focuses mainly on the individual. This ignores the fact that many conditions (and not in the last place stress) have their roots in socio-economic circumstances. These factors are out of the scope of curative interventions, but are possibly changed by preventive efforts.

Besides the above-described arguments for the preventionist position, there exist legal incentives too. In Europe, the European Union has implemented laws that have highlighted the need for healthier work environments. In 1989, the Union issued the "framework directive on health and safety at work", which became to resemble a kind of European work environment act. It included quite specific employer directives regarding the quality of the job content, well being of workers at work, medical examination, training, information and worker participation. The consequence of these legal developments was that many member states implemented penalties to make sure that companies would guarantee a healthy working environment in line with the EU directives. In the Netherlands, for example, the so-called "Pemba law" obliges enterprises to directly pay for each worker that becomes occupationally disabled. In the US, too, industries are facing financial responsibilities for their occupationally disabled, although the legal situation there is markedly different. Through a workers compensation system, employees have the option to sue their employers for missed income due to occupational disability (Kohler Moran et al., 1995). Compared to a Social Security pension, very high sums can be the result. In both the US and the EU, therefore, the legal situation causes a shift in focus, so that companies are becoming more and more responsible for their stress problems.

2.2 Characteristics of prevention programs

In categorising a preventive intervention, a number of distinctions can be made. Following a distinction by Hosman (1998), the following 8 dimensions can be distinguished: (1) type of actor, (2) target group, (3) goals, (4) determinants, (5) instruments, (6) system level, (7) time and (8) setting. These dimensions are not only important in describing and categorising different interventions but also in answering the

question of which aspects of an intervention are crucial for its effectiveness. For this latter purpose, an analysis is needed that links program characteristics to effect. To prepare this, however, such characteristics shall first have to be coded using a classification system in which many of the here described distinctions are reflected. In this chapter, each of the 8 research dimensions shall be elaborated in more detail:

2.2.1 Actor/program provider

The dimension of actors pertains to the people that provide the stress intervention. Because stress can be a very costly economic factor, it are often organisations that initiate a stress intervention. Sometimes, the manager of the human resources department notices problems and decides to take action (e.g. in the study of Malkinson, Kashmir and Weinberg, 1997). However, besides the Personnel Development office, it are often (mental) health professionals or external consultants that provide the intervention.

2.2.2 Target group

This pertains to the question on what kind of group the intervention shall focus. Regarding this factor, a number of aspects are important. One distinction, which also returns in the coding system, is the difference between ultimate and intermediate target group. Here, the group in which the intervention is carried out in order to prevent future problems is called the ultimate target group. In some cases, however, programs use a so-called intermediate target group. In this case, the program uses an additional group that has an important influence on the "problem group". In the case of organisations, for example, stress is often caused by organisational constraints, such as unsupportive supervisors. Because the problem group might not be able to alter these constraints, interventions sometimes target other groups that do have this possibility. In the example of the unsupportive supervisors, they could be given a sensitivity training. Here, the workers that are stressed are the ultimate target group, whereas the unsupportive supervisors constitute the intermediate target group.

Another important factor is whether the target group is characterised by a high amount of risk. It is quite common that intervention programs are targeted at high-risk groups. In the present coding system, this was the case when the target group possesses a lot of risk factors³, or when there are already many stress related cases⁴. When the program was targeted at a population for which no higher risk was known or stated, this was coded as normal risk⁵ (see appendix 3 for an operationalisation of this variable). The question of high-risk is important because this affects some very important strategic decisions. Budgets for prevention are limited, so it would be very valuable information whether it is more efficient to spend the efforts on a smaller group of higher risk people or to spread the intervention over a larger population. A hypothesis that shall be researched is that **programs aimed at high-risk groups are more effective than programs that do not**. There are a number of reasons why this could be the case. Firstly, people that are at high-risk for a problem could well be more motivated to participate in an intervention. Many authors, such as Bunce (1997), suggest that programs are more effective when they have more motivated participants. Another reason why high-risk programs could be more effective is related to the so-called span of change. People that are at risk for a problem will often be characterised by a greater number of risk factors. According to the span of change concept, such people therefore have more room for improvement.

2.2.3 Types of goals

Of course, the nature of an intervention is very dependent of its target. In this dimension, different types of goals can be distinguished. So-called *distal goals* are the end targets for the preventionist. In the case of stress interventions, this would most often be stress, although related conditions like anxiety and depression can also be addressed. *Proximal goals* are intermediate goals that have to be accomplished

³ Selective preventive intervention.

⁴ Indicative preventive intervention.

⁵ Universal preventive intervention.

in order to reach the distal goal. For example, increasing a worker's social support system can be a proximal goal in order to reduce stress (distal goal). A very important other class of goals are *social end goals*. These goals are the most important aims for other involved parties in the social network of the participants. An example of this kind is reduced absenteeism. This is a social end goal because for most companies, this would be the eventual focus, where stress is just a *proximal goal* to achieve this.

Another goal distinction is the one between primary, secondary and tertiary prevention (Houtman and Kompier, 1995; Murphy (1988), in Cooper and Sadri, 1995). In *primary prevention*, the goal is to prevent the incidence⁶ of a problem, whereas in *secondary prevention* the aim is to intervene as quickly as possible once a problem has started to manifest itself. Alternatively, secondary prevention can aim at the prevention of new episodes, which is called relapse prevention. Finally, *tertiary prevention* is aimed at mitigating the side effects of a condition. An example from another context would be the integration of schizophrenics into society. Here, social exclusion is the "side effect" of the disease. Social integration would prevent this from happening.

A third goal distinction that can be made is the discrimination between psychological, behavioural and physiological goals. In the current essay, this distinction is important because a hypothesis is that **programs are only effective on those variables that are closely linked to their method**. Support for such a prediction comes from Murphy's 1996 review on worker-oriented interventions, in which he found very outcome specific effects of the different methods (see the discussion for a closer look at Murphy's results). The rationale behind this requires some elaboration, because most stress models (such as the IS-model) do not at first sight predict this to happen. In theory, stress is a process that is inherently linked with behavioural, psychological and physiological outcomes. As was seen in chapter 1, this assumption is backed up by numerous empirical sources. Theoretically, an intervention that would intervene in the stress process would thus result in changes on all outcome levels. However, different methods can place different accents. They concentrate on certain processes that are thought to be essential in the working mechanisms of their methodology, and these processes are more closely monitored and guided than other ones. This focussed attention is thought to result in differential effects. For example, biofeedback is an approach that is distinguished by an accent on physiological changes. Therefore, the effect of biofeedback is predicted to be mainly physiological as well.

It is quite difficult to make such exact predictions for all of the covered methods. Although some methods are easily distinguished as either psychological or physiological, others, such as exercise and meditation, are more difficult to assess. Exercise, for example, involves physiological changes (such as transpiration, elevated heart rate, etc.) but also invokes an altered psychological state (some kind of a "feeling good" experience). Therefore, it is difficult to predict where the most effect will occur. The same is true for meditation, which involves both physiological and mental calmness. For this reason, there could not be specific hypotheses in the sense of: "method A will lead to outcome X". Rather, the hypothesis simple states that there is a differential effect, leaving the question regarding the precise direction and nature of this effect to future research.

2.2.4 Risk and protective factors

These are the factors that the program addresses in order to manipulate its goals. When remembering the IS-model on stress, several of such possible intervention factors appear. For example, the way a person copes with his surroundings or his perception of these circumstances can be addressed. In the coding system, this dimension shall be covered under the categorisation "risk and protective factors addressed". In the essay, no specific hypothesis on risk and protective factors was formed. The working factors behind an intervention are often unspecified or hypothetical and are therefore difficult to code.

⁶ Number of new cases

2.2.5 Instruments

Because a description of methods and techniques would take too much space to be discussed here, this topic will be dealt with separately in section 2.3.

2.2.6 System level

Regarding the level that is addressed by the program there are 3 possibilities: (1) the micro, or individual, level, (2) the meso, or organisational/group level, and (3) the societal macro level, which encompasses the whole of society. The distinction is highly relevant for the stress intervention field. Here, many researchers have divided stress prevention programs into worker-oriented interventions, which target the stressed employee (micro level,) and system- or organisation-oriented interventions (meso level), which focus on the whole organisation (e.g. Houtman and Kompier, 1995). The system level dimension shall also be coded in the classification system. Regarding this variables, two hypotheses shall be examined. As was seen in the IS-model, both the individual and the environmental context exert a broad influence on the stress process. However, only organisational interventions can affect the very roots of the stress process: the stressor. On the other side, there has been less research and experience with these kinds of intervention. Also, there does not yet exist such a clearly defined methodology as with individual interventions. Thus, because there are pros and cons for both approaches, it shall not be predicted that one is better than the other. Rather, the hypothesis is that **organisational methods are just as effective as individual-oriented ones**. This would be in line with numerous authors (e.g. Daniels, 1996), who state that the two approaches are equally important in reducing stress.

The second hypothesis regarding system level has to do with the relation between short and long-term effects. As can be seen in the IS-model, the stress process is influenced by a feedback loop that creates the possibility for long-term effects. Because the stress processes begins with the occurrence of a stressor, organisational interventions seem to be the only possible way to directly stop this development. Because the stressor is taken away in organisational interventions but stays present in individual ones, it is hypothesised that the effect of individual programs wears out, whereas organisational efforts continue to be effective. Therefore, the hypothesis that shall be researched is that, **at follow-up, organisational interventions shall be more effective relative to individual ones than at post-test**.

2.2.7 Time aspects

The factor time shall be coded in the classification system in the following ways. First, the program duration, or the time (number of weeks) it is implemented, shall be specified. Moreover, the length of the intervention in number of sessions and the distribution of these sessions shall be coded. For example, a program could be run 8 times with a distribution of one session per week. Of possible importance is also the availability of so-called *booster sessions* and *fading systems*. Programs with a fading system gradually reduce the frequency of sessions until this reaches zero or almost zero. Booster sessions are a limited amount of extra meetings that are provided after the original program has ended. For example, after a stress intervention that has been run for 2 months 1 time a week, participants would come together once per year to check results and "refresh" the knowledge that was learned during the program. Despite their potential usefulness, the presence of booster sessions and fading systems seems to be a rarity in the field of stress interventions (in fact, no study that was covered in the essay reported having either one of them).

Besides the distribution of the sessions and the presence of fading systems and booster sessions, another distinction is of importance. This is the question of the time relative to working hours that the intervention is being carried out. Here, there are two options. Firstly, the intervention can take place within working hours. Secondly, sessions can be located at some time point after the ending of the working day. This has potential implications for program effectiveness. On the one hand, there is the possibility that interventions being carried out within working hours are more effective. The rationale could be that companies that provide their employees with interventions within working time communicate the message that the worker is a valued and recognised asset of the organisation. The reverse could also be the case, however. One possible reason is provided by Kolbell (1995), who states that interventions during working hours might actually interfere with the learning process. Employees would see such interventions as yet another work-related demand that interferes with other assignments and tasks. Also,

research in the learning theory tradition has found that learning and stress levels relate to each other according to an “inverted U” function. This principle, which is known as the Yerkes-Dodson law (Lieberman, 1999), implies that learning takes place best under moderate stress levels. Thus, stress can facilitate learning, until it reaches such levels that further increases interfere with the process (hence the “inverted U”). Thus, when workers are already too stressed by their daily work and are given some workshop or stress management course, they might learn under unfavourable stress levels.

The essay will try to address the question of whether running an intervention after or within working hours is more effective by including this distinction as one of the variables in the analysis. **The hypothesis that shall be tested is that interventions after working hours are more effective.** Because the essay sees stress interventions (with the exception of, maybe, exercise) essentially as learning processes, it is assumed that learning principles such as the inverted U function do also apply and make interventions within working hours less effective. Since the distinction after/within working hours is not very relevant to organisational interventions, this prediction shall only be researched for individually focused programs.

The second hypothesis for the effect of time on effectiveness is that **there is a positive relationship between duration of individual-oriented programs and effectiveness.** Such a dose-effect relationship would resemble findings from time-limited psychotherapy research, where such positive results have often been found (e.g. Barkham et al., 1996). Carlson and Hoyle (1993) provide a reason for this prediction by emphasising the “skill frame of reference” of stress interventions. As noted above, (individual) stress programs are essentially a learning process focussed at new skills. Because the learning of a skill requires its practice and application (which would take a certain amount of time), it is expected that an intervention becomes more effective the longer it is carried out.

2.2.8 Setting of the intervention

“Setting” pertains to the question where an intervention is carried out. In the case of stress interventions, the setting is often the company itself, which provides room for meetings and training sessions. Also, it is possible that training is carried out on other locations, such as training centres or universities. This is often the case when participants come from different organisations, which makes it very difficult to run the intervention “on site”. Like the “working hours” distinction, it is hypothesised that programs that are run on site differ from off site interventions in effectiveness. In this case, **the hypothesis is that (individual-oriented) programs carried out on site shall be more effective.** In the reasoning behind this prediction, the concept of transfer of learning is important. As reviewed by Lieberman (1999), transfer from the context of learning to the context where the learned has to be performed occurs best when these two contexts are more similar to each other. Thus, interventions on the job would have the perfect setting because the context of learning and the context of performance are exactly the same. Following this rationale, they are expected to be more effective.

2.3 Methods for stress reduction

There exist many different methods to achieve reductions in stress. Many of them were not especially developed for this purpose but were “borrowed” from related field such as clinical psychology (Murphy 1996). In this section, a distinction across methods shall be presented. In this distinction, they shall be grouped according to a focus on either the individual worker or on his environment. For the worker-oriented category, there have already been a number of categorisation attempts. This has resulted in a number of widely accepted categories (e.g. biofeedback). In the environment-oriented category, there does not exist such a large quantity of theoretical work. Studies examining these strategies are very rarely found (Cooper and Sadri, 1995). Therefore, it is more difficult to present a categorisation of clearly outlined methods. For these types of programmes, therefore, a distinction according to program’s targets was made.

2.3.1 Individual methods

Muscle relaxation

The first individual method is muscle relaxation, which is also regarded as a separate category by Murphy (1996). The technique involves focusing one's attention on muscle groups in order to relax them. It is learned to identify even small amounts of physical tension and then to release this. Often, to practice these skills, muscle groups are first being clenched and then relaxed. A special variation of this technique is so-called progressive muscle relaxation (PMR). Here, participants successively work through the various muscle groups of the body, thus cleaning it from stress and tension.

Biofeedback

The second technique of biofeedback is widely accepted as a distinct method (e.g. Van der Hek and Plomp, 1997; Murphy, 1996; Matteson and Ivancevich, 1987; Dipboye et al., 1994). It is based on feeding back information from the body to the person involved. This information comes from machines that are attached to the person and that give indications of the state of different psychophysiological functions, such as heart rate. By providing such information, biofeedback training is claimed to result in being able to control bodily process, including those that play an important role in the experience of stress.

Exercise

Exercise involves regular fitness or aerobic activities⁷. The intention is to enhance the physical condition of the participants, because healthy and fit individuals have a lower baseline arousal and fewer excitation fluctuations than non-fit individuals (Dipboye et al., 1994). Because of this, they are expected to be more resistant to stress. Other authors who treat exercise as a separate method are Matteson and Ivancevich (1987) and Dipboye et al. (1994).

Meditation

The fourth technique of meditation is based on mental calmness. While maintaining a passive attitude towards intruding thoughts, one focuses on some calming word or sound (also called mantra). During this, people are being instructed to breathe as regularly and calmly as possible. This is claimed to evoke a "relaxation response" and, after practice, the person is supposed to be able to evoke it at will (Murphy, 1996). Although many authors have combined muscle relaxation and meditation into a broader "relaxation" category (e.g. Matteson and Ivancevich, 1987; Van der Hek and Plomp, 1997; Dipboye et al., 1994), this essay follows Murphy, who, in his 1996 review, treated them as different methodologies. Support for this came from his results, which showed that muscle relaxation had resulted in different effects than meditation (see also the discussion).

Cognitive-behavioural training

Cognitive-behavioural skill training involves techniques that aim to modify (1) the way we think about stressful situations (i.e. appraisal) and ourselves (cognitive restructuring) and (2) the skills that people use to manage these stressors (Murphy, 1996; Van der Hek and Plomp, 1997). An example that places more emphasis on the first aspect is Ellis' Rational Emotive Therapy (RET). A widely used variant that is more directed at skill training is Meichenbaum's stress inoculation training (SIT), which involves the three stages of (1) education, (2) learning new coping skills and (3) putting them into practice. Although the two aspects of cognitive restructuring and skill training may appear very distinct, many authors regard them as belonging into the same category (e.g. Murphy 1996; Matteson and Ivancevich, 1987; Dipboye et al, 1994; Van der Hek and Plomp, 1997). Indeed, because they both deal with the way people cope with a stressor (both mentally and behaviourally), they are often found together.

Other methods

In the present sample of studies, the "other category" included psycho education, massage, classroom management training, cultural training, human relations training and hypnotic techniques. Because of their rare use, they are not coded separately. In fact, in the present sample of studies, none of these methods was more than one time encountered. With regard to the effectiveness of this rest category, it is very difficult to make predictions. Murphy found his "other" category to be consistently effective on all type of stress outcomes except for physiological measures. For his category of organisational/job outcomes, it even was the most effective approach. However, because this category is very

⁷ Aerobic exercise is defined as any physical activity that produces an elevated respiration, heart rate and metabolic rate for 20-30 minutes (Dipboye et al., 1994).

heterogeneous, it might well be possible that the current analysis results in different conclusions. Therefore, no specific hypothesis was stated.

Combinations of methods

Of all the studies reviewed by Murphy in his 1996 article, combinations were the most frequently used interventions. This resembles the trend that many present studies tend to be more eclectic in their methods and do not focus on just one or two variables anymore. Of the 39 individual-oriented studies in the present analysis, 15 (38,5%) used at least two different methods. In the essay, **the hypothesis is that programs that use multiple methods are more effective**. Again, the concept of span of change is helpful in understanding the rationale. Every method uses a certain set of techniques to accomplish a specified goal. To take effect, some determinants of change (see the section on risk and protective factors) have to be manipulated. Because, usually, different methods act on different determinants of change, the use of multiple methods would enlarge the possible span of change.

