Advertising Effectiveness Measurement: Intermediate Constructs and Measures

Lars Bergkvist

AKADEMISK AVHANDLING

Som för avläggande av ekonomie
Doktorsexamen vid Handelshögskolan i
Stockholm
Framläggs för offentlig granskning
onsdagen den 31 maj 2000,
kl. 10.15 i Sal Ragnar Handelshögskolan,
Sveavägen 65

ADVERTISING EFFECTIVENESS MEASUREMENT: INTERMEDIATE CONSTRUCTS AND MEASURES

STOCKHOLM SCHOOL OF ECONOMICS
EFI, THE ECONOMIC RESEARCH INSTITUTE

EFI Mission

EFI, the Economic Research Institute at the Stockholm School of Economics, is a scientific institution which works independently of economic, political and sectional interests. It conducts theoretical and empirical research in management and economic sciences, including selected related disciplines. The Institute encourages and assists in the publication and distribution of its research findings and is also involved in the doctoral education at the Stockholm School of Economics.

EFI selects its projects based on the need for theoretical or practical development of a research domain, on methodological interests, and on the generality of a problem.

Research Organization

The research activities are organized in nineteen Research Centers within eight Research Areas. Center Directors are professors at the Stockholm School of Economics.

ORGANIZATION AND MANAGEMENT	
Management and Organisation; (A)	Prof Sven-Erik Sjöstrand
Center for Ethics and Economics; (CEE)	Adj Prof Hans de Geer
Public Management; (F)	Prof Nils Brunsson
Information Management; (I)	Prof Mats Lundeberg
Man and Organisation; (PMO)	Acting Prof Jan Löwstedt
Industrial Production; (T)	Prof Christer Karlsson
ECONOMIC PSYCHOLOGY	
Center for Risk Research; (CFR)	Prof Lennart Sjöberg
Economic Psychology; (P)	Prof Lennart Sjöberg
MARKETING	
Center for Information and Communication Research; (CIC)	Adj Prof Bertil Thorngren
Center for Consumer Marketing; (CCM)	Associate Prof Magnus Söderlund
Marketing, Distribution and Industrial Dynamics; (D)	Prof Lars-Gunnar Mattsson
ACCOUNTING, CONTROL AND CORPORATE FINANCE	
Accounting and Managerial Finance; (B)	Prof Lars Östman
Managerial Economics; (C)	Prof Peter Jennergren
FINANCE	
Finance; (FI)	Prof Clas Bergström
ECONOMICS	
Center for Health Economics; (CHE)	Prof Bengt Jönsson
International Economics and Geography; (IEG)	Prof Mats Lundahl
Economics; (S)	Prof Lars Bergman
ECONOMICS STATISTICS	
Economic Statistics; (ES)	Prof Anders Westlund
LAW	
Law; (RV)	Prof Erik Nerep

Chairman of the Board: Prof Sven-Erik Sjöstrand. *Director:* Associate Prof Bo Sellstedt,

Adress

EFI, Box 6501, S-113 83 Stockholm, Sweden • Internet: www.hhs.se/efi/
Telephone: +46(0)8-736 90 00 • Fax: +46(0)8-31 62 70 • E-mail efi@hhs.se

ADVERTISING EFFECTIVENESS MEASUREMENT: INTERMEDIATE CONSTRUCTS AND MEASURES

Lars Bergkvist

STOCKHOLM SCHOOL OF ECONOMICS
EFI, THE ECONOMIC RESEARCH INSTITUTE

Dissertation for the degree of Doctor of Philosophy, Ph.D.
Stockholm School of Economics 2000

ISBN 91 - 72 58 - 531 - 5

Keywords: Advertising effectiveness
Advertising research
Advertising pretests
Advertising posttests
Copy testing
Cognitive response measurement
Adjective checklists
Attitude toward the ad
Brand Beliefs

Cover layout: Håkan Solberg, Mediaproduktion AB
Cover photo: Lars Bergkvist
Printed by: Elanders Gotab, Stockholm, 2000
Distributed by: EFI, The Economic Research Institute
Stockholm School of Economics
P.O. Box 6501, SE - 113 83 Stockholm, Sweden

Acknowledgements

I started working on this thesis a little less than three years ago and during this time I have interacted with many people who have, in many different ways, contributed to it.

Without Mi, there would never have been a thesis. Your continuous support – from the wording of questions to moral encouragement when my self-confidence failed – has meant everything (not to mention what you mean to me in all other aspects of life). Thank you.

My supervisors' contributions have been many and invaluable. John Rossiter had an idea that resulted in the empirical studies. He has also taught me the meaning of a professional approach to academic research. For this, and his comments on research designs and draft versions of his thesis I am very grateful. Magnus Söderlund scrutinized numerous drafts, offered many valuable suggestions for improvement, and recommended literature that otherwise would have gone unexplored. Claes-Robert Julander encouraged me with questions.