2.3.2 Organisational methods

As many authors state, when compared to worker-oriented interventions, there is a lack of efforts aimed at the environment (Daniels, 1996). This is reflected in the sample of the present study, where only 11 of 47 programs were directed at the environment of the worker. Moreover, the different methods used by such environmental approaches are less sharply outlined. Because of this, they shall not be grouped according to the method that they use but to the goals at which they are aiming. Of the 11 organisation-oriented studies in the present analysis, 3 could be coded as oriented at the social support network, 3 at the enhancement of control and 5 at the reduction of stressors on the job. This division in stressor reduction, social support and control is supported by the structure of the so-called Karasek model of the work environment (Gaillard, 1996). In this model, working conditions can be described along three dimensions, namely work load (mental demands), decision latitude (control) and social support from colleagues and supervisors. Below, a short description of these different dimensions shall be given.

Programs aimed at social support

These kinds of intervention seek to strengthen a worker's network of social support⁸. That social support is able to buffer stress is a common finding across stress research (e.g. Ogden, 1996; Kahn and Byosiere, 1992). In the IS-model (chapter 1), social support therefore has a very important moderating role on the stress process. Social support can work in a variety of ways. In terms of benefits, Ogden (1996) outlines the following resources that people can acquire: self-esteem, information, social companionship and instrumental support (physical help). Although social support can have many roots, a very important aspect is support within the job/organisation (e.g. from colleagues and co-workers). Programs that target this variable do so mostly by setting up support groups. In these cases, groups are set up where workers can talk about their jobs and provide each other with help and advice.

Programs aimed at control

These interventions aim to increase the amount of control, or decision latitude, that a worker has in organising his job. Such interventions are based on findings that control is an important moderator of the stress process. For example, studies on both humans and animals have been shown that control is a good predictor of all kinds of health outcomes (Gaillard, 1996). Moreover, according to Ogden (1996), control has an influence on the subjective experience of stress. Finally, people that show a greater amount of control engage more in active coping styles, thereby changing the environment and preventing ill health (e.g. Ogden, 1996). Programs aimed at increasing control can achieve this in a number of ways. Some programs arrange extra staff meetings, so that personnel are provided with a greater amount of say regarding important decisions.

Programs aimed at stressor reduction

This method involves direct changes in environmental stressors. The rationale behind these kinds of interventions is relatively simple. Because stressors are at the basis of the stress process, reductions in stressors lead to a reduction in stress, regardless of mediating and moderating variables. One approach

⁸ There does not yet exist an universally-accepted definition of social support. It seems clear however that aspects such as the number of social contacts and their perceived helpfulness, intimacy and availability play an important role (Brehm and Kassin, 1996).

to achieve this is job redesign, which is aimed at changing the characteristics of the job. This would for example lead to inserting extra pauses in very demanding working hours. Another example is socio-technical change, where the focus is on changing the organisational system.

Summary

To briefly summarise this section, a number of different dimensions of prevention programs were covered. Of all these dimensions, numerous aspects were reviewed. For example, a program can be aimed at a high-risk target group, which might influence eventual effect. Regarding the instruments that are used to bring about this effect, numerous methods were outlined. These can be grouped according to their focus on either the worker or the organisational environment. In the worker-oriented category, there exist the methods of muscle relaxation, biofeedback, exercise, meditation and cognitive-behavioural training. Also, some programs employ relatively rare methods such as massage therapy. Finally, in many cases, a combination of the diverse methods is used. Studies that target the environment for stress reduction are less often found than worker-oriented interventions, in a ratio of approximately 1/5. Also, it is more difficult to group them according to methods. However, these programs can be grouped according to the environmental factor they try to manipulate. Here, some programs are aimed at increasing social support, some at enhancing control and participation, and others at directly reducing the stressors themselves.

2.4 Implementing the program

When a program has been constructed and is ready for implementation, it faces a context where it is to be set up. In these surroundings, it is confronted with a specific target group, whose characteristics exert an influence on the intervention process. Examples of such characteristics come from some of the analysed studies in which the intervention company was suddenly struck by a major organisational change. For example, Pavett, Butler, Marcinik and Hodgdon (1987) studied a stress intervention on a navy ship. During the course of the intervention, the ship failed a training test, which resulted in extra time on sea. The authors reported that this influenced performance and training results. In a case such as this, it is very difficult to assess the value of the program. All the amount of extra stress might completely overrule small stress reductions achieved by the program. Thus, the characteristics of the organisation are very important determinants of success. Another example of this is the importance of having managerial and organisational support for the program (Kompier, Geurts, Gründemann, Vink, and Smulders, 1998; Houtman en Kompier, 1995) because such support prevents all kinds of administrative barriers and the withdrawal of resources.

With regard to the characteristics of the target group that influence program effectiveness, a number of important factors can be outlined. In the coding sheet, features such as age, education, income, sex, immigration status and family structure are listed. These factors are thought to play a role in the implementation of the programme. For highly educated personnel, a cognitive intervention might result in different outcomes than for lower educated workers. Sometimes, program providers recognise this principle and tailor their interventions to the specific needs of their target group. Because, in these cases, such adjustments are integrated in the program itself, they were coded as program characteristics. Sometimes, however, a program encounters attitudes and opinions that have not been integrated in the program beforehand. In these cases, factors such as the motivation of the target group have to be considered as context characteristics. As stated before, motivated workers are more likely to become highly involved in the program, which, according to Murphy (1996), affects eventual effect. Such an expectation is also supported by Bunce, who in his 1997 article points to the importance of so-called process variables (factors such as the satisfaction of the participants) in predicting effect.

Chapter 3: Method

3.1 Introduction: Why perform a meta-analysis?

Thus far, overviews of the stress process (chapter 1) and preventive programs (chapter 2) have been given. From the evidence that was discussed there, multiple hypotheses were generated. As noted in the introduction, there are numerous ways to do address such questions. Roughly, three distinct methodologies exist: (1) the traditional experiment, (2) the narrative literature review or (3) the meta-analysis. To begin with the first option, the study could address its research questions by conducting an experiment. The core of this method consists of manipulating some independent variable⁹ and observing its effect on the dependent variable¹⁰. While holding constant all other variables (with the use of an equivalent control group), the impact of an intervention can thus be determined. At least, that is the theory. Unfortunately (at least for the scientist...), human behaviour is difficult to manipulate, common definitions are unavailable, and methods, techniques and sample characteristics vary over studies (Wolf, 1986). Therefore, in the social sciences, “single experiments” only rarely provide definite answers to research questions.

An alternative for the experiment is provided by the so-called narrative literature review. Here, a scholar collects a representative sample of the literature on a certain topic and then gathers an impression of the current state of the art. This approach is called “narrative” because it is based on the subjective expert decision of the researcher instead of on statistical analysis. Compared with the single experiment, the literature review has the important advantage that it is based on more than one study and is therefore more reliable. However, as with the previous approach, there are also important difficulties. As Wolf (1986) describes, narrative literature reviews are “notorious for depending on subjective judgements, preferences, and biases of the reviewers (p. 10).” Therefore, conflicting interpretations of the literature among different reviewers are not uncommon. For this reason, it is not altogether undeserved that these narrative approaches are often criticised for their unreliability and lack of scientific rigour.

Because the individual experiment and the traditional literature review have so much possible difficulties, the present study looked for an alternative approach. This alternative was found in the third of the above-described alternatives: the meta-analysis. Such a meta-analysis allows for a quantitative classification, integration and analysis of empirical studies and their results (Wolf, 1986). Moreover, it overcomes much of the problems that are faced by the previously covered techniques. Together with the narrative literature review, it has the advantage over the single experiment that an integrative conclusion over a greater number of studies is possible. Contrasted with the narrative literature review, however, the meta-analysis is more objective and less liable to bias. Moreover, because the meta-analysis is a quantitative technique, it makes possible two other things. Firstly, a meta-analysis is capable of grouping patterns of effect size in order to better examine inconsistencies across studies. Such inconsistencies in effect size can be statistically identified with the use of a so-called homogeneity test (see below). Secondly, it is possible to statistically examine characteristics of studies as potential determinants of effect (Wolf, 1986). Because the study is interested in general patterns across interventions and identifying determinants of effect, it was chosen to perform a meta-analysis.

3.2 Description of the analysis

In the section above, the choice for the meta-analysis technique was explained. In the rest of this chapter, the consecutive steps of this method shall be described and elaborated. These are the following five phases: (1) the formulation of a research question, (2) a literature search and selection of studies, (3) the coding of these studies, (4) the calculation of effect sizes and (5) the reviewing and interpretation of these results (Wolf, 1986). In the sections below, each of these 5 steps shall be outlined in more detail.

⁹ The variable that is altered.

¹⁰ The measure the researcher is interested in. In case of an intervention effect, the dependent variable changes as a result of the experimental manipulation.

3.2.1 Research questions

As stated above, the first step in performing a meta-analysis is the formulation of the research question(s) in which the purpose and goals of the research project are articulated. For the present study, three interrelated objectives shall be addressed. These three research questions are: (A) estimating a general effect size for stress interventions, (B) analysing these results for outliers and patterns of inconsistencies and (C) addressing the effect hypotheses in order to identify possible effect predictors. Below, each of these questions shall be described in more detail.

Research question A: magnitude of effect size

As stated above, the first research question addresses the magnitude of the average effect size across studies. For this purpose, it is first necessary to obtain a standardised measure of effectiveness. However, study evaluations are often based on conceptually different outcome measures (e.g. depression versus blood pressure), so that it is not always clear which of these measures can be used in assessing the global effect. Here, a meta-analysis is faced with the so-called “apples and oranges problem”, or the problem that it is impossible to compare fundamentally different outcome types in the same analysis (Wolf, 1986). In order to address this problem, a number of distinct outcome categories were created and separately analysed (as recommended by Wolf, 1986). As such, the present study discriminated between the following 5 outcome groups: (1) psychological outcomes, (2) physiological outcomes, (3) behavioural outcomes, (4) job/organisational satisfaction and (5) cognitive stress coping. Below, these different outcome types are described in more detail.

-Psychological outcomes included all measures that are related to mental health (as described in chapter 1). Mental health was defined here as symptoms (e.g. feelings of stress or frustration) or syndromes (e.g. depression or anxiety) of “internalised”¹¹ distress.

-Physiological outcomes were conceptualised as stress-related consequences of physiological and biological nature. These included both psychophysiological¹² indicators (e.g. blood pressure) and more long-term health consequences (e.g. illness, sleep patterns, etc.). Because the study was primarily interested in distal outcomes, proximal goals such as fitness and muscle strength were excluded.

-Behavioural outcomes were thought of as every factor that was listed in chapter 1.2.2. However, only those behaviours for which there existed a clearly desired direction (i.e. behaviours which could be either positively or negatively valued) were included.

-Job/organisational satisfaction included everything related to positive or negative feelings towards the organisation or job (satisfaction, support, commitment, involvement, etc.). “Objective” characteristics, such as reductions in stressors, were left out.

-Following Bamberg and Busch (1996), the category of cognitive stress coping was added. This category consisted of the factors that were covered in chapter 1.4 and included coping style, self-efficacy, etc.

Despite the fact that all of the above outcomes represent serious and important consequences of stress, the present study was mainly interested in mental health outcomes. Because only psychological outcomes were available for *all* studies (see the section below on inclusion/exclusion criteria), the mean psychological effect was the only measure that was reliable enough for firmer conclusions. The other mean effect sizes should be more interpreted as trends.

Research question B: Homogeneity analysis

In this second research question, the individual effect sizes of step A were further analysed to check whether the results are homogeneous (as was also done by Bamberg and Busch in their 1996 meta-

¹¹ This means that psychological dysfunction should be focused at the self, such as in the case of depression. “Externalised” symptoms such as aggression and alcoholism are excluded.

¹² Psychophysiological means that these are indicators that are directly reflected in fluctuations in psychological arousal.

analysis). Homogeneous results mean that all study outcomes can be regarded as reliable estimators of a single population effect size. It would also mean that there is less need to assume that the “apples and oranges problem” is biasing the analysis. When results are not homogeneous, however, the data will be further explored. In this case, it is important to find out which studies cause the heterogeneity. By analysing such studies, it might be possible to identify patterns of inconsistencies, which can provide important additional information regarding the effect hypotheses (Wolf, 1986). This information can then be used in addressing research question C.

Research question C: Decisions on research hypotheses

The third research question pertains to addressing the effect hypotheses that were formulated in the previous chapter. Doing this is a means of determining the presence of effect predictors. In accordance with the hypotheses, attention is directed at the 9 variables that are listed below (for a full description see previous chapter). It shall be analysed whether they are linked with program effectiveness. The major part of this analysis shall be based on quantitative statistical procedures such as ANOVAs and Multiple Regression Analysis, with the hypothesised effect predictor as an independent variable and the relevant effect size as the dependent. Moreover, when the homogeneity analysis would show that the sample is heterogeneous, it is possible that an additional qualitative analysis takes place (see 3.2.5). When this is the case, the conclusions from the statistical analyses will be supplemented by qualitative data. Thus, when the sample is heterogeneous, both quantitative and qualitative methods can be used to draw conclusions on the presence of effect predictors.

Table 3.1
Research hypotheses

Hypothesis
1 Programs that are aimed at high-risk groups are more effective than programs that are not
2 Programs are only effective on those variables that are closely linked to their method
3 Organisational methods just as effective as individual-oriented ones
4 In the period from post-test to follow-up, organisational interventions will increase in effectiveness relative to individual ones.
5 Interventions after working hours are more effective
6 There is a positive relationship between duration of individual programs and effectiveness
7 Programs carried out on site are more effective
8 Programs that use multiple methods are more effective
9 There is an impact of study design on effect size

3.2.2 Literature search and selection of studies

When research questions for the analysis are formulated, the next step is to select and obtain the data material on which conclusions can be based. In the case of a meta-analysis, this raw material is formed by empirical studies. However, for most meta-analyses, the step of searching and selecting appropriate studies is problematic (Wolf, 1986). Here, potential dangers of subjectivity and selectivity arise. Moreover, perhaps the greatest threat to validity in this step is the so-called file drawer problem. This problem refers to the fact that, often, only research projects with positive and significant results are published. In cases of non-significant or negative intervention effects, reports typically end up in university file drawers, from which they are difficult to retrieve. For a meta-analysis, this means that estimates of

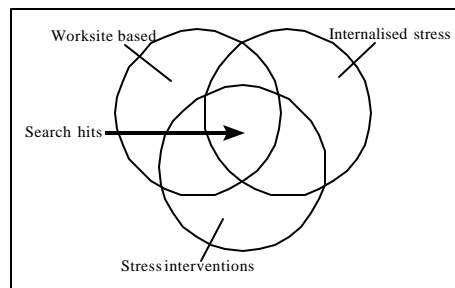
effect size are typically inflated. Because of these problems, the present study took numerous steps to preserve validity. Below, the search and selection procedure shall be described in more detail.

Construction and execution of the computerised search procedure

In order to reduce selectivity bias, it was tried to retrieve as much of the articles as possible from universally accessible research databases. This was done using a logically constructed and stepwise search strategy. The construction of this search procedure can be described in the following way. First, possible search terms were identified. For this matter, the thesaurus of PsycLit (the psychological database) was browsed for key words that are related to the stress concept (as described in chapter 1 and 2). This resulted in a preliminary list of search terms. After this, this list was used to perform pilot searches. Consecutively, the pool of studies that resulted from this try out search was compared to that of other literature reviews and meta-analyses, such as those of Murphy (1996), Van der Hek and Plomp (1997) and Bamberg and Busch (1996). Differences in results were analysed and resulted in ideas for additional terms. This process was repeated until the results of the search procedure were reasonably in accordance with the other sources. For example, compared with Murphy's results, try outs continued until the computerised search procedure covered all but 9 usable studies from Murphy's analysis sample of 83¹³ articles.

The final result of the process described above was a search strategy that can be described in the following way. First, key words relating to stress and its symptoms were inserted (see appendix 1 for a complete list of all terms). The first cluster of these key words included search terms that were related to internalised psychological symptoms of stress (see 3.2.1, footnote 6 for how this was defined). The second cluster consisted of key words that are indicative of a stress management intervention, such as "biofeedback", "meditation" and "primary prevention". Thirdly, a third cluster aiming at worksite interventions was inserted. This cluster consisted of terms such as "occupational stress" and all kinds of descriptions for personnel, such as "aerospace personnel" and "business and industrial personnel". This cluster was included to select only those interventions that are implemented in a worksite context¹⁴. After these three clusters were inserted, they were combined to search for studies that shared at least one key word of every category. The result of this combination can be depicted in the Venn-diagram (figure 3.1) below. Because the result of this grouping resulted in too many studies to be inspected individually (180885 articles), they were further narrowed down using a compound of search terms that aimed at experimental studies¹⁵ (see also appendix 1).

Figure 3.1. Venn diagram of computerised search procedure



¹³ Note that Murphy's (1996) inclusion criteria did were different from the present analysis. Because Murphy also included non-experimental studies, his sample was enlarged by many studies that were ignored in the current search procedure.

¹⁴ Worksite interventions were defined as carried out within the context of a profession and dealing with general workplace distress. Because of this latter criterion, studies that targeted only specific outcomes, such as flight anxiety, were excluded

¹⁵ Experimental studies were defined as interventions with both an experimental group (i.e. receiving the intervention) and a control group (excluded from fundamental manipulations; examples are a waiting list or no-treatment control group). As an essential feature of the experiment, intervention and control group need to be equivalent on important variables.

The above-described procedure was run on two scientific databases. Because the study was primarily interested in the effects of stress interventions on psychological outcomes, the main search efforts concentrated on the psychological PsycLit database (accessed from the University of Nijmegen computer system). Parallel, however, a similar search was performed in Medline. In PsycLit, 188 "hits" came out of the search process, whereas the result in Medline was 111 studies. In addition to these computer hits, 9 additional articles were taken from the Murphy (1996) review. Furthermore, in the Bamberg and Busch meta-analysis (1996), 3 extra articles came up. And finally, 4 articles that were cited by other authors but did not show up in any of the computerised searches, nor in the other reviews, were included. All together, the resulting pile of articles thus counted 315 studies.

Inclusion/exclusion criteria

After the systematic search had resulted in the above-described 315 studies, the pile of articles had to be screened manually in order to determine which of them could be included in the analysis. Below, the criteria for exclusion and inclusion are listed:

-First of all, studies that did not contain an intervention program evaluation were excluded. This was defined a program that is assessed according to some quantitative outcome variable.

-Studies that still pertained to other settings than a worksite context (see footnote 9 for how this criterion was defined) were also left out. These included for example interventions that were run on students.

-Moreover, studies that were too old to resemble an up-to-date description of the field were excluded. The cut-off point for this criterion was set at the year 1980.

-Fourthly, the program needed to be evaluated on the basis of at least one mental health outcome (symptom or syndrome of internalised distress; see 3.2.1).

-Studies regarding the post-traumatic stress syndrome (PTSS)¹⁶ were excluded. PTSS is a distinct psychopathological condition that differs from stress in both symptoms and aetiology.

-For reasons of retrieval, the study needed to be available through the University of Nijmegen library system or the Dutch IBL article copying service.

-Studies that were written in an unknown language could not be included. In the present study, these were studies in other languages than English, German or Dutch.

-Studies without equivalent experimental and control group were also left out. Primarily, this equivalence was defined as involving a randomised assignment to groups. In a limited amount of cases, it was also sufficient when a study corrected post-test values for biasing factors (such as differences at pre-test). Finally, those cases in which the experimental and control group were analysed for pre-test differences and found equivalent were also included¹⁷.

-Finally, when studies did not present enough statistical information to calculate d , they were excluded as well (see 3.2.4 for the exact prerequisites). However, because the present research focused primarily on mental outcomes of stress, studies were only excluded when there was a lack of information to calculate the psychological effect sizes.

¹⁶ A DSM-IV condition that occurs as a reaction to highly threatening events. In order to arrive at this diagnosis, a clients should be bothered by frequent re-experiences, avoidance of trauma-related stimuli and an increase in tension.

¹⁷ One study (3/p/#24) used a control group that was selected on the basis of its equivalence. An *inspection* of the pre-test means showed only very small differences (less than half a Standard Deviation). This study was included because it was thought to meet the requirements for a calculation of d sufficiently enough to defend a decision in favour of retaining statistical power.

When the above-described inclusion/exclusion criteria were applied on the pool of 315 possible studies, this resulted in the following picture. Of these articles, 107 contained no description of an evaluated study. Furthermore, 19 articles had to be dropped because they were written in unknown languages. Finally, three studies appeared as “double hits” (i.e. the same article appeared both in the PsycLit and the Medline search). Thus, the preliminary number of studies was 186. Of those articles, 3 had to be dropped because they were dated before 1980 and 1 study because it was focused on PTSS. Nine studies could not be retrieved from the available library systems. Furthermore, 56 studies were excluded because they were not implemented in a worksite context. In these excluded studies, stress interventions were carried out with following target groups: immigrants (1), medical patients (20), psychiatric patients (4), the unemployed (2), children (10), the elderly (4), soldiers (2), students (7), parents (3) and communities (3). In the next exclusion step, 31 research reports were left out because they did not target mental health related outcomes. Excluded targets were: knowledge (1), competency (9), posture (2), alcoholism and drug abuse (5), aggression (4), back pain (2), economic costs (1), computer skills (1), death anxiety (1) and physiological outcomes (5). Finally, 27 studies were excluded because they lacked a randomised control group and another 21 studies were unusable because they did not provide sufficient statistical data. Thus, there were a total of 277 excluded studies, which brought the final number of usable articles to 38. Because nine of these articles included two intervention samples, the final N of the analysis was 47 (exact bibliographical data can be found in the reference list at the end of the essay).

Addressing the file drawer problem

Although it is difficult to eliminate the file drawer problem completely, there are two possibilities to reduce its impact. Firstly, a meta-analyst can try to overcome the problem by collecting as many as possible studies from unpublished sources. For this purpose, there are a number of channels, such as the retrieval of unpublished dissertations (e.g. with the use of Dissertation Abstracts), and reports and presentations that are given at professional meetings. Although this solution is potentially very useful, the present meta-analysis was faced with too many time and infrastructure constraints (e.g. the Nijmegen library system did not have access to international dissertations) to retrieve such unpublished material. Therefore, the present study needed to resort to the second solution of calculating the so-called fail-safe N^{18} . This is a statistic that conveys the number of studies with negative outcomes that would be needed to cancel the conclusion that a certain intervention effect exists. When this fail-safe N is large enough, the possibility that the positive effect sizes are a mere artefact of the file drawer problem is highly improbable. However, although a large fail-safe N adds credence to the existence of a real intervention effect, it remains probable that the current findings will overestimate effect size.

3.2.3 Coding system and procedure

Construction of coding system

In order to classify and code the selected studies for the analysis, a specialised coding system developed by the Nijmegen Prevention Research Centre (NPRC) was used. This system was created with the purpose of coding and classifying different kinds of preventive interventions in a systematic way. The coding system is the product of an intensive and long collaboration between the NPRC and professor Hendricks-Brown from the university of South-Florida. Before its use in the present essay, the coding system had already been successfully used in a number of other research areas, including depression and school-based interventions. In the design of the system, great attention to considerations of validity was paid. First of all, experienced prevention scientists assessed the construct validity of the system. On the basis of this, numerous revisions took place. The system that was used for the present study already belonged to the third full revision. Secondly, eventual users of the system are given the opportunity to test it on a “neutral” study (i.e. not belonging to the field in which they were going to use it). This provides the opportunity for feedback to the constructors, and also for getting further acquainted with the frame of reference behind the device. Finally, a coding book is available that assists the coder in

¹⁸ This shall be done using the equation of Orwin (in Wolf, 1983/1986): $N_{fs} = \frac{N(\bar{d} - \bar{d}_c)}{\bar{d}_c}$. In this

equation, N_{fs} is the fail-safe N, \bar{d} the general effect size and \bar{d}_c the criterion value.

making deliberate and valid decisions. To reduce conceptual confusion, the coding system is furthermore provided with a separate definition book.

Construction of two additional variables

Because the coding system was designed for preventive interventions in general, area-specific distinctions sometimes remain uncovered. In the present study, this was the case for the variables “working hours” and “location” (see chapter 1 for a full description). For these two factors, the system needed to be supplemented with additional variables. For the “working hours” factor, a three level variable was created that could be coded as either “within”, “after” or “not clear”¹⁹. When the intervention was carried out during a pause, this was coded as “within”. This was done because when the pause break is filled with a stress intervention, personnel have still received no opportunity to recover from build-up stress. For the second factor “location”, the construction process was executed in much a similar way. The construction process resulted in a three level variable (again including a “not clear” level) that coded whether the program was carried out on or off site. Together with the other hypotheses, this resulted in a coding sheet (see appendix 2 for an example) that consisted of the variables that are listed below.

1. General information

This block of variables was mainly composed of general bibliographical information, such as the author of the article, the magazine in which it appeared, the ISBN/ISSN number, year of publication, etc.

2. Status

Here, the factors of “location” and “working hours” were coded. As stated above, these were three-level variables with a separate level for cases in which the study's information was not clear enough to be reliably coded.

3. Time

Here, three aspects were coded: the number of sessions, the duration of the program (in weeks) and the session time (in minutes). This resulted in three continuous variables. Programs for which one of these variable was regarded as irrelevant (e.g. stressor reduction interventions without any sessions) were assigned a missing value. Sometimes, the coding of the number of sessions was complicated by the use of home practice. Here, it was decided to count the number of plenum sessions. Only when home practice was the dominant component of the intervention, the nr. of practice sessions was counted. Moreover, this was only done when participants were given a relatively brief instruction and then left to practice a certain minimum a week.

4. Risk status

This was coded in a variable involving the levels “high-risk” and “normal/not stated”. An intervention fell in the first category when its target group was either characterised by the presence of important risk factors (selective preventive intervention) or the experience of an elevated level of symptoms (indicative preventive intervention). The reason to group these two cases together was that they were encountered combined extremely often. Support for this decision comes from Murphy, who did the same in his 1996 review.

5. System level

This was coded in a three level variable. When a program was aimed at the individual worker, this was coded as “micro”, whereas interventions that were aimed at structural factors were classified as “meso”. Where both levels were involved, the study was coded as “mixed”.

6. Methods

On the individual level, the method variable could be coded into 7 possible levels. These were muscle relaxation, biofeedback, exercise, meditation, cognitive-behavioural skill training and “others” (see chapter 2). When there was more than one of these method used, this was coded as a combination methodology. For organisational interventions, methods could be coded along

¹⁹ This last level was added because exploration of the data showed that information regarding this working hours distinction was often ambiguous or non-existing.

three categories: programs that altered social support, programs that targeted the level of control and programs that were focused on stressor reduction.

7. Research design

The quality of a study's research design was coded with the use of 9 separate Cochran items. The content of these items covered the following topics: (1) definition of aims, (2) definition of the intervention, (3) randomly assigned control group, (4) equivalent control group, (5) reporting of the number of subjects in trial, (6) existence of pre-intervention data, (7) existence of post-intervention data, (8) reporting of attrition rates and (9) the presentation of results for all the outcomes (see appendix 3 for the exact operationalisations of the items). For the total score on this variable, the number of items scored with "yes" was summed up. This resulted in a continuous variable with a possible range of 0-9.

8. Protocol

This two-level variable coded whether the intervention was reportedly structured along some source of written material. This written source needed to be a clearly specified book, research article or manual. A simple "based on" was not enough. In all other cases, the variable was coded with no/not stated.

9. Profession

Here, the profession of the participants was coded. In the present sample, there were 8 possible levels of this variable. In separate categories, the professions of (1) health care worker (e.g. nurses), (2) educator (teachers), (3) manager, (4) blue collar industrial worker, (5) office worker and (6) law enforcement personnel were scored. An (7) "other" category was constructed for professions that were found less than three times. This included social service workers, bus drivers and postal workers. (8) When there was more than one profession represented in the sample, this was coded as "mixed".

3.2.4 Calculations of effect size

As a measure for effect size, Cohen's d was used. This is a frequently used standardised computation of effect size that is based on the post-test difference between the experimental and the control group (Wolf, 1986). This was done because the present analysis only included studies with an experimental design. Because the experimenter ideally randomises control and treatment groups, they are expected to be equal on all imaginable variables. Any difference at post-test will therefore be due to the intervention and not to any other influences. In some cases, however, sample size may be too small so that randomisation might not result in completely equivalent groups. Because of this, it was chosen to use adjusted means for the calculation of d as much as possible. Such adjusted means control for eventual differences at pre-test, and according to Taylor and White (1992), they provide a more accurate basis for calculating effect size. In the calculation of d , the difference at post-test is divided between the pooled Standard Deviation (see equation 1 below)²⁰. To reduce the possibility of computation errors, all of the calculations were performed using the Microsoft Office 2000 Excel package. When the study contained more than one outcome category, effect sizes were averaged in order to arrive at a general summary of effect. Moreover, when this was possible, effect sizes for the follow-up period were calculated²¹.

Equation 1
Calculation of effect size using Cohen's d

$D = \frac{M_1 - M_2}{S_p}$	<p>Where S_p equals:</p> $S_p = \sqrt{\frac{(N_1 - 1)SD_1^2 + (N_2 - 1)SD_2^2}{N_1 + N_2 - 2}}$
-----------------------------	--

²⁰ When the post-test Standard Deviation was not reported, the pre-test SD was used.

²¹ When there was more than one follow-up, the follow-up that was closest to post-test was taken.

As was stated in 3.2.2, the last exclusion criterion of the present study was the unavailability of adequate statistical data. As can be seen above (equation 1), Cohen's d is calculated by using means and Standard Deviations. However, many studies fail to provide these data, although they do report some other outcome statistics. Often, such studies only report F or t -values for comparisons between the experimental and control groups. Because there are formulas for converting these statistics to d (Wolf, 1986; Taylor and White, 1992; Ray and Shadish, 1996; Lyons, 1998), these studies were not always excluded. Alternative equations can be reliably used because they are logically derivable from equation 1 and produce equivalent results (Taylor and White, 1992). However, because they can only be used when this did would not violate the mathematical logic, they cannot be applied to all F and t statistics. For the statistics of multiple factors (e.g. two-way) ANOVAs and gain score ANOVAs, this led to exclusion. The same was true for unconventional tests such as the Mann-Whitney U test and regression models. For these tests there exists no accepted method for conversion to d .

In one special case, the unavailability of means and Standard Deviations did not lead to the exclusion of the study. In a selective number of studies, no summary statistics were provided. Instead of this, only the conclusion that statistical tests did not yield any "significant intervention effect" was communicated. Following the recommendations of Cohen (in Wolf, 1977/1986) and Durlak (1997), in these cases, effect size was estimated as zero.

At the end of this section, a point has to be made about the possibility of weighing effect size by sample size. In such a weighting procedure, studies with a larger number of participants are given more influence in the calculations of effect size. This is done because such studies provide more stable, reliable results. Although this might be a useful alternative for many meta-analyses, the present study did not perform such a sample correction because of two reasons. Firstly, the assumptions of the meta-analysis technique imply that error fluctuations (that are indeed larger with smaller samples) are levelled out across the analysis. When sample size is greater than 10, a weighting procedure does therefore not make much of a practical difference (Green and Hall (1985) in Wolf, 1986). The other important reason for using uncorrected values is that sample size is not the neutral factor that it might be in many other research fields. In these other cases, it is possible that effect size does not *meaningfully* depend on the size of the research sample. In stress interventions, however, it is not that simple. Smaller samples mean more attention and involvement from the researcher, who is in most cases also the program provider. This increased attention can cause higher effect sizes.

3.2.5 Analysing, reviewing and interpretation of the results

When the calculations of the previous step have been executed, the final step is to draw conclusions from them. In this section, the process by which these conclusions will be reached is described in more detail.

Research question A: magnitude of effect size

For the first question, drawing a conclusion involves the following procedure. First of all, the above-described calculations will result in a mean effect size for each outcome category. In interpreting these figures, the guidelines of Cohen (in Wolf, 1986) shall be used. These state that an effect size of 0.20 can be regarded as small, 0.50 as medium and 0.80 as large. In order to check how solid the results of this step are, the fail-safe N shall be computed. When the fail-safe N is large enough, this means that the results of the analysis are sufficiently powerful to withstand numerous potential studies that contradict these outcomes. This in turn means that they are less endangered by the file drawer problem (see earlier).

Research question B: Homogeneity analysis

The outcomes of the homogeneity analysis shall be treated in the following way. When results are homogeneous, the analysis shall move to step C and statistically elaborate the effect hypotheses. If they are heterogeneous, they shall be addressed in more detail. In this case, the following procedure shall be followed. By inspecting the individual τ^2 -statistic²² of all individual studies, it shall be analysed which

²² Indicative of the matter in which the study deviates from the general effect size pattern.

studies cause this heterogeneity. When all studies have a large τ^2 , this would mean that there are probably two or more groups of studies with distinct outcomes patterns. In this case, performing separate meta-analyses for each separate group would be a better option. However, it is also possible that there is only a small amount of studies with extreme τ^2 -values. This would indicate that the heterogeneity is caused more by outliers than by the existence of multiple groups. Regarding the question of what to do with the results of such outlier studies, there is still some debate (Wolf, 1986). Although some authors (e.g. Hunter, Schmidt and Jackson (1982) in Wolf, 1986) recommend excluding such studies from the final analysis, there are strong arguments for leaving them in. It can be argued that heterogeneity is a common finding across studies by different investigators using different methods (Harris and Rosenthal (1985) in Wolf, 1986). Therefore, even Hedges (Hedges (1984) in Wolf, 1986) admits "it is not necessarily inadvisable to draw inferences from heterogeneous effects." Furthermore, a practical argument for keeping outliers in the analysis is that this preserves sample size. Finally, it is assumed that studies fluctuate in effect size according to certain intervention characteristics. Thus, excluding outliers –however extreme– would ignore an interesting source of variation. Thus, when there is only a limited amount of outlier studies, these shall be left in the statistical analysis.

Research question C: Decisions on research hypotheses

There are many statistically methods to address the hypotheses of the third research question. These methods vary across the properties of research data. In the present analysis, this resulted in the following picture. Firstly, when the hypothesis involved the comparison of different groups (e.g. after versus within working hours), the ANOVA procedure was used. When the hypothesis involved a group x time interaction effect, a special kind of ANOVA, namely the repeated measures ANOVA, was employed. In the present study, this was the case for the hypothesis that organisational interventions will increase in effectiveness (relative to individual ones) in the period from post-test to follow-up. Here, post-test and follow-up were the two levels of the factor time, whereas the group factor was the system level at which the program was targeted. Another type of analysis was involved when the independent variable was continuous. This was the case for the variables "duration", "design" and "publication year". In these instances, a regression analysis was performed. The necessary calculations for these analyses were performed using the SPSS 9.0 computer software package. When evaluating and reviewing results, a significance level of 0.05 was adopted. However, because the sample size of 47 was still quite small, close attention was also paid to so-called "marginally significant findings" (i.e. findings with a significance level between .05 and .10).

When the homogeneity test (research question B) would result in the identification of a small number of study outliers, the quantitative procedures shall be supplemented with a so-called qualitative analysis (as suggested by Light and Pillemer (1984) in Wolf, 1986). This would be done to inspect the outliers and see if meaningful patterns can be discerned. Because there are only few agreed-on rules for applying qualitative analysis, only its general structure can be described here. According to the handbook that was published by the US National Science Foundation (1997), qualitative analysis involves three distinct steps. The first step involves the process of data reduction, in which relevant material for the analysis is defined and retrieved. For the present study, this would mainly be done by the homogeneity analysis, which is able to identify which studies represent outliers and require a closer look. For the second step of data display, the outliers would be displayed in a matrix that summarises their most important features. In principle, it is the intention to compare them to an appropriate reference group in order to see what distinguishes both extremes from each other. For example, when the outliers are very ineffective, they should be compared to a group of successful studies. In the third and final step, the displayed data would be used to draw conclusions on the research question(s). In the present essay, the question would be whether a qualitative analysis supports, supplements or contradicts the findings from the statistical analyses. In this way, a qualitative analysis can lead to an increased understanding of the research matter and provide an aid in generating new hypotheses (Wolf, 1986).