A number of people were particularly important in helping the thesis to take shape. Anne commented on many a version of this thesis with consistent stringency. Jonas was my discussion partner in statistical as well as broad scientific matters. Per-Olov Edlund, Larry Percy, Richard Wahlund, Kelly Shaver and Ulf Essler generously shared with me their expertise. I am greatly indebted to all of you.

I am grateful for the financial support the research project has received from *Annonsörföreningen*, *Tidningsutgivarna*, *Sveriges Reklamförbund*, *Godkända Annonsförmedlares Förening* and the *Stockholm School of Economics*. I would

also like to thank the people in these organizations who, by participating in the reference group, also contributed intellectually to the project.

My fellow Ph.D. candidates at the Center for Consumer Marketing – Hanna, Fredrik, Micael, Helena, Jens, Per and Anna – participated in discussions about the design of the empirical studies, and other issues. I would also like to thank Sassa and Rebecca for their help with numerous practicalities.

My two-month visit at the Australian Graduate School of Management in Sydney was enriched by the wonderful people I met there. I am indebted to them, and to Fred and Mary, and Pat and Eugene, who made the non-studious side of life in Sydney very pleasant.

I am very grateful for the professional help I received on three other fronts: Anders Klapp at *Jefferson* did the artwork on the ads in the empirical studies; Maria De Liseo at *Ipsissima Verba* worked hard to weed out the worst of my abuse of the English language and to increase the readability of the text; the SSE library managed to find every article and book I needed, no matter how obscure.

Finally, I thank the students who participated in the empirical studies and Britt-Marie Östling for caring when nobody else did. I am also grateful to Jonas, Iggy Pop and Ulf.

Interaction with talented and intelligent people is, alas, not a guarantee against errors. Those that remain in this thesis should be attributed to the author.

Stockholm, April 18, 2000

Lars Bergkvist

Table of Contents

1 Introduction

1.1 Background and problem area

Advertising is a quick way to spend money. On January 31 1999 it was possible to spend $1.6 million in 30 seconds, which was the price for showing a 30-second commercial during the broadcast of the Super Bowl XXXIII (Grover, 1999). Advertising is also a way to spend large sums of money. The amount of money some large advertisers spend on advertising is impressive. For example, in 1998 Procter & Gamble spent more than $3,700 million on advertising (Procter & Gamble, 1998), while Unilever spent more than $5,500 million on advertising (Unilever, 1998). Another example is the luxury goods conglomerate LVMH, which spent an amount equal to 11 percent of their sales on advertising in 1998 (LVMH, 1998). Even in a small market such as Sweden, the sums of money spent on advertising are substantial. In 1998 the ten largest advertisers all had gross advertising media budgets in excess of $23 million (Info, 1999).

The question is whether the money spent on advertising is money well spent. Does increased advertising expenditure lead to increased sales or, for the advertiser, other desirable outcomes? Some attempts have been made to directly relate advertising to purchases or sales, mainly in studies based on field experiments in which the effects of other factors on sales have been controlled for. In one of these studies, one exposure to a newspaper ad was compared to no exposure at all. The experiment showed that within 30 hours of exposure, purchases in the group exposed to an ad was higher than in the group not exposed, albeit the effect was small (Bogart, Tolley and Orenstein, 1970). Interestingly, the study by Bogart et al. also showed that 3 of the 24 ads in the study caused a decline in purchases. A number of experiments reported in Blair (1987) showed that increased media weight, i.e., increased spending on

advertising, led to increased sales in 6 out of 20 cases. Similar results were reported from another study, in which it was found that less than half of the brands in a number of tests with increased media weight achieved a statistically significant increase in sales (Lodish, Abraham, Kalmenson, Livelsberger, Lubetkin, Richardson and Stevens, 1995).

The effect of advertising on sales can also be investigated by testing the effect of a reduction in advertising expenditures. Results in a study by Aaker and Carman (1982) show that a reduction in advertising expenditures does not automatically lead to a decrease in sales. They report that out of eleven experiments testing reduced advertising weight, ten showed that advertising could be reduced without a decrease in sales. Another study found that neither increasing nor decreasing advertising weight had much impact on sales, while changes in other variables, e.g., copy strategy or media mix, often led to quick, substantial sales increases (Eastlack and Rao, 1989).

Judging from these earlier studies, it would appear that advertising sometimes has a positive effect on sales, sometimes no effect at all, and sometimes even a negative effect. The uncertain outcome means that an advertiser cannot take for granted that an advertising campaign will work as intended or that an increase in advertising expenditure will lead to positive outcomes. This points to a need for systematic evaluation of the effectiveness of advertising, to find out how individual ads or advertising campaigns work.