Chapter 4: Results

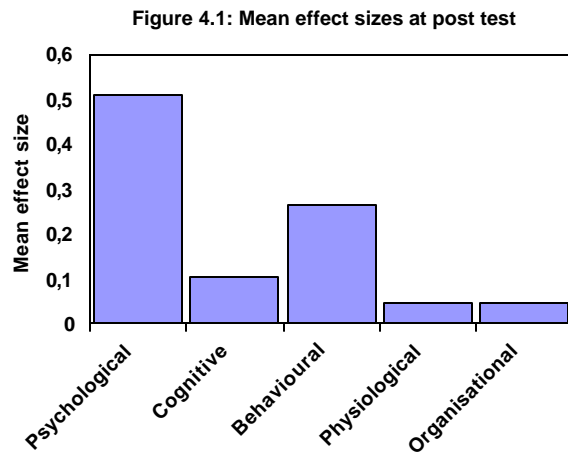
4.1 Research question A: Magnitude of effect size

Effect size at post-test

Regarding the intervention effect sizes at post-test, the following picture emerged. The greatest effect occurred for the psychological outcomes, for which a value of 0.51 was found (see table 4.1). This figure was based on a total number of 2,287 participants over 47 studies (average $N = 49$, $SD = 43$). Following Cohen's guidelines (0.20 is small, 0.50 is medium and 0.80 is large effect), this can be interpreted as a medium effect size. The sturdiness of this effect size can be assessed by looking at the fail-safe N . With a criterion of 0.50 d , the fail-safe N takes the value of 1.0, which means that only one negative study would reduce the average effect to a value under 0.50. When the criterion value is set at 0.20, however, a fail-safe N of 73.1 emerges. This signifies that a very large number of negative studies would be needed to reduce the mean effect size to Cohen's "small" level. After psychological effect sizes, the second greatest effect occurred in the behavioural category (see figure 4.1), which noted an effect size of 0.26 (small effect). However, the fail-safe N indicated that only 3.5 negative studies would be needed to reduce this value to a level below 0.20, which signifies that this outcome is not very robust. The other categories of cognitive, physiological and organisational outcomes noted effect sizes of 0.11, 0.05 and 0.05 respectively. The impact on these variables can therefore be regarded as very small to zero.

Table 4.1
Mean effect sizes at post-test

Outcome	Psychological	Cognitive	Behavioural	Physiological	Organizational
Mean d	0.51 (N = 47)	0.11 (N = 8)	0.26 (N = 11)	0.05 (N = 21)	0.05 (N = 19)



Effect size at follow-up

With regard to the follow-up measures, an initially spectacular picture emerged. An average 24 weeks ($SD = 27$) after post-test, the psychological values increased dramatically. The behavioural and physiological measures also showed considerable gains, whereas the impact on cognitive measures reversed from positive to negative (see table 4.2).

Table 4.2
Mean effect sizes at follow-up

Outcome	Psychological	Cognitive	Behavioural	Physiological	Organizational
Mean <i>d</i> (N)	0.76 (N = 15)	-0.17 (N = 4)	0.48 (N = 5)	0.15 (N = 5)	0.04 (N = 9)

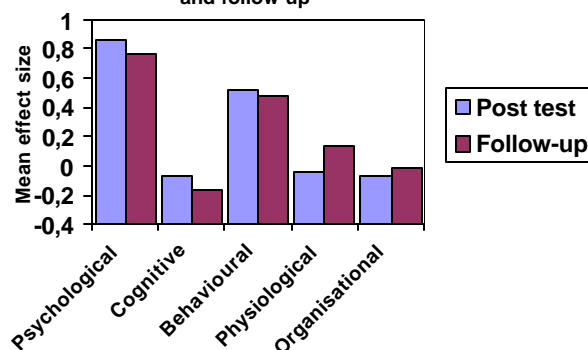
Closer inspection of the data revealed another picture, however. Because not all studies included a follow-up test, it could well be that programs with and without a follow-up already differed in effectiveness at pre-test. To test this explanation, a 2-way ANOVA was computed with the availability of follow-up as the independent variables and post-test effect size as the dependent variable. For the categories with great enough sample size to perform such an analysis (i.e. around 15-20), the analysis was significant for psychological values ($F(1) = 6.70$, $p = .01$), but not for physiological ($F(1) = 0.14$, $p = .71$) or organisational values ($F(1) = 1.45$, $p = .25$). This means that studies with a follow-up measure already had higher psychological post-test values (mean $d = 0.87$) than those who did not (mean $d = 0.34$). Therefore, paired comparisons were made (between studies that reported both follow-up and post-test). When this is done, another picture emerges (see table 4.3 and figure 4.2).

Table 4.3
Paired effect size means and Standard Deviations at post-test and follow-up

Mean <i>d</i> (SD)	Psychological	Cognitive	Behavioural	Physiological	Organizational
Post-test	0.87 (1.01)	-0.08 (0.44)	0.52 (0.66)	-0.05 (0.13)	-0.07 (0.35)
Follow-up	0.76 (0.78)	-0.17 (0.59)	0.48 (0.62)	0.14 (0.19)	-0.02 (0.36)
N	15	4	5	4	7

As can be seen, the psychological effect size now changed from 0.87 to 0.76, which indicates a slight decrease (see figure 4.2). For the other categories, physiological and organisational measures changed from -0.05 to 0.14 and -0.07 to -0.02 respectively. This indicates that the effect on physiological interventions increased at follow-up, whereas effectiveness on organisational outcomes remained more or less stable. For the categories of cognitive and behavioural outcomes, analysis of the difference analysis between studies with and studies without a follow-up was not possible, which makes a comparison over time extremely problematic. However, trends were apparent in that cognitive outcomes decreased from -0.08 to -0.17, whereas behavioural measures remained somewhat constant (0.52 at post-test versus 0.48 at follow-up).

Figure 4.2: Pairs of mean effect sizes at post test and follow-up



4.2 Research question B: Homogeneity analysis

As stated in chapter 3, a test of homogeneity was performed in order to see whether the calculated outcomes can be treated as valid estimates of the same global effect size. Alternatively, the test yields evidence for the existence of either multiple groups or one or more outlier. Firstly, however, an exploration of the data was performed. Because the equation for homogeneity weights studies for sample size, it was necessary to identify sample size outliers. These needed to be excluded because, otherwise, studies with extreme sample sizes would exert an exaggerated influence on the results. Against a mean sample size of 49, these were studies 3/p/#36 (N = 187) and 3/p/#139 (N = 245). The final analysis was therefore based on 45 studies.

As can be seen below (table 4.4), the sample is significantly heterogeneous ($\chi^2 = 13.73$, $p < .0005$). Further inspection shows that this heterogeneity is caused by five studies (studies 3/p/#85, 3/p/#145a, 3/p/#157a, 3/p/#157b and 3/p/#134), which is reflected by their significant p -values (marked with *). Four of these studies can be characterised by extremely high effect sizes. These are studies 3/p/#157a and b (Sharp and Forman, 1985), study 3/p/#134 (Stanton, 1988) and study 3/p/#145a (Tunnecliffe, Leach and Tunnecliffe, 1986). The fifth outlier study that contributed to the heterogeneity outcome (study 3/p/#85 by Freedy and Hobfoll, 1994) is characterised by a small effect size. The Freedy and Hobfoll study does not represent a very extreme value, though. The fact that it was nonetheless identified in the homogeneity analysis is due to its large sample (N = 87).

Table 4.4
Test of homogeneity: individual and cumulated chi square statistics and probability values

SPSS nr.	Study nr.	Effect size	Chi square	Cumulative Chi Square	P^{23}
35	3/p/#133	+0.56	0.02	0.02	
22	3/p/#151b	+0.43	0.04	0.06	$P > .25$
...
13	3/p/#20	-0.02	3.23	55.08	$P < .10$
19	3/p/#85	+0.01	5.44	60.52	$P < .02^*$
29	3/p/#145a	+3.20	8.31	68.83	$P < .005^*$
16	3/p/#157a	+1.99	11.00	79.83	$P < .0005^*$
17	3/p/#157b	+2.04	11.49	91.32	$P < .0005^*$
20	3/p/#134	+1.93	13.73	105.05	$P < .0005^*$

*Causing significant heterogeneity at $p < .05$

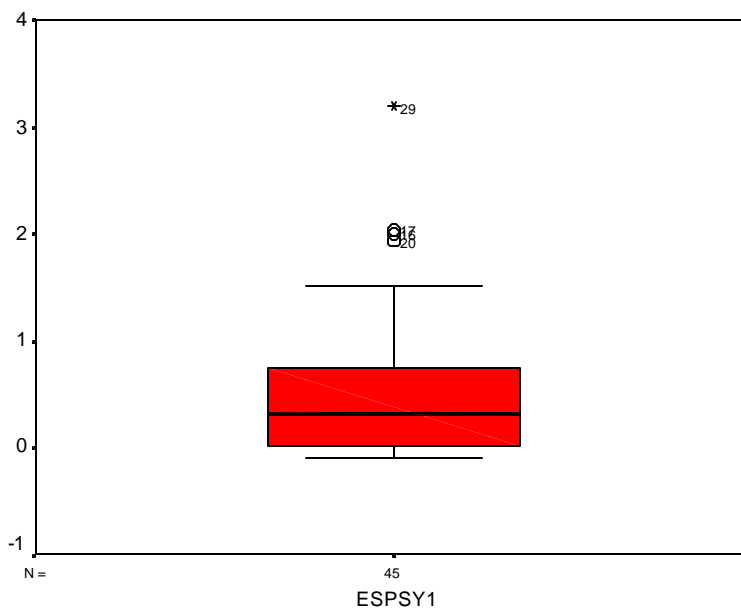
After concluding that five outlier studies in the present sample cause heterogeneity, it had to be decided how to treat the findings of these studies. As stated previously, there are two possibilities. Firstly, it is possible that the sample can be characterised by two or more groups. In this case, the analysis is confronted with the already-mentioned "apples and oranges problem", which would make running one and the same meta-analysis on these different groups problematic. Secondly, the identified studies could represent outliers that do not so much constitute a separate group, but are for some reason more effective than the other interventions. In order to decide which is the case, the results were graphed in a box-plot, which is depicted below (figure 4.3). Here, it becomes clear that the 4 positive outliers represent values that lie well above the normal range. Within this subgroup, the most effective study even takes a

²³ P-values were assigned according to Moore and McCabe (1994).

separate position with an effect size that is more than one unit of d higher than the other outliers, which are clustered together around a value near $2d$.

On the basis of this information, it is still very difficult to rule out any of the two possible scenarios. There are three studies that are clustered tightly together can indicate that a separate class of interventions is at work. However, the fact that two of these studies come from the same article suggests that intervention characteristics, which are for a large part stable across studies of the same article, may play a role. Thus, the extreme studies might not belong to a distinct group but only reached extreme values because of certain intervention characteristics. This would mean that they are best treated as a limited collection of outliers. Because of this last argument, it was decided to keep these studies in the statistical analysis. However, as stated in the previous chapter, they were further explored with the use of a qualitative analysis.

Figure 4.3
Graphed distribution of psychological effect sizes



4.2 Research question C: Decisions on research hypotheses

4.2.1 Quantitative analysis

Hypothesis 1: High risk programs are more effective

As stated in the previous chapter (3.2.5), the first hypothesis was researched with a one-way ANOVA. As can be seen in the table below, high-risk programs did indeed have a higher effect size (0.53) than non-high-risk programs (0.41). However, this difference did not reach adequate levels of significance ($F(1) = 0.22$, $p = .65$). This means that hypothesis 1 was not confirmed.

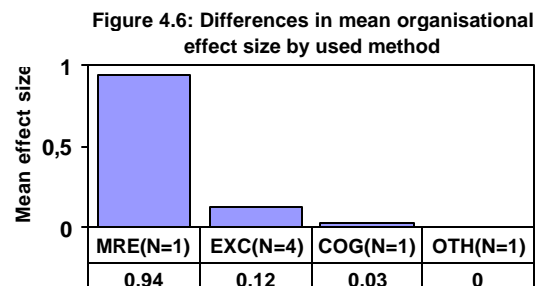
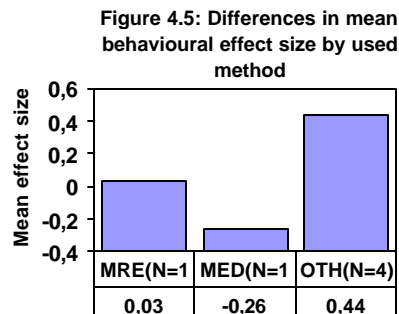
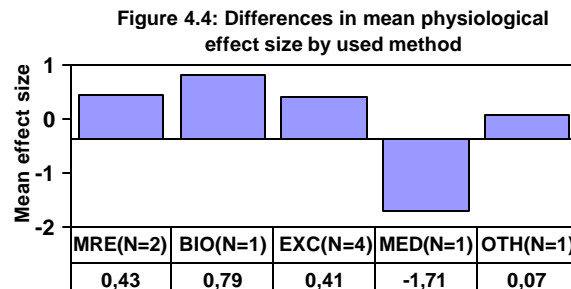
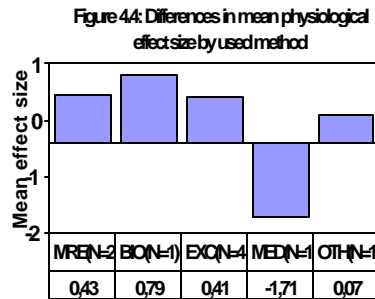
Table 4.4
Means, Standard Deviations and ANOVA summary for the effect of “risk status” on psychological effect size

Groups	N	Mean (SD)	Df	F	P
High risk	39	0.53 (0.74)	1	0.22	.65

Normal/not stated	8	0.41 (0.30)			
-------------------	---	-------------	--	--	--

Hypothesis 2: Programs are more effective on method-related variables

Because it was not clear whether the methods in the "other" category²⁴ would form a coherent collection, the second hypothesis was researched twice: one time including the "other" category and one time without. Moreover, in both tests, the physiologically-oriented methods were grouped together. Unfortunately, for all outcome categories except psychological measures, sample sizes were insufficient to perform an adequate statistical analysis. For these outcome variables, only trends could be investigated. As was expected, cognitive-behavioural interventions were most effective on cognitive outcomes (see figure 4.3²⁵). Meditation produced a slightly negative effect here. Also in accordance with the hypothesis is the fact that exercise, muscle relaxation and biofeedback were the most successful methods for physiological measures (see figure 4.4). Contrary to expectations, however, meditation produced a very large negative effect in this category. In the latter case, however, results were only based on one study, which makes conclusions difficult to draw. Finally, although there had been no predictions for these categories, differences in behavioural and organisational effect were found. For behavioural measures, "other" interventions were most successful, whereas meditation produced a negative change (see figure 4.5). Muscle relaxation was most effective in changing organisational variables, whereas cognitive-behavioural interventions and exercise produced smaller effects. In the same category, there was no change for programs that used "other" methodologies (see figure 4.6).



Apart from the inspection of trends, the only outcome category that could be evaluated statistically was psychological outcomes. Here, the hypothesis would predict the largest effect for cognitive-behavioural programs. Inspection of the means confirmed this prediction. A very large effect size was also obtained for muscle relaxation programs, whereas exercise and "other" methods reached more moderate effects (see figure 4.7). As can be seen in the results (table 4.5), however, these differences were not

²⁴ I.e. not belonging to the main categories of muscle relaxation, biofeedback, exercise, meditation and cognitive-behavioural training.

²⁵ For the tables 4.3 - 4.6, the following abbreviations were used. BIO = biofeedback, COG = cognitive-behavioural skill training, EXC = exercise, MED = meditation, MRE = muscle relaxation, OTH = "other" methods.

statistically significant. ($F(5) = 0.70, p = .63$). For the analysis without the "other" category, results were similar ($F(4) = 0.73, p = .59$). This means that the second hypothesis was also rejected.

Figure 4.7: Differences in mean psychological effect size by used method

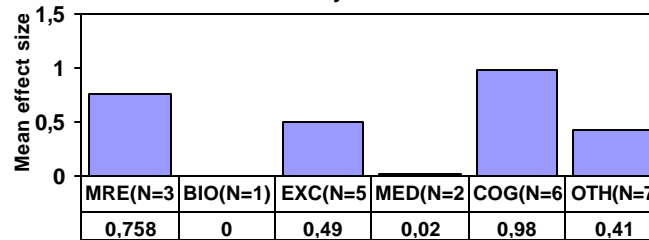


Table 4.5
ANOVA summary for the effect of "method" on psychological effect size

Between subjects factor	Df	F	P
Method	5 (4)*	0.79 (0.73)*	.63 (.59)*

*Without the "other" category

Hypothesis 3: Organizational interventions are just as effective as individual ones

Below, the results are presented for the ANOVAs with "system level" as the independent variable. As stated earlier, analyses were only possible in samples that were greater than 15-20 studies. This was the case for psychological, organisational and physiological outcomes. For the two latter categories, results were non-significant, although trends were apparent. Inspection of the means showed that individual interventions showed greater effect sizes for physiological outcomes (table 4.7), whereas organisational interventions were more effective for organisational outcomes (table 4.8). For psychological effect size, however, there did appear a marginally significant effect ($F(1) = 3.92, p = .06$). However, inspection of the means showed that these differences were not in the expected direction (table 4.6). Individual interventions were more effective ($d = 0.54$) than organisational ones ($d = 0.12$). Because of this, hypothesis 3 was discarded.

Table 4.6
Means, Standard Deviations and ANOVA summary for the effect of "system level" on physiological effect size

Groups	N	Mean (SD)	Df	F	P
Micro	15	0.08 (0.66)	1	0.07	.80
Meso	5	-0.00 (0.00)			

Table 4.7
Means, Standard Deviations and ANOVA summary for the effect of "system level" on organisational effect size

Groups	N	Mean (SD)	Df	F	P
Micro	11	0.01 (0.54)	1	0.22	.65

Meso	7	0.11 (0.19)			
------	---	-------------	--	--	--

Table 4.8
Means, Standard Deviations and ANOVA summary for the effect of “system level” on psychological effect size

Groups	N	Mean (SD)	Df	F	P
Micro	35	0.54 (0.61)	1	3.92	.06
Meso	8	0.12 (0.12)			

Hypothesis 4: The relative efficacy of organisational interventions increases at follow-up

As stated in chapter 3, this hypothesis was tested with a repeated measures ANOVA. When the hypothesis would be confirmed, a significant interaction effect would take place, with better relative means for organisational interventions at follow-up than at post-test. As can be seen below (table 4.9), there were trends in the expected direction, with organisational methods being 25% less effective as individual ones at post-test, compared with only 16,9% at follow-up. This can be seen more clearly in figure 48, where the effect lines of both types of interventions run convergent. However, because the interaction effect was non-significant ($F(1) = 0.19, p = .67$), hypothesis 4 is rejected (see table 4.9).

Figure 4.8: Differential effect of system level on psychological effect size at post test and follow-up

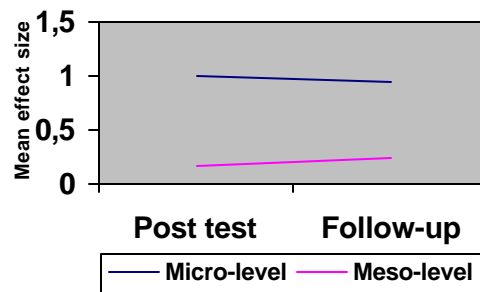


Table 4.9
Means, Standard Deviations and ANOVA summary for the interaction effect of “system level” and “time” on psychological effect size

Groups	N	Mean (SD)	Df	F (interaction)	P	
Post-test	-micro	9	1.00 (0.89)	1	0.19	.67
	-meso	5	0.17 (0.12)			
Follow-up	-micro	9	0.95 (0.84)			
	-meso	5	0.24 (0.34)			

Hypothesis 5: Interventions after working hours are more effective

Because the rationale behind this hypothesis was based on a learning paradigm, the position of the exercise method became somewhat unclear. Since exercise does not directly imply the learning of new skills, it was questionable whether the hypothesis would still apply. Therefore, the hypothesis was tested on two samples: one including exercise and one without. This was only possible for psychological effect

size, however, because the other measures all had sample sizes lower than 15. As can be seen in table 4.10, the observed means were in the expected direction: programs performed after working time had much higher effect ($d = 0.88$) than programs run within the working hours ($d = 0.38$). However, the results were not statistically significant ($F(1) = 2.64, p = .13$).

Table 4.10
Means, Standard Deviations and ANOVA summary for the effect of “working hours” on psychological effect size

Groups	N	Mean (SD)	Df	F	P
During working hrs	10	0.38 (0.40)	1	2.64	.13
After working hrs	7	0.88 (0.85)			

When programs using the exercise method are left out of the sample, a slightly different picture emerges (see table 4.11). Again, differences in effect size are in the expected direction, but without “exercise”, results attain marginal significance ($F(1) = 3.29, p = .09$). This means that for the sample that does not include exercise, hypothesis 5 is partially confirmed.

Table 4.11
Means, Standard Deviations and ANOVA summary for the effect of “working hours” (without “exercise”) on psychological effect size

Groups	N	Mean (SD)	Df	F	P
During working hrs	8	0.29 (0.34)	1	3.29	.09
After working hrs	7	0.88 (0.85)			

Hypothesis 6: Programs carried out on site are more successful

This hypothesis was tested in a similar way as 5. Again, analyses were performed on samples with and without exercise. The results are presented below (tables 4.12 and 4.13). Inspection of the means show that trends are not in the expected direction: on site programs are less effective ($d = 0.37$) than of site programs ($d = 0.76$). Moreover, these results were marginally significant ($F(1) = 3.15, p = .09$). This strongly disconfirms hypothesis 6. The same is true for the analysis without exercise, where results did not reach statistical significance either ($F(1) = 2.95, p = .11$).

Table 4.12
Means, Standard Deviations and ANOVA summary for the effect of “location” on psychological effect size

Groups	N	Mean (SD)	Df	F	P
On site	14	0.37 (0.49)	1	3.15	.09
Off site	6	0.76 (0.35)			

Table 4.13
Means, Standard Deviations and ANOVA summary for the effect of “location” (without “exercise”) on psychological effect size

Groups	N	Mean (SD)	Df	F	P
On site	11	0.32 (0.49)	1	2.95	.11
Off site	5	0.76 (0.39)			

Hypothesis 7: Programs of longer duration and more sessions are more effective

The hypothesis that programs of longer duration are more effective was tested with a linear regression model. As predictors, the variables “number of sessions” and “duration in weeks” were used, whereas the dependent variable was the psychological effect size. Below (table 4.14), the results of this analysis are presented. As can be seen, trends were in the unexpected direction, with programs becoming less effective as program length increases (see negative regression-weights). However, these results were not statistically significant ($F(2) = 0.42, p = .66$). Hypothesis 7 was not confirmed.

Table 4.14
MRA summary for the effect of “session number” and “duration” on psychological effect size

Regression model	Beta regression weights (p -value)	N	Df	F	p
Nr. of sessions +	-0.09 ($p = .64$)	38	2	0.42	.66
Duration in weeks	-0.10 ($p = .60$)				

Hypothesis 8: Programs that use multiple methods are more effective

For this hypothesis, only the analyses regarding psychological and physiological effect size had large enough samples to be run adequately. The results for the psychological outcome category showed a trend in the expected direction ($d = 0.64$ for multiple methods against 0.56 for single methods), but this difference was not statistically significant ($F(1) = 0.11, p = .74$). For physiological measures, the trend was in the reverse direction ($d = 0.18$ for single methods versus 0.06 for multi methods) although these findings were also non-significant ($F(1) = 0.73, p = .41$). Therefore, hypothesis 8 was rejected (see table 4.15).

Table 4.15
Means, Standard Deviations and ANOVA summary for the effect of “multi-method” on psychological effect size

Groups	N	Mean (SD)	Df	F	P
Single method	24	0.56 (0.78)	1	0.11	.74
Multi method	15	0.64 (0.65)			

Table 4.16
Means, Standard Deviations and ANOVA summary for the effect of “multi-method” on physiological effect size

Groups	N	Mean (SD)	Df	F	P
Single method	9	0.18 (0.83)	1	0.73	.41
Multi method	7	-0.10 (0.20)			

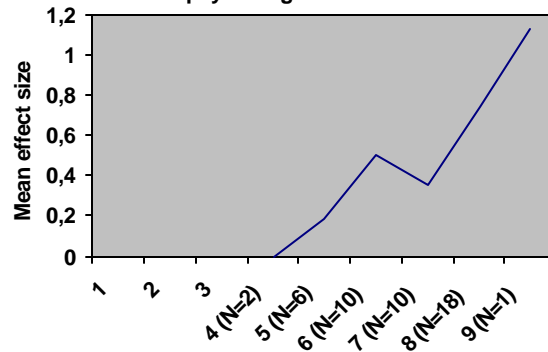
Hypothesis 9: The quality of a study’s research design is related effect

A linear regression analysis was performed on psychological effect size to see whether study quality was significantly related with effect. Below, the results of this analysis are presented. As can be seen (table 4.8), a significant effect was found ($F(1) = 4.94, p = .03$), whereas the regression weight (0.32) shows that this relation was in the positive direction. This means that better designed studies were more effective (see also figure 4.9). Because of this finding, hypothesis 9 was confirmed.

Table 4.17
MRA summary and regression weight for the effect of “research design” on psychological effect size

Regression model	Beta regression weight	N	Df	F	P
Design rating	0.32	47	1	4.94	.03

Figure 4.9: Relation between design rating and mean psychological effect size



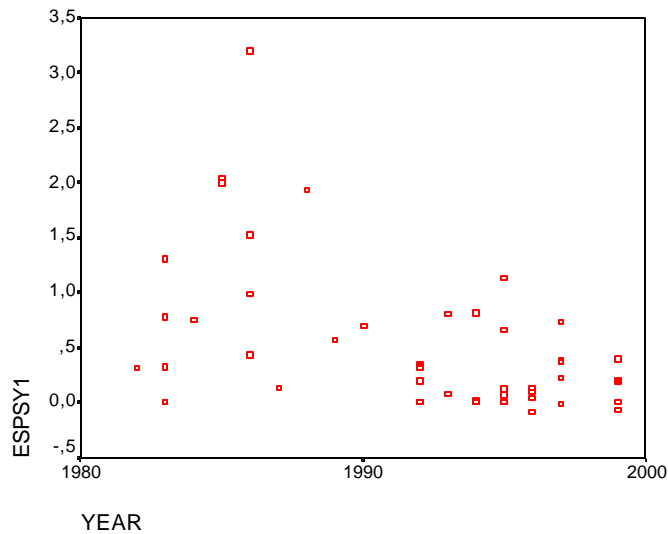
Impact of the year of publication on effect

Although there was no hypothesis regarding this variable, the impact of “year” on psychological effect size was also calculated. In a linear regression analysis, a significant relation was obtained (see table 4.18). There was a negative relationship between “year” and effect size (regression weight of -0.46) that reached statistical significance ($F(1) = 11.81, p = .01$). This means that the later the study is published, the less effective its results are (see figure 4.10; for a possible interpretation of this results see the discussion).

Table 4.18
MRA summary and regression weight for the effect of “publication year”
on psychological effect size

Regression model	Beta regression weight	N	<i>Df</i>	<i>F</i>	<i>p</i>
Publication year	-0.46	47	1	11.81	.01

Figure 4.10
Relationship between publication year and psychological effect size



4.2.2 Qualitative analysis

As stated above, research question B resulted in the identification of a group of 4 positive outliers. To determine what characterised these studies, they were compared against a reference group of the 4 least effective studies. Besides study #85, which was already identified in the homogeneity analysis, these were studies #6 (Winzelberg and Luskin, 1999), #20 (Aust, Peter and Siegrist, 1997) and #49 (Shulman and Jones, 1996). With this comparison, an important part of the first step of a qualitative analysis, the phase of data reduction, is accomplished. To further reduce the complexity of the analysis material, the focus was narrowed down to a selective number of relevant factors and variables. Firstly, these were all the essay's hypotheses that were also addressed in the quantitative review. Moreover, the qualitative part focused on a number of additional factors of which the impression existed that they were potential effect predictors. Mostly because of their complexity, these factors had not been classified in the coding system. These new variables arose during the reading and inspecting of the two groups of articles. They were generated, evaluated and revised in an iterative (i.e. back and forth) process that is typical of a qualitative analysis. When the list of variables was thus generated, the analysis displayed the relevant evidence for both groups in adjacent columns (see tables 4.19 for the hypotheses variables and 4.20 for the additional variables). On the basis of this, additional conclusions on effect predictors became possible.

Table 4.19
Qualitative comparison on the hypothesis variables between the 4 least and the 4 most effective studies

Factor	4 least effective studies	4 most effective studies	
Working hours	#85 Not stated #20 After working hours #49 Within working hours #6 Not stated	#145a #134 #157a #157b	After working hours Not stated After working hours After working hours
Location	#85 On site #20 Not stated #49 On site #6 Not stated	#145a #134 #157a #157b	On site Not stated Not stated Not stated
Duration in weeks	#85 5 weeks #20 12 weeks #49 6 weeks #6 4 weeks	#145a #134 #157a #157b	5 weeks 4 weeks 4 weeks 4 weeks
Risk status	#85 high #20 high #49 high #6 high	#145a #134 #157a #157b	high high high high
Micro/meso level	#85 micro #20 mixed #49 micro #6 micro	#145a #134 #157a #157b	mixed micro micro micro
Used method	#85 COG ²⁶ #20 MRE + COG	#85 #134	COG MRE + OTH

²⁶ The meaning of the abbreviations is the same as in chapter 4.2.1, tables 4.3 – 4.6.

	#49 OTH #6 MED	#157a MRE + COG #157b OTH
Study design	#85 6 #20 6 #49 7 #6 5	#145a 8 #134 6 #157a 8 #157b 8
Publication year	#85 1994 #20 1997 #49 1996 #6 1999	#145a 1986 #134 1988 #157a 1985 #157b 1985

As can be concluded from table 4.19, the results from the qualitative analysis largely resemble those of the quantitative one. For example, the majority (3 of 4) of the highly effective studies were carried out after working hours, against only one of the 4 least effective interventions. This corresponds with the significant effect that was found in the statistical analysis. Moreover, 2 of 4 little effective interventions were performed on site against only one highly effective study, which again confirms the results of the quantitative analysis. For the variable "duration", the ineffective studies lasted 6.75 weeks on average, whereas the effective ones were finished within a mean 4.33²⁷ weeks. Against a median of 6 and a Standard Deviation of 3.54 for the overall sample of individual interventions, this is not a difference to call for a further investigation. Further, there was no difference between the 2 groups on the variables "risk status" (all high) and "system level" (3 micro and 1 mixed in both groups). This again reflects the findings from the statistical analysis, where no influence of these factors was found either. Also in accordance with quantitative findings, the used method did not make a difference. Quite on the contrary, the very effective and very ineffective studies used a remarkable similar pattern of methodologies. Finally, the same effects as in the quantitative analyses were found for the factors "design" and "publication year". Well-designed and relatively early studies were the most effective.

Table 4.20
Qualitative comparison on additional variables between the 4 least and the 4 most effective studies

Factor	4 least effective studies	4 most effective studies
Use of specific distal goals	#85 Emotional exhaustion, depression	#145a Stress
	#20 Coping, mood, ill health	#134 Stress
	#49 Anxiety	#157a Anxiety
	#6 Stress, anxiety, self efficacy beliefs	#157b Anxiety

²⁷ When averages were involved, study #157a and #157b were treated as one to avoid overrepresentation.

Profession or situation specific	<p>#85 No</p> <p>#20 No</p> <p>#49 No</p> <p>#6 Partly (teacher stress and self efficacy but general anxiety)</p>	<p>#145a Yes (teacher stress)</p> <p>#134 No</p> <p>#157a Partly (both teacher and general anxiety)</p> <p>#157b Partly (both teacher and general anxiety)</p>
Education and experience of program provider	<p>#85 Doctoral student</p> <p>#20 Experienced psychologist and sociologist</p> <p>#49 Students</p> <p>#6 Not stated</p>	<p>#145a Postgraduate consultants</p> <p>#134 Not stated.</p> <p>#157a 3 school psychologists</p> <p>#157b 3 school psychologists</p>
Self assessed stress level	<p>#85 Not stated</p> <p>#20 Not stated</p> <p>#49 Unclear (only stated that participants scored above average on pre-test anxiety questionnaire)</p> <p>#6 Not stated</p>	<p>#145a Participants rated their jobs as stressful (specials pre-test question)</p> <p>#134 Participants sought help</p> <p>#157a Yes (only those were invented that experienced problems)</p> <p>#157b Yes (only those were invented that experienced problems)</p>
Audiovisual media	<p>#85 No</p> <p>#20 No</p> <p>#49 No</p> <p>#6 No</p>	<p>#145a No</p> <p>#134 Audiocassette</p> <p>#157a Videotapes</p> <p>#157b Videotapes</p>
Skill goal	<p>#85 General skills</p> <p>#20 General skills + Cognitive restructuring</p> <p>#49 None</p> <p>#6 RISE response²⁸</p>	<p>#145a Collaborative behavioural consultation technique</p> <p>#134 Ego enhancing technique</p> <p>#157a General skills</p> <p>#157b Classroom management skills</p>

²⁸ A simple meditation technique using sound as a focussing device and three corollary techniques (Winzelberg and Luskin, 1999).

Modelling	#85	No	#145a	Yes (modelling of consultants)
	#20	No	#134	No
	#49	No	#157a	Yes (modelling of videotape)
	#6	No	#157b	Yes (modelling of videotape)

As can be seen in the table above (4.20), a qualitative inspection of patterns and regularities yielded a number of interesting new variables that appeared effect-related. Below, a short description of these patterns shall be given. Firstly, it is an interesting finding that most of the ineffective studies addressed multiple distal goals (in 3 of 4 cases), whereas all the effective interventions were focussed on a single outcome. Moreover, the current results provide a slight indication that positive outliers were more focussed on profession- and situation-specific goals, whereas this was much less the case in little effective studies. Furthermore, differences between the two groups emerged when the experience of the program providers was compared. Whereas little effective studies were often lead by students (2 of 4), the majority of the effective ones was carried out by graduated experts. Another marked difference arose for the factor of the self-assessed state of the participants. In all of the very effective studies, participants rated *themselves* as highly stressed whereas this was only the case in one ineffective intervention. Also, no ineffective intervention made use of audiovisual media, whereas 3 of 4 effective interventions did. Regarding the use of modelling as an intervention technique, the same picture emerged, with only one effective study that did not use this technique. Finally, the last difference pertains to the proximal goals of the intervention. Here, 3 of 4 positive outliers used very specific skills (e.g. a certain ego-enhancement procedure). Of the ineffective interventions, however, only one study targeted such well-described skills.

Chapter 5: Discussion

In the previous chapter, the results of the statistical and qualitative analyses were presented. In this chapter, these results shall be linked back to the research questions. Moreover, the implications of the study's conclusions for the theory on stress and prevention shall be discussed. After this, the discussion shall focus on limitations and strengths of the current research design. Finally, at the end of the chapter, general conclusions and recommendations for future research and intervention practice shall be given.

5.1 Overview of results

5.1.1 Research question A: Magnitude of effect size

In the previous chapter, three research questions were addressed. In research question A, the current essay sought to determine the average effect size of stress intervention programs. It was found that this varied over different outcome variables. At post-test, stress interventions had a medium effect on psychological variables. A high fail-safe N^{29} showed that these findings were relatively robust. Small effects were found for behavioural studies, whereas cognitive, physiological and organisational measures only reached very small to zero effects. For those studies that also included a follow-up, psychological, cognitive and behavioural effect sizes slightly decreased at the second measuring point. For psychological effect sizes, these follow-up results can be characterised as "very effective". However, these follow-up studies already had higher psychological values at post-test. Of the other follow-up outcome measures, physiological and organisational effect sizes increased slightly. Overall, the conclusion of these findings is that the current sample is most effective on psychological values, less effective on cognitive and behavioural outcomes and least successful on physiological and organisational variables. Effect sizes are reasonable well preserved over time, although most outcome measures do suffer slight decreases.

5.1.2 Research question B: Homogeneity analysis

In the analysis of homogeneity, it was shown that results were significantly heterogeneous. This heterogeneity was caused by 5 studies, of which 4 represented extremely effective interventions. These were two cognitive-behavioural interventions (SIT and classroom management) by Sharp and Forman (1985), an ego-enhancement intervention by Stanton (1988) and a so-called collaborate behavioural consultation (CBC) program by Tunnecliffe et al. (1986). The fifth study (Freedy and Hobfoll, 1994) was characterised by very poor results. The next question, then, was what there should be done with these 5 programs and their results. Because the heterogeneity was only caused by a relatively small group of studies, it was concluded that the present sample is best treated as homogeneous with a small group of outliers. It was chosen not to exclude these 5 outliers because this would ignore an interesting source of variation. To give the outlier studies extra attention, however, an additional qualitative analysis of the differences between very effective and very ineffective interventions was performed.

5.1.3 Research question C: Decisions on research hypotheses

For the third research question, several hypotheses were addressed. Using a number of statistical analyses, the goal was to determine factors that have an influence on study effect size. Here, the present essay found the following results. First of all, there were some results for which there had been no specific hypotheses. This was the case for the variable "publication year", for which a negative, statistically significant, relationship with study effectiveness was found. Secondly, there were many variables that had been expected to be effect-related, but for which no statistical difference was found. Although there were trends in the predicted direction, it did not make a statistical difference (1) whether or not the program was directed at high-risk groups, (2) which method it used, (3) how many weeks the program lasted or the number of sessions that were held and (4) whether or not the intervention used multiple methods or just a single one. Moreover, (5) compared to individual interventions, organisational ones did not increase in effect from post-test to follow-up. Thirdly, there were variables for which the essay's findings pointed in the opposite direction of the hypotheses. This was the case for the outcome

²⁹ 73.1 against a criterion value of 0.20 (small) effect size.

that interventions that are run off site are more effective. Moreover, in contrast to the study's hypotheses, individual-oriented programs were found to be statistically more effective than organisational interventions. As a last category, there were hypotheses that were confirmed by the present data. As predicted, programs that were carried out after working hours were significantly more effective than interventions during working hours. Also, the quality of study design was found to be related to effect.

In the qualitative analysis of the differences between very effective and very ineffective studies, the outcomes of the statistical analysis were largely confirmed. Besides the research hypotheses, however, the analysis also pointed at a number of additional effect-related factors. Here, the first of these extra variables was that very effective interventions were distinguished by a focus on only one type of outcome category. In three of the four very effective studies, these targeted variables were (at least partly) profession or context specific. With respect to their target group, the participants in all four effective studies regarded *themselves* as stressed, which was at most the case in only one ineffective intervention. Furthermore, in three of the positive outliers, the intervention was led by graduated experts. Because many ineffective studies had "only" students as providers, this represents a large difference in experience. Marked differences were also found in the fact that all effective studies used modelling and audiovisual media in their methods, which was not the case for any of the less effective studies. Finally, very effective studies were more often characterised by specific, well-described skills, whereas the goals of ineffective studies were much more broad and general.

5.2 Implications for prevention theory

5.2.1 Earlier effect reviews

Murphy's 1996 review

Murphy (1996) performed a review on 101 stress interventions that were published in 64 articles. Like in the present essay, these articles were identified through computerised search techniques and prior reviews, and also through contacts with stress experts. In his results, the following picture emerged. For biological/somatic and organizational measures, no large effects were found. This finding is replicated in the present study, which likewise found only small effectiveness in these areas. For psychological outcomes, Murphy observed greater effect sizes, especially on measures of anxiety. This is also paralleled by the current research, although the present results do not differentiate between different types of psychological outcomes. The current category of behavioural outcomes was not covered in Murphy's analysis. Thus, overall, the conclusion can be drawn that the current research confirms the results of Murphy.

On another level, however, comparisons result in a somewhat more equivocal picture. Whereas Murphy found differential effects of his intervention methods, there were no statistical differences in the current sample. Inspection of trends showed that some results confirmed Murphy's findings, whereas others did not, or were even in the opposite direction. Like in Murphy's study, biofeedback was only effective on physiological outcomes, and cognitive-behavioural interventions were most effective on psychological variables. Unlike Murphy's results, however, the latter method was not successful in affecting organisational measures. Another contradiction was that combinations of methods were effective on all outcomes in Murphy's sample, but negatively effective on organisational and physiological measures in the current study³⁰. Furthermore, for this "other" category, Murphy found positive results for all outcomes except for physiological measures. In the present sample, this category was also ineffective on organisational methods. Finally, Murphy evaluated meditation as very effective, whereas in the present essay, its results were negligible or even negative. Thus, when comparing the different methods on their relative effectiveness, the outcomes of the present study were only partly similar to Murphy's results.

In appraising the conclusions from the above-described comparison, there are a number of factors that can shed light on the differences, but also call for caution in endorsing similarities too easily. First of all, Murphy ran his analysis on a different kind of sample because studies that did not include a control group were also included³¹. Secondly, Murphy also used a different analysis. In assessing effect, he

³⁰ Not shown in the results chapter. For organisational outcomes, an effect size of -0.258 was obtained and for physiological measures an effect size of -0.095 .

³¹ This partly explains his larger sample of interventions.

counted the number of studies with positive and significant results. To determine the effect of stress interventions on an outcome category, he calculated the percentage of such significantly effective studies. The present analysis, on the other hand, calculated effect on the basis of control and experimental group differences. There are two consequences of this different methodology. In the first place, "significance counting" treats every significant effect as a positive result but does not consider the size of this effect. However, especially in very large samples, even small differences can account for significant results. This means that a significant effect does not automatically mean that a program has exerted a substantial and real influence. Furthermore, the second consequence of Murphy's way of calculation is that negative effects are not taken into account. In some cases in the present study, such harmful results were found. This has certainly influenced the outcomes of the effect size calculations.

Bamberg and Busch' 1996 meta-analysis

Another source of evaluating the current results comes from the Bamberg and Busch (1996) meta-analysis. Unlike Murphy, their review uses the same statistical procedures as the present essay, which makes comparison less problematic. In their study, they analysed the effect of 27 work-based "cognitive-behavioural" interventions on a number of outcome variables. Two of these outcomes are especially interesting because they were also included in the present analysis. These outcomes were psychological stress symptoms, for which they found an effect size of 0.56, and coping cognitions, for which an effect of 0.32 was calculated. When trying to compare these results to the present study, however, the problem of a proper point of reference arises. Because they also included 2 biofeedback interventions, 9 relaxation trainings, 3 fitness programs and 2 social support groups, their sample of "cognitive-behavioural" studies is much more heterogeneous than in the current category. Because Bamberg and Busch' sample practically included all individual methods from the current coding system, it seems most reasonable to compare their results with the worker-oriented category of the present study. Here, Bamberg and Busch found a psychological effect size of 0.56, which is very similar to the 0.51 that was achieved in the present analysis. Some differences arise for cognitive coping outcomes, however. For this category, the present analysis found a smaller effect size (0.11) than Bamberg and Busch (0.32). However, this discrepancy must be seen in relation to the small sample size of this category (N = 7), which makes effect size fluctuations more probable. Furthermore, the present sample of individual interventions counted two meditation studies, a method that Bamberg and Busch did not include in their analysis. Therefore, sample differences might have played a role. When meditation is left out, the cognitive effect size adopts a value of 0.23, a value that is more similar to the Bamberg and Busch results.

5.2.2 Implications for effect hypotheses

High-risk

The hypothesis that high-risk programs would be more effective was based on solid reasoning and predictions from many authors (e.g. Bunce, 1997). Moreover, in the Murphy review (1996), 94% of the high-risk programs achieved some significant results against only 79% of universal preventive interventions. Despite of this, predictions were not confirmed. There are a number of potentially distorting factors to explain this effect. Important is of course the above-described difference between the review techniques of Murphy and the present analysis. Besides this, some important factors point in the direction of validity considerations. In the present study, it was difficult to code the "risk status" variable in a reliable way. In the present coding system, programs were assessed as "high-risk" when (1) their target group was characterised by important risk factors or (2) when they were experiencing an elevated symptomatology level. As was shown in chapter 1, modern stress theories such as the IS-model state that stressors only lead to adverse outcomes when a person appraises these factors as threatening. Hence, the availability of a risk factor or stress symptom may not be enough to conclude that a person has an elevated risk for a stress-related condition. Moreover, in many cases the question of high versus normal risk was answered by other parties than the workers themselves (e.g. management or the study's researchers). The workers themselves, however, might not have shared this impression, which makes any further use of the term "stress" problematic. Thus, until more clarity is provided, it cannot be ruled out that an otherwise conceptualised "high-risk" variable might still show a relation with effectiveness. In line with this, the qualitative analysis suggested that programs are more effective when they are targeted at a population that regards *itself* as stressed. Because of this, future research should examine the role of subjective appraisal in the working mechanisms of stress interventions.

Dose-effect relation

Another effect hypothesis that was not statistically confirmed was the existence of a positive dose-effect relationship. In other words, the current stress interventions did not become more effective the longer they were carried out. On the contrary, a negative relationship between "duration" and intervention effectiveness was found. Because it was possible to code this variable quite unambiguously, an explanation shall have to reach beyond methodological considerations. Apparently, the rationale behind the hypothesis was not adequate and needs to be revised. In this rationale, two assumptions played a central role. The first one is that stress interventions would share some similarity with time-limited psychotherapy, where such dose-effect relationships have often been found (e.g. Barkham et al., 1996). Apparently, this association is problematic. Unlike stress interventions, psychotherapeutic sessions are mostly carried out on an individual basis (i.e. involving only therapist and client), require a greater quality of the therapeutic relationship, involves more serious problems (e.g. suicidal thoughts), etc. These differences may have produced the distinct outcome patterns that were observed in this study. The second assumption was that programs would behave according to the so-called skill frame of reference (Carlson and Hoyle, 1993). Because many stress interventions involve the learning of skills, a longer time to practice and rehearse would lead to better results. Because such an effect was not found, this rationale shall have to be re-examined. Central in a possible explanation are so-called non-specific factors. Many authors have already pointed to the fact that such factors (e.g. increased attention from the company and researchers) play a central role in the working mechanisms of stress interventions (Bunce and West 1996; Bunce, 1997). Because these factors do not so much depend on the time a program is carried out, extra sessions or weeks might not further increase effectiveness. Thus, because the two assumptions can be challenged, the present findings call for a critical re-examination of the skill frame of reference. For example, future research shall have to address the possibility of non-specific factors.

Learning principles

In the rationale behind the "location" and "working hours" hypotheses, aspects of learning theory played a central role. The present results provide a mixed picture, however. When applied to individual programs, the so-called "Yerkes-Dodson Law" correctly predicted that programs run after working hours are more successful. However, the generalisation principle, which would predict that programs that are carried out on site are more effective, was not mirrored in the present results. Contrary to predictions, off site interventions were more successful. Because these conflicting results even reached marginal significance, the assumptions behind the hypothesis have to be questioned. One possibility is that off site locations are more likely to be optimally designed for an intervention situation, and provide better materials and facilities. Moreover, there will more likely be an "intervention atmosphere". Because most people associate training resorts and universities with professionalism, participants' confidence and motivation are increased and, as stated earlier, motivation is a possible determinant of program success. Furthermore, through a process of cognitive dissonance, motivation might also be increased by the effort of travelling to the off site intervention location. Alternatively, it might provide additional barriers for participation so that only highly motivated workers are willing to participate. Besides increased motivation, the second and perhaps most important reason might well be that off site interventions provide the worker with a "safe distance" from the stress in the workplace. This safe distance could promote experimenting with new behaviour patterns. Furthermore, away from job demands, workers could feel more relaxed and intrinsically motivated. Off site programs could therefore be more successful. This would not only explain the findings of the "location" hypothesis, but can also be applied to the "workings hours" hypothesis. After all, interventions were also more effective when they are implemented at a safe distance from the working hours. To test this explanation, however, further research is necessary.

System level

Another result that was in the opposite direction of the research hypothesis is the finding that individual programs were more effective than organisational ones. Because this result even reached marginal significance, it is worth questioning the rationale behind the hypothesis' assumptions. Here, a comparison with a study of Reynolds (1997) can be helpful. In her research, Reynolds contrasted an individual counselling intervention with a set of organisational measures that were aimed at increased workplace control. Results were in the same direction as in the present analysis: individual interventions were more effective. In a discussion of her findings, Reynolds pointed at a number of factors that are relevant for the present study as well. For a successful implementation of organisational, she states,

effective and efficient means to change the aetiological determinants of a certain problem must be available. There exist, however, very few methods that have been unequivocally proven to be effective in changing organisational stressors. Moreover, such methods are faced with all kinds of organisational constraints, such as managerial attitudes and a low level of resources. When they would not result in structural improvements, structural measures could even produce harmful side-effects, as they raise employee expectations of better working conditions without satisfactorily fulfilling these claims. The resulting frustration would wipe out any possible effect. Thus, because methods have not yet been reliably established, organisational interventions might fail to result in much effect. Individual programs, on the other hand, can rely on well-established, evidence-based methods with a long tradition of use, and an acceptance by staff and workers. Moreover, they can be easily implemented while leaving the structure of the organisation intact. Therefore, although organisational programs might be better in ethical and moral terms, individual programs appear to be better in terms of efficiency.

Span of change

As discussed above, the hypothesis that high-risk programs are more effective was not confirmed by the present results. Also, as can be read back in the results, it was not confirmed that multiple method interventions are more successful. In both hypotheses, the span of change concept played a major role. The fact that both hypotheses are disconfirmed asks for a critical re-evaluation. For the variable "risk status", it has already been pointed out above that its conceptualisation might have been problematic. Secondly, there also might have been conceptualisation problems for the "multi-method" factor. Possible indications for this are provided by the qualitative results. Here, it appeared that programs with a distinct and well-outlined methodology were more effective. Moreover, programs that appeared as positive outliers were mostly aimed at a single stress management skill. Thus, successful interventions had a more one-dimensional methodology and might therefore be easier to grasp for participants. Moreover, such interventions could be more able to concentrate the available time and resources on just a limited amount of factors, whereas spreading the intervention over multiple areas might give rise to a "too much of too little" situation. Because, in such cases, none of the risk factors is *really* targeted, the span of change might actually become smaller. When this is true, the predictions of the span of change concept would still be valid. Therefore, although it is true that the results from the analysis disconfirmed the span of change hypotheses, this can be explained in a number of other ways. Because it is possible that other conceptualisations would have led to different results, a full rejection of the span of change concept seems premature.

Publication year

Regarding the time of publication, the study reached the conclusion that study effectiveness declined with the year in which it was published. This is a very intriguing finding that was not predicted by any hypothesis. Moreover, it is difficult to interpret because there are so many possible explanations. One of them is that the "publication" effect results from an overrepresentation of the positive outliers in the later half of the 1980s. However, results show that excluding these outliers from the sample would still result in a significant influence ($F(1) = 6.63, p = .01$)³². What then, is the working factor here? There are a number of other factors that might provide an explanation. Firstly, it is possible that the effect of a higher publication year can be explained by its correlation with other factors. This explanation was tested by calculating η^2 and F statistics for the relationship between year of publication and all other factors for which a (marginally) significant effect on psychological effect size has been shown (i.e. "location", "design", "working hours" and "system level"). As can be seen in the table below, none of these factors showed a significant relation. It is therefore not unlikely that an explanation must involve the characteristics of the variable itself. Such a negative time effect is counter-intuitive, as it is more readily expected that scientific progress and improved experience would have lead to a greater effectiveness. Therefore, possible explanations might have to be searched in context characteristics such as changing societal developments and opinions. As Tim Newton has eloquently outlined in his 1995 book *"Managing" stress: emotion and power at work*, the stress concept has experienced a striking popularisation. Moreover, advice columns, stress gurus and "popular" psychology have contributed to the dissemination of all kinds of "remedies" against stress, such as positive thinking and relaxation. Finally, providing stress management interventions has become a little industry in itself, with many companies fighting over a piece of the very lucrative market (Houtman and Kompier, 1995). The effect of these developments is very difficult to assess but, as is indicated by the present results, it is possible that some sort of

³² This analysis was not published in the results section.

habituation effect has taken place. Because knowledge of stress is already widely spread, lectures might not be as enlightening as they were in earlier days. Moreover, some stress management skills might already have adopted by a large proportion of the public. As a result, stress interventions bring less new information and can therefore achieve less additional effect. As stated above, however, final conclusions cannot be easily drawn and a more scientific explanation of the present finding shall have to await future research.

Table 5.1
Overview of statistical relations between publication year and other effect predictors

Variable	Location	Working hours	Research design	System level
Statistic	$f^2 = 9.22$	$f^2 = 12.59$	$r = .00$	$f^2 = 17.48$
<i>p</i> -value	$p = .60$	$p = .32$	$p = .99$	$p = .99$

Research design

The finding that a study's research design has an impact on effect size is not new to the present study. Surprisingly, however is the nature of this relation: whereas in most other meta-analyses, such a relation is in the negative direction (Wolf, 1986), better-designed studies in the present essay were more effective. Before trying to explain these findings, it is necessary to inspect the properties of the "design" variable. Although, theoretically, the lowest possible value was a score of zero, this minimum was never achieved: no study reached values lower than 4. The maximum value was often reached, however. An inspection of the individual items can explain this phenomenon. For two "design" items, a positive score was certain beforehand because the present inclusion criteria had already excluded studies that failed to comply with their methodological demands. These two items were: "findings for post-test" and "defined aims". Furthermore, the variables "findings for pre-test" and "randomly allocated control" were also almost always scored with "yes". Therefore, fluctuations in "design" scores must have come from the following 5 items: (1) "definition of intervention", (2) "equivalent comparison group", (3) "number of subjects in trial", (4) "reported attrition rates" and (5) "findings for all outcomes".

With these variable properties in mind, there are a number of potential mechanisms that can play a role in explaining the current finding. Firstly, it could well be that a study's design resembles the amount of work a program provider has invested in the intervention. This is possible because a well-constructed study does not only include a good methodological design, but also well-devised materials, good preparations and background knowledge, etc. Secondly, an explanation of this finding comes from experiences with varying presentation styles across different research outcomes. From reading the present articles, the impression arose that studies that did not reach positive and significant results often reported these findings in an incomplete way. For example, they would only concentrate on the limited amount of significant results or on trends in the expected direction (e.g. by bar graphs, percentages, etc.). This would result in a "no" on the items "findings for all outcomes" and "findings for post-test" and might therefore explain part of the "design" effect.

5.3 Limitations of the study

In interpreting the results from the present study, there are a number of limitations that need to be kept in mind. Both in the data collection and in the analysis procedure, there existed factors for which a biasing effect cannot be completely ruled out. These factors were: (1) ambiguities in the coding procedure, (2) the fact that the study used a relatively small sample size, (3) the decision to keep the outliers in the analysis and (4) the validity of the qualitative analysis. In the section below, these different problems shall be explained and elaborated in more detail.

5.3.1 Ambiguities in the coding procedure

Whereas for coding some factors, such as "year" and "duration in weeks", coding was relatively easy, other factors invoked more serious problems. Many times, study information was vague or ambiguous. In

such cases, an objective coding procedure was greatly troubled and subjective interpretations were made necessary. For example, in some studies, the assignment to experimental groups was only partly random. In such cases, only the control or experimental group would be randomly allocated. Would such a study receive a yes or a no on the "design" item "randomly allocated groups"? Because cases such as this one required subjective decision, the danger of bias arises. The study tried as much as possible to evade these dangers by only coding a variable when a clear decision, based on the text of the article, was possible. Moreover, with the use of a definition and coding book, coding criteria were handled as uniform as possible. Finally, the present classification system was based on a widely used and often revised coding system that had already been successfully used for other studies. However, because the coding was only done by one researcher, the danger of bias still remained. Due to time constraints, double coding (i.e. having a second coder repeat the procedure and check for discrepancies) was not possible. In order to reach firmer conclusion, such measures would be necessary.

5.3.2 Small sample

A second limitation of the present research is the relatively small sample ($N = 47$). This made identifying small group differences very difficult. This was especially difficult when the analysis involved only a limited part of the entire sample (e.g. only individual methods). The problem was further sharpened by the fact that it was chosen to exclude studies that could not be unambiguously coded. The small sample that was the result size had very important consequences for the current results. For a large number of variables, the N was not large enough (i.e. around 15 or higher) to allow for a statistical analysis. The result was that for many research hypotheses only trends (inspection of means) could be discussed. Such trends might have reached significance if sample sizes had been larger. The result was that a number of interesting patterns needed to be relatively disregarded because they could not be provided with an adequate scientific foundation. This is regrettable, but unfortunately also indicative of the state of the field. There is still a great need for more evaluation studies.

5.3.3 Keeping the outliers in the quantitative analysis

It was stated earlier, the present study performed a homogeneity test in order to check whether the study was characterised by a coherent group of effect sizes. The outcome of this analysis was that there were five studies that significantly contributed to heterogeneous outcomes. Four of these studies were positive outliers, whereas the remaining one was characterised by very poor results. It was chosen to keep these outlier studies in the analysis. This was done in accordance with a number of other meta-analysts (e.g. Wolf, 1986) and was backed up by the following arguments. Firstly, study outcomes had been previously grouped in different categories, which had already greatly reduced the possibility that the analysis would be comparing "apples with oranges". Secondly, studies were coded on a number of variables that were hypothesised to be linked with effect. Because of this, effect size variations were not eschewed, since they provided an opportunity to test the study hypotheses. Thirdly, the outlier studies were further inspected using a qualitative analysis. This way, the reasons for their large effect size could be examined. Despite of these arguments, however, a certain amount of caution is still required. When it is true that the 4 positive outliers did belong to a conceptually different group, this would have contaminated the analysis. This is the more dangerous because extreme studies play a relatively large role in the calculations of a statistical analysis. Thus, although there were solid arguments for leaving the outliers in the analysis, the above-described problems call for some caution in interpreting the results.

5.3.4 Validity of the qualitative analysis

As stated above, a qualitative analysis was performed to investigate the difference between the 5 least and the 5 most effective studies. This analysis resulted in the conclusion that there are a number of additional factors that might be related to intervention success. However, these qualitative conclusions differ from the statistical results in terms of validity. Whereas decisions in a quantitative analysis are largely left over to calculations and significance values, the major research tool in a qualitative analysis is the researcher himself. This means, however, that conclusions are threatened by human bias. This does not mean that qualitative analysis is an inherently unreliable tool, influenced only by the whim of the scientist. On the contrary, good qualitative analysis is both systematic and internally disciplined, and there are numerous validity measures to safeguard the results from the above-described dangers

(Berkowitz, 1997). To achieve this, the analysis needs to be self-aware, reflective and honest so that third parties can "walk through" thought processes and assumptions. The present study invested much effort to comply with these prerequisites. Thus, the results of the current analysis are by no means "merely subjective". However, because qualitative analysis is and stays a human enterprise, results should be treated with the appropriate caution.

5.4 Recommendations for future theory and practice

At the end of the current essay, some recommendations for future theory and practice shall be given. These recommendations shall be discussed in the following three categories: (1) implications for future research, (2) implications for the stress and intervention theory and (3) implications for future practice. In the following section, these recommendations shall be elaborated in more detail.

5.4.1 Implications for future studies

Suggestions for future improvement of similar studies can be summarised under the following points: (1) a double coding procedure and (2) a larger study sample and (3) revision of some of the coding constructs.

Recoding of variables

Because of the above-described ambiguities in coding, it would greatly improve validity if these variables would be double-coded by other, independent, researchers. As stated earlier, such a double coding procedure would eliminate many sources of subjective bias and make results more reliable.

Greater sample

In order to reach more definite conclusion on some of the observed trends, future studies would need a larger sample so that differences can be more easily detected and statistically confirmed. Although accomplishing such greater statistical power is not an easy task, there are several possible solutions. Firstly, the search procedure can be intensified so that more effort is invested in the search and retrieval of articles. For example, surveys on stress experts could be performed to identify additional studies. Secondly, coding can be performed more "loose" and interpretative, so that less ambiguous study information will not need to be excluded from the analysis. For the research variables of the quantitative analysis, such more interpretative coding procedures were not used because this was thought to threaten validity too much. However, in other cases (e.g. when a second coder is present), such a measure might be justified to preserve possibly important studies. Thirdly, a way to achieve a greater N might be to include evaluations without a control group. As was described by Carlson and Schmidt (1999), a statistical analysis would be able to calculate an "average control group effect". Such a procedure would involve the subtraction of pre- and post-test scores for the control group and dividing these differences through the Standard Deviation. After this, control group effects would be averaged over all studies. This average control group effect could then be used to adjust the results of no-control group studies. The possibility of such a procedure, however, is not yet clear and must await future research.

Operationalisation of constructs

As became clear in the present analysis, some of the present constructs turned out to be insufficient to account for variations in program effectiveness. Most clearly, this was the case with the "at risk" variable. Programs with a target group that is characterised by a high number of risk factors or an elevated symptomatology were not more effective than other programs. However, the qualitative analysis provided an indication that risk status might still be an important effect predictor when it is considered as the subjective taxation that participants themselves have of their stress situation. In line with stress theory, such as Lazarus' stress and coping model, it is apparently more important how people appraise their situation than how this situation "objectively" is. Therefore, future conceptualisations of risk status might gain more predictive power when they are used in this more subjective, appraisal-like way.

5.4.2 Implications for stress and intervention theory

In order to further strengthen the theoretical foundations of current stress interventions, a number of recommendations can be made. In the following section, two of such recommendations shall be

described in more detail: (1) a continuation of research efforts and (2) a re-implementation of the current "model programs".

More research is needed

To begin, it is greatly needed that more research on stress interventions is performed. Firstly, such efforts will have to be directed on the stress process itself. The fact that there is still no commonly accepted theory or definition of stress greatly slows down theoretical progress. Moreover, many of the mechanisms in the stress process remain unclear. Future research shall have to contribute more advanced theories, whose predictions can be tested in laboratory and field experiments. Of special concern here is the creation of unifying theories and models. As stated in the American Psychological Association publication *Job stress interventions* (Murphy, Hurrell, Sauter and Keita, 1995), there is a great need for such integrative models. The IS-model that was introduced in the second chapter can serve as a very good starting point for such an enterprise. However, the IS-model needs much theoretical and practical follow-up work before it can reach its full potential as an integrative framework. In order to serve this purpose, many of its constructs and processes will need revisions and clarifications. Of special importance here is the role of physical processes. This important factor is not sufficiently accounted for in present models. Although much is known on the short-term bodily reactions to external stressors (e.g. the so-called general adaptation syndrome GAS), there is only scarce knowledge on the interaction between psychological and physiological factors in the development of chronic stress reactions (e.g. "wear-and-tear"). Because of this, full explanations of the working mechanisms of, say, exercise are still failing. To achieve this, an interdisciplinary cooperation between different sciences such as psychology, medicine and biology is needed.

Re-implementing model programs

In order to reach more solid conclusions on the "efficacy"³³ of present interventions, they will have to be replicated in other contexts and with other target groups. Of special concern here are the 4 outlier studies that showed extremely high effectiveness. These outlier intervention might represent "model programs" that should be applied in other settings as well. However, only when it is demonstrated that these programs are effective in a wider range of situations will it be possible to state that the success of these interventions was not only due to a talented provider or an extraordinarily motivated target group. Only then, there can be a question of real model programs.

5.4.3 Implications for future practice

On the basis of the current essay, a number of recommendations for future preventive programs can be given. There has to be distinguished, however, between the results of the qualitative and those of the quantitative analysis. Whereas the latter result are reasonably "hard", the former ones are based on human judgement and need therefore to be regarded with appropriate caution. With this in mind, program designers are recommended to take the present outcomes into account when devising new interventions. Firstly, on the basis of the qualitative results, it may be advisable that programs are devised along a clearly described, one-dimensional end goal. It can also be suggested to direct intervention efforts on profession- and situation-specific end goals instead of on more general stress outcomes. Moreover, the learning of a stress management skill might be the most effective option for an intervention target. For an adequate choice of target group, people that regard *themselves* as stressed appear the best candidates, whereas intervention providers may need to be people that are appropriately educated and experienced. Promising techniques are the use of video materials and modelling techniques. Secondly, on the basis of the quantitative results, it appears that preventive stress programs should be carried out at a "safe distance" from the stressing work place, meaning after working hours and off site. Furthermore, it seems more successful to aim such programs on changes in the individual worker than on alterations in the organisational context. Until scientific knowledge and practical experience have improved the methods for organisational change, attention for the individual worker and his body, thoughts and feelings is the most effective and practical way to relieve one of the most serious and pervasive problems of the new millennium: stress.

³³ Efficacy is the generalised effectiveness of a program over different contexts, providers and target groups.

References

- Bamberg, E. & Busch, C. (1996). Betriebliche Gesundheitsförderung durch Streßmanagementtraining: eine Metaanalyse (quasi-)experimenteller Studien. *Zeitschrift für Arbeits- u. Organisationspsychologie*, *40*, 127-137.
- Barkham, M., Rees, A., Stiles, W.B., Shapiro, D.A., Hardy, G.E. & Reynolds (1996). Dose-effect relations in time-limited psychotherapy for depression. *Journal of consulting and clinical psychology*, *64*, 927-935.
- Beehr, T.A. (1976). Perceived situational moderators of the relationship between subjective role ambiguity and role strain. *Journal of applied psychology*, *61*, 35-40.
- Berkowitz, S. (1997). Analyzing Qualitative Data. In J. Frechtling & L. Sharp-Westat (Eds.), *User-friendly handbook for mixed method evaluations* (chapter 4) [ONLINE]. Available: http://www.ehr.nsf.gov/EHR/REC/pubs/NSF97-153/CHAP_4.HTM [2001, October 25].
- Brehm, S.S. & Kassing, S.M. (1996). *Social psychology*. Boston/Toronto: Houghton Mifflin Company.
- Bunce, D. (1997). What factors are associated with the outcome of individual- focussed worksite stress management interventions? *Journal of occupational and organizational psychology*, *70*, 1-17.
- Bunce, D. & West, M. A. (1996). Stress management and innovation interventions at work. *Human-Relations*, *49*, 209-232.
- Caplan, R.D. & Jones, K.W. (1975). Effects of work load, role ambiguity, and Type A personality on anxiety, depression, and heart rate. *Journal of applied psychology*, *60*, 713-719.
- Carlson, R.C. & Hoyle, R.H. (1993). Efficacy of abbreviated progressive muscle relaxation training: a quantitative review of behavioral medicine research. *Journal of consulting and clinical psychology*, *61*, 1059-1067.
- Carlson, K.D. & Schmidt, F.L. (1999). Impact of experimental design on effect size: findings from the research literature on training. *Journal of applied psychology*, *84*, 851-862.
- Cooper, C.L. (1983). Identifying stressors at work: recent research developments. *Journal of psychosomatic research*, *27*, 369-376.
- Daniels, K. (1996). Why aren't managers concerned about occupational stress? *Work and stress*, *10*, 352-366.
- Dipoye, R.L., Smith, C.S. & Howell, W.C. (1994). *Understanding industrial and organizational psychology: an integrated approach*. Orlando, FL: Harcourt Brace College Publishers.
- Durlak, J.A. & Wells, A.M. (1997). Primary prevention mental health programs for children and adolescents: a meta-analytic review. *American journal of community psychology*, *25*, 115-152.
- Ganster, D.C., Fusilier, M.R. & Mayes, B.T. (1986). Role of social support in the experience of stress at work. *Journal of applied psychology*, *71*, 102-110.
- Hosman, C. (1998). *Inleiding in de preventieve gezondheidszorg*. Nijmegen: Research Groep Preventie en Psychopathologie KUN.
- Houtman, I.L.D. & Kompier, M.A.J. (1995). Courses on work stress: a growing market, but what about their quality? In Murphy, L. R. Hurrell, J. J., Sauter, S.L. & Keita, G.P. (Eds.), *Job stress interventions* (pp. 337-349). Washington DC: American Psychological Association.
- Kahn, R.L. & Byosiére, P. (1992). Stress in organizations. In Dunnette, M.D. & Hought, L.M. (Eds.), *Handbook of industrial and organizational psychology*, vol. 3 (pp. 571-650). Palo Alto: Consulting Psychologists Press.
- Kessler, R.C., Price, R.H. & Wortman, C.B. (1985). Social factors in psychopathology: stress, social support and coping processes. *Annual review of psychology*, *36*, 531-572.
- Kohler Moran, S., Wolff, S.C. & Green, J.E. (1995). Worker's compensation and occupational stress: gaining control. In Murphy, L. R. Hurrell, J. J., Sauter, S.L. & Keita, G.P. (Eds.), *Job stress interventions* (pp. 355-368). Washington DC: American Psychological Association.
- Kolbell, R.M. (1995). When relaxation is not enough. In Murphy, L. R. Hurrell, J. J., Sauter, S.L. & Keita, G.P. (Eds.), *Job stress interventions* (pp. 31-43). Washington DC: American Psychological Association.
- Kompier, M.A.J., Geurts, A.E., Gründemann, R.W.M., Vink, P. & Smulders, P.G.W. (1998). Cases in stress prevention: the success of a participative and stepwise approach. *Stress Medicine*, *14*, 155-168.

- Lazarus, R.S. (1995). Psychological stress in the workplace. In Crandall, R., Perrewe, P.L., et al. (Eds.), *Occupational stress: a handbook* (pp. 3-14). Philadelphia: Taylor & Francis.
- Lieberman, D.A. (1999). *Learning: behavior and cognition*. Belmont: Wadsworth/Thomson Learning.
- Lyons, L.C. (1998). *Meta-analysis: methods of accumulating results across research domains* [ONLINE]. Available: <http://www1.monumental.com/solomon/MetaAnalysis.html> [2001, October 25].
- Matteson, M.T. & Ivancevich, J.M. (1987). *Controlling work stress: effective human resource and management strategies*. San Francisco: Jossey-Bass.
- Murphy, L.R., Hurrell, J.J., Sauter, S.L. & Keita, G.P. (Eds.). (1995). *Job stress interventions*. Washington DC: American Psychological Association.
- Murphy, L.R. (1996). Stress management in work settings: a critical review of the health effects. *American Journal of Health Promotion*, 11, 112-135.
- Newton, T. (1995). *"Managing" stress: emotion and power at work*. London: Sage Publications.
- Ogden, J. (1996). *Health psychology: a textbook*. Buckingham: Open University Press.
- Ray, J.W. & Shadish, W.R. (1996). How interchangeable are different estimators of effect size? *Journal of consulting and clinical psychology*, 64, 1316-1325.
- Reynolds, S (1997). Psychological well-being at work: is prevention better than a cure? *Journal of psychosomatic research*, 43, 93-102.
- Taylor, M.J. & White, K.R. (1992). An evaluation of alternative methods for computing standardized mean difference effect size. *Journal of experimental education*, 61, 63-72.
- Van der Hek, H. & Plomp, H.H. (1997). Occupational stress management programs: a practical review of published effect studies. *Occupational medicine*, 47, 133-141.
- Veltman, N.E., Ruiters, M. & Hosman, C.M.H. (1996). *Stemming maken: versterken van weerbaarheid en verminderen van depressieve klachten bij jongeren*. Nijmegen: Research Groep Preventie en Psychopathologie KUN.
- Wolf, F.M. (1986). *Meta-analysis: quantitative methods for research synthesis*. Beverly Hills: Sage Publications.
- Zoike, E. (1999). *Krankheitsarten 1998-1999*. Essen: BKK Bundesverband.

List of included studies

- Aust, B., Peter, R. & Siegrist, J. (1997) [3/p³⁴/#20]. Stress management in bus drivers: A pilot study based on the model of effort reward imbalance. *International Journal of Stress Management*, 4, 297-305.
- Bertoch, M. R., Nielson, E. C., Curley, J. R. & Borg, W. R. (1989) [3/p/#133]. Reducing teacher stress. *Journal of Experimental Education*, 57, 117-128.
- Cecil, M. A. & Forman, S. G. (1990) [3/p/#126a 3/p/#126b]. Effects of stress inoculation training and coworker support groups on teachers' stress. *Journal of School Psychology*, 28, 105-118.
- Cooley, E. & Yovanoff, P. (1996) [3/p/#56]. Supporting professionals at risk: Evaluating interventions to reduce burnout and improve retention of special educators. *Exceptional Children*, 62, 336-355.
- Doctor, R.S., Curtis, D. & Isaacs, G. (1994) [3/Mu/#11]. Psychiatric morbidity in policemen and the effect of brief psychotherapeutic intervention – a pilot study. *Stress medicine*, 10, 151-157.
- Freedly, J. R. & Hobfoll, S. E. (1994) [3/p/#85]. Stress inoculation for reduction of burnout: A conservation of resources approach. *Anxiety, Stress and Coping: An International Journal*, 6, 311-325.
- Friedman, G. H., Lehrer, B. E. & Stevens, J. P. (1983) [3/Nw/#60]. The effectiveness of self directed and lecture/discussion stress management approaches and the locus of control of teachers. *American Educational Research Journal*, 20, 563-580.
- Ganster, D. C., Mayes, B. T., Sime, W. E. & Tharp, G. D. (1982) [3/p/#177]. Managing organizational stress: A field experiment. *Journal of Applied Psychology*, 67, 533-542.
- Gronningsaeter, H., Hytten, K., Skauli, G., Christensen, C.C. & Ursin, H. (1992) [3/p/#110a + 3/p/#110b]. Improved health and coping by physical exercise or cognitive behavioral stress management training in a work environment. *Psychology and Health*, 7, 147-163.
- Hall, E., Hall, C. & Abaci, R. (1997) [3/p/#24]. The effects of human relations training on reported teacher stress, pupil control ideology and locus of control. *British Journal of Educational Psychology*, 67, 483-496.
- Hammer, M. R. & Martin, J. N. (1992) [3/p/#108a + 3/p/#108b]. The effects of cross cultural training on American managers in a Japanese American joint venture. *Journal of Applied Communication Research*, 20, 161-182.
- Higgins, N. C. (1986) [3/p/#151a 3/p/#151b]. Occupational stress and working women: The effectiveness of two stress reduction programs. *Journal of Vocational Behavior*, 29, 66-78.
- Hinman, M., Ezzo, L., Hunt, D. & Mays, A. (1997) [3/p/#29]. Computerized exercise program does not affect stress levels of asymptomatic VDT users. *Journal of Occupational Rehabilitation*, 7, 45-51.
- Jackson, S. E. (1983) [3/p/#173]. Participation in decision making as a strategy for reducing job related strain. *Journal of Applied Psychology*, 68, 3-19.
- Kawakami, N., Araki, S., Kawashima, M., Masumoto, T. & Hayashi, T. (1997) [3/p/#36]. Effects of work related stress reduction on depressive symptoms among Japanese blue collar workers. *Scandinavian Journal of Work, Environment and Health*, 23, 54-59.
- Kolbell, R.M. (1995) [3/p/#63]. When relaxation is not enough. In Murphy, L. R. Hurrell, J. J., Sauter, S.L. & Keita, G.P. (Eds.), *Job stress interventions* (pp. 31-43). Washington DC: American Psychological Association.
- Landsbergis, P. A. & Vivona Vaughan, E. (1995) [3/p/#73a + 3/p/#73b]. Evaluation of an occupational stress intervention in a public agency. *Journal of Organizational Behavior*, 16, 29-48.
- Lee, S. & Crockett, M. S. (1994) [3/p/#82]. Effect of assertiveness training on levels of stress and assertiveness experienced by nurses in Taiwan, Republic of China. *Issues in Mental Health Nursing*, 15, 419-432.
- Lindquist, T. L. & Cooper, C. L. (1999) [3/p/#5]. Using lifestyle and coping to reduce job stress and improve health in "at risk" office workers. *Stress Medicine*, 15, 143-152.
- Malkinson, R., Kushnir, T. & Weisberg, E. (1997) [3/p/#19]. Stress management and burnout prevention in female blue collar workers: Theoretical and practical implications. *International Journal of Stress Management*, 4, 183-195.

³⁴ "p" = PsycLit, "mu" = Murphy's references, "E" = cross reference, "nw" = errata PsycLit, "me" = MedLine

- Mikkelsen A. & Saksvik P.O. (1999) [3/Me/#6a + 3/Me/#6b]. Impact of a participatory organizational intervention on job characteristics and job stress. *International Journal of Health Services*, 29, 871-93.
- Murphy, L. R. (1983) [3/p/#172a + 3/p/#172b]. A comparison of relaxation methods for reducing stress in nursing personnel. *Human Factors*, 25, 431-440.
- Norvell, N. & Belles, D. (1993) [3/p/#104]. Psychological and physical benefits of circuit weight training in law enforcement personnel. *Journal of Consulting and Clinical Psychology*, 61, 520-527.
- Orpen, C. (1984) [3/p/#158]. Managerial stress, relaxation and performance. *Journal of Management Development*, 3, 34-47.
- Pavett, C. M, Butler, M., Marciniak, E. J. & Hodgdon, J. A. (1987) [3/p/#139]. Exercise as a buffer against organizational stress. *Stress Medicine*, 3, 87-92.
- Payne, B.D. & Manning, B.H. (1990) [3/#E/191]. The effects of cognitive restructuring on preservice teacher's anxiety about teaching. *Contemporary educational psychology*, 15, 261-267.
- Peters, K. K. & Carlson, J. G. (1999) [3/p/#2]. Worksite stress management with high-risk maintenance workers: A controlled study. *International Journal of Stress Management*, 6, 21-44.
- Pronk, S. J., Pronk, N. P., Sisco, A., Schiller Ingalls, D. & Ochoa, C. (1995) [3/p/#74]. Impact of a daily 10 minute strength and flexibility program in a manufacturing plant. *American Journal of Health Promotion*, 9, 175-178.
- Pruitt, R.H. (1992) [3/Mu/#9]. Effectiveness and cost efficiency of interventions in health promotion. *Journal of advanced nursing*, 17, 926-932.
- Sharp, J. J. & Forman, S. G. (1985) [3/p/#157a + 3/p/#157b]. A comparison of two approaches to anxiety management for teachers. *Behavior Therapy*, 16, 370-383.
- Schaubroeck, J., Ganster, D.C., Sime, W.E. & Ditman, D. (1993) [3/E/#195]. A field experiment testing supervisory role clarification. *Personnel psychology*, 46, 1-25.
- Shulman, K. R. & Jones, G. E. (1996) [3/p/#49]. The effectiveness of massage therapy intervention on reducing anxiety in the workplace. *Journal of Applied Behavioral Science*, 32, 160-173.
- Stanton, H. E. (1988) [3/p/#134]. Relaxation, deepening and ego enhancement: A stress reduction "package." *Australian Psychologist*, 23, 315-322.
- Thomason, J. A. & Pond, S. B. III (1995) [3/p/#65a + 3/p/#65b]. Effects of instruction on stress management skills and self management skills among blue collar employees. In Murphy, L. R. Hurrell, J. J., Sauter, S.L. & Keita, G.P. (Eds.), *Job stress interventions* (pp. 7-20). Washington DC: American Psychological Association.
- Tunnecliffe, M. R., Leach, D. J. & Tunnecliffe, L. P. (1986) [3/p/#145a + 3/p/#145b]. Relative efficacy of using behavioral consultation as an approach to teacher stress management. *Journal of School Psychology*, 24, 123-131.
- Von Baeyer, C. L. & Krause, L. (1983) [3/p/#170]. Effectiveness of stress management training for nurses working in a burn treatment unit. *International Journal of Psychiatry in Medicine*, 13, 113-126.
- Winzelberg, A. J. & Luskin, F. M. (1999) [3/p/#6]. The effect of a meditation training in stress levels in secondary school teachers. *Stress Medicine*, 15, 69-77.

Appendix 1: List of used PsycLit search terms

Cluster 1: Key words relating to stress interventions

PsycLit nr.	Number of hits	Search term
1	2812	"Relaxation-" in DE ³⁵
2	1329	"Stress-Management" in DE
3	59	"Anxiety-Management" in DE
4	6162	"Behavior-Modification" in DE
5	6158	explode ³⁶ "Cognitive-Techniques"
6	511	"Preventive-Medicine" in DE
7	1337	"Primary-Mental-Health-Prevention" in DE
8	322	"Relapse-Prevention" in DE
9	3826	"Participation-" in DE
10	270	"Participative-Management" in DE
11	2660	explode "Biofeedback"
12	303	"Muscle-Relaxation" in DE
13	308	"Progressive-Relaxation-Therapy" in DE
14	851	"Meditation-" in DE
15	84	"Job-Enrichment" in DE
16	3073	"Occupational-Guidance" in DE
17	101	"Work-Rest-Cycles" in DE
18	75	"Working-Space" in DE
19	62	"Work-Week-Length" in DE
20	968	"Employee-Assistance-Programs" in DE
21	945	"Communication-Skills-Training" in DE
22	467	"Assertiveness-Training" in DE
23	244	"Human-Relations-Training" in DE
24	959	"Management-Training" in DE
25	105	"On-the-Job-Training" in DE
26	2406	"Personnel-Training" in DE
27	142	"Self-Instructional-Training" in DE
28	1770	"Social-Skills-Training" in DE
29	38	"Work-Adjustment-Training" in DE
30	3416	explode "Exercise"
31	12429	"Prevention-" in DE
121	40	"Microcounseling-" in DE
122	2240	"Group-Counseling" in DE
123	720	"Psychotherapeutic-Counseling" in DE
124	17158	"Counseling-" in DE
125	50934	"Education-" in DE
126	1298	explode "Inservice-Training"
127	2536	explode "Self-Help-Techniques"
128	1895	explode "Relaxation-Therapy"

Cluster 2: Key words relating to stress symptomatology

PsycLit nr.	Number of hits	Search term
-------------	----------------	-------------

³⁵ "in DE" signifies that the term was inserted using the key word thesaurus.

³⁶ "Explode" signifies that all lower order search terms are included. For example, stress exists of numerous subordinate classes of strain, which are represented by lower order terms such as physiological stress, occupational stress, etc.

33	11566	"Depression-Emotion" in DE
34	20802	"Anxiety-" in DE
35	183	"Work-Related-Illnesses" in DE
36	25447	explode "Stress"
37	4524	"Occupational-Stress" in DE
38	162	"Irritability-" in DE
39	1298	"Psychiatric-Symptoms" in DE
40	6287	"Self-Esteem" in DE
41	2049	"Well-Being" in DE
42	278	"Hardiness-" in DE
43	319	"Self-Confidence" in DE
44	11199	"Coping-Behavior" in DE
45	1329	"Stress-Management" in DE
46	54706	explode "Mental-Disorders"
47	6478	explode "Mental-Health"

Cluster 3: key words relating to worksite stress interventions

PsycLit nr.	Number of hits	Search term
49	183	"Work-Related-Illnesses" in DE
50	4524	"Occupational-Stress" in DE
51	4674	#49 or #50
52	1514	explode "Aerospace-Personnel"
53	6144	explode "Business-and-Industrial- Personnel"
54	106	"Disabled-Personnel" in DE
55	7224	explode "Educational-Personnel"
56	4916	explode "Government-Personnel"
57	6501	explode "Health-Personnel"
58	2546	explode "Law-Enforcement-Personnel"
59	1153	explode "Legal-Personnel"
60	6674	explode "Management-Personnel"
61	10064	explode "Medical-Personnel"
62	12031	explode "Mental-Health-Personnel"
63	481	explode "Nonprofessional-Personnel"
64	450	explode "Paramedical-Personnel"
65	735	explode "Religious-Personnel"
66	302	explode "Service-Personnel"
67	343	explode "Technical-Personnel"
68	848	explode "Volunteer-Personnel"

Cluster 4a: key words relating to evaluations

PsycLit nr.	Number of hits	Search term
76	12577	Evaluate
77	33326	evaluated
78	2411	evaluates
79	134936	examined
80	32385	examines
81	22615	examination
82	34410	determine
83	1783	determines
84	26947	determined
85	118286	test
86	52999	tests

87	51209	tested
88	59369	evaluation
89	698	looked at
90	1338	looks at
91	64185	describes
92	50367	described
93	15106	description

Cluster 4a: key words relating to effect

PsycLit nr.	Number of hits	Search term
95	177623	effects
96	36079	impact
97	28152	outcome
98	103201	effect
99	32166	effectiveness
100	26039	outcomes

Cluster 4c: key words relating to interventions

PsycLit nr.	Number of hits	Search term
71	58098	program
72	36053	intervention
73	67214	training
74	2308	workshop

Cluster 5: keywords relating to experimental design

PsycLit nr.	Number of hits	Search term
103	775	waiting list
104	470	reference group
105	12288	control group
106	40441	controls
107	1884	no treatment
108	2016	comparison group
109	2333	control condition
110	179	placebo condition
111	10599	placebo
112	1159	control conditions
113	153	placebo conditions
114	334	reference groups
115	4466	control groups
116	22825	experiment
117	74797	experimental

Appendix 2: Shortened version of the coding sheet

1. General information

Article number	3 (stress)	P / Mu / Me / Nw / E	#
Author(s)			
Year of publication			
Journal			
ISSN / ISBN			
Vol (nr)			
Pages			

2. Status

Working hours	Within	After	N.S.	P.
Location	On site	Off site	N.S.	P.

3. Effect size

Post-test

Follow-up

Time		
N		
Mean ES		
-Psychological		
-Cognitive mediators		
-Biological		
-Organizational		
-Behavioral		

4. Time

Number of sessions	
Duration of program	
Session time	

5. Risk status

Risk status	High-risk	Not stated
Specify		Page:

6. System level

System Level	Individual (micro)	Organizational (meso)
--------------	--------------------	-----------------------

7. Methods

	Individual focussed						Organisation focussed		
	MRE	BIO	EXC	MED	COB	OTH	SOC	CON	STR
Multi-level	YES						NO		
Multi-method	YES						NO		

8. Design

Research Design	1	2	3	4	5	6	7	8	9
-----------------	---	---	---	---	---	---	---	---	---

9. Protocol

Protocol	YES	NO/NOT STATED
----------	-----	---------------

10. Profession

--

Appendix 3: Operationalisation of “design” items

-Defined aims. This was scored with “yes” when the study specified what goals is tried to achieve.

-Defined interventions. Only those programs were scored with a “yes” that clearly defined both the nature of the intervention and the basic characteristics of the program provider.

-Randomly allocated control. This was coded with “yes” when participants were assigned at random to experimental and control group.

-Equivalent comparison group. Only those studies that analysed differences between experimental and control group and found them equivalent received a “yes”.

-Number of subjects in trial. Here, the number of subjects that were initially contacted needed to be reported also. The item was not coded with a “yes” when only the number of participants in the eventual experimental and control group was stated.

-Pre-intervention data. Coded with “yes” when pre-test data were reported.

-Post-intervention data. Coded with “yes” when post-test data were reported.

-Reported attrition rates. This was assigned a “yes” when the difference between initial enrolment and final analysis sample at post-test could be retrieved. Such attrition could either be explicitly reported or inferred from the reported Ns in the results section.

-Findings for all outcomes. When there were only reported findings for a selected number of variables (most often only those that were significantly effective), this item was scored with “no”.

Appendix 4: Description of outliers

Sharp and Forman (1985): A comparison of two approaches to anxiety management for teachers.

In this study, Jane Sharp and Susan Forman evaluated the effects of two anxiety management approaches for teachers. The intervention was provided by three school psychologists (one of them being Sharp). The participants were 60 experienced volunteer teachers from the same district, a southeast metropolitan area in the U.S.A. In this recruitment, only participants with negative emotional and physical stress reactions were invited to participate (high-risk). The program lasted 4 weeks and 8 sessions were carried out. Both interventions involved 2hrs group sessions of 10 participants each and were focussed on two very distinct outcome variables, namely self-reported and behaviourally-manifested *teacher* anxiety. The two interventions can be described in the following way. The first program was based on Meichenbaum's stress inoculation training (SIT) and some earlier work of Forman herself. The teachers first received some theoretical background of the program and were then trained in both relaxation and rational restructuring techniques (multi-method). They were assisted in developing their own *personal* set of coping skills. These skills were practiced with the use of videotape role-playing. The second program also started with a discussion of the theoretical background. After this, subjects identified student behaviour problems in their particular classrooms. They then received instructions in behaviour management techniques. As in the first intervention, these were rehearsed using videotape role-play.

Stanton (1988): Relaxation, deepening and ego-enhancement: A stress reduction "package."

The second study was designed and provided by Harry Stanton. The intervention had never before been carried out in a group setting. In the present study, however, he used it on a group of 20 nurses from different hospitals, who had sought help for their stress related problems (high-risk). All the participants were married and female and did at least have 2 years of nursing experience. Sessions were being held for a 4 week period with 1 session per week. The first meeting lasted some 50 minutes but in the three succeeding ones, this was reduced to 20 minutes. The technique Stanton used included elements of relaxation and guided visualisation. There were five stages: (1) physical relaxation with the help of breathing, (2) mental calmness, (3) disposing of "rubbish", (4) removal of barrier (negative forces and elements in one's life) and (5) enjoyment of a special place, where subjects are to fill themselves with "a stream of positive images". The last 4 stages all made use of visualisation techniques that bore close resemblance to guided hypnosis. This intervention was entirely provided by letting subjects listen to a standardised taped while they sat comfortably in a "pleasantly furnished room". The effect of the study was investigated on one single outcome measure: the stress profile. At the post measure, 68% of the nurses said they still practiced, which indicates a belief in the usefulness of the technique. It was reported that the five techniques could both be applied separately and as a "package".

Tunnecliffe, Leach and Tunnecliffe (1986): Relative efficacy of using behavioural consultation as an approach to teacher stress management.

The intervention was carried out by 2 outside consultants. Participants were an entire school staff of 7 teachers plus the principal. There were 2 males and 5 females and all teachers rated their profession as stressful (high-risk). It were the local schools themselves that had requested the intervention, which involved 5 weekly 1.5 hrs sessions that were held in the school and after working hours. The program instructed participants in guidelines for collaborative problem solving. This was first being modelled by the consultants and then left to the teachers themselves. In this learning process, there was minimal "expert" guidance, nor were the participants given any specific problem solution. Instead, the focus was on general problem solving. This was done to maximize the "collaborative" element of the intervention: together, the teachers would have to be able to use the SBC technique on an ongoing basis after the intervention had ended. The participants were contracted to keep using the technique in regular staff meetings, so that the intervention was institutionalised in the school environment (systemic change). The program was evaluated on one outcome measure: the teacher occupational stress factor questionnaire. High attendance rates were obtained (80-100%).