One way to evaluate the effectiveness of advertising would be to relate it to sales or other desired behavioral outcomes. The problem with this approach is that it is inherently difficult to relate advertising to sales or other behavioral outcomes as sales are dependent not only on advertising, but also on the other variables in the marketing mix, e.g., product, price or distribution (cf. Kotler, 1997). Further,

the effects tend to be spread out over time (Lilien, Kotler and Moorthy, 1992). On an individual level, purchase decisions and actual purchases are influenced by a number of individual and environmental factors, e.g., culture, social class, personal influences and knowledge (cf. Howard, 1989; Engel, Blackwell and Miniard, 1995). The situation is similar for the relationship between advertising and other behavioral outcomes, e.g., inquiries, showroom visits or increased willingness to pay more for a product. These behaviors are in most cases also influenced by several factors other than advertising. Focusing only on behavioral variables in the evaluation of advertising effectiveness would also mean that other possible goals of advertising, e.g., building brand equity (cf. Aaker, 1992; Keller, 1993), would be excluded from the evaluation.

The uncertainty in the outcome of advertising campaigns in combination with the difficulties in directly relating advertising to behavioral outcomes points to the need for alternative measures of advertising effectiveness. These alternative measures tend to come in the form of various intermediate measures of advertising effectiveness, that have been developed over the years. Intermediate measures of advertising effectiveness are intermediate in the sense that they measure the mental effects, cognitive or affective, that are caused by advertising and that are assumed to precede desired behavioral outcomes such as purchase (cf., Vakratsas and Ambler, 1999).

Intermediate measures of advertising effectiveness are important not only to practitioners of advertising, but also to academics. In most research on how advertising works, irrespective of whether it comes in the form of laboratory experiments (e.g., Axelrod, 1963; Petty, Cacioppo and Schumann, 1983; Broach, Page and Wilson, 1997) or studies with real-market data (e.g., Barclay, Doub and McMurtrey, 1965; Rossiter, 1981; Batra, Lehman, Burke and Pae, 1995), the dependent variables tend to be intermediate measures. In these studies

it is assumed, explicitly or implicitly, that intermediate measures of advertising effectiveness are related to behavioral outcomes of advertising.

Many of the intermediate advertising effectiveness measures have a fairly long history. In a brief account of the origins of copy research, Lipstein (1984) refers to an example from 1924 in which *brand recall* is used as a measure of advertising effectiveness. Two other well-known measures can be traced almost as far back: *Ad recognition* was introduced in the 1930s and *day-after-recall* in the 1940s.

There appear to have been few, if any, radical changes in intermediate advertising effectiveness measures used in research in the last 30 years. The intermediate variables in research on advertising effectiveness measurement per se, or research on how various variables influence advertising effectiveness, tend to be chosen from a limited number of measures, most of which were developed before 1970. Studies that are in some way related to intermediate advertising effectiveness measures typically include two or three of the following measures: *ad awareness* (e.g., Zielske, 1959; Barclay et al., 1965; Bogart et al., 1970; Clancy and Kweskin, 1971; Brown, 1985; Naccarato and Neuendorf, 1998; Stapel, 1998), *brand awareness* (e.g., Plummer, 1972; Briggs and Hollis, 1997; Pieters and Bijmolt, 1997), *purchase intentions* (e.g. Wright, 1973; Kalwani and Silk, 1982; Batra et al., 1995), *pre-post persuasion* (e.g., Buzzell, Kolin and Murphy, 1965; Stewart and Furse, 1986; Blair, 1987; Laskey, Fox and Crask, 1995), *brand attitude* (e.g., DuBois, 1974; Brown, 1985; Briggs and Hollis, 1997), *attitude toward the ad* or *liking the ad* (e.g. Crane, 1964; Greyser, 1973; Gardner, 1985; Biel and Bridgwater, 1990), *adjective checklists* (e.g., Wells, 1964; Leavitt, 1970; Aaker and Bruzzone, 1981; Zinkhan and Burton, 1989; Hollis, 1995) or *cognitive response measurement* (e.g., Wright, 1973; Shavitt and Brock, 1990; Wansink, Ray and Batra, 1994).

Further, practitioners appear to use similar sets of intermediate measures. An overview of the measures used in the British *Advertising Effectiveness Awards* between 1980 and 1994 revealed the use of intermediate measures such as ad awareness, brand awareness and brand attitude (Broadbent, 1995). Similarly, textbooks on advertising and advertising effectiveness measurement indicate a lack of radical development, even if the textbooks differ somewhat with respect to the intermediate effectiveness measures they discuss (e.g., Lucas and Britt, 1963; Johnson, 1974; Aaker and Myers, 1982; Rossiter and Percy, 1997; Kumar, Aaker and Day, 1999).

The stability in intermediate measures of advertising effectiveness over the years does not necessarily indicate a lack of progress within the field of advertising effectiveness measurement. Instead, the focus appears to have been on issues other than the intermediate measures used. Two important developments, for example, have been *continuous tracking*, developed by the U.K. research company Millward Brown in the 1970s (e.g., Brown, 1985; Hollis, 1995; Broadbent, 1999), and the use of *single-source data* (e.g., Pedrick and Zufryden, 1991; Jones, 1995; Lodish et al., 1995; Schroeder, Richardson and Sankaralingam, 1997).

Given that intermediate measures of advertising effectiveness are of central importance to both practitioners and academics, it is hardly satisfactory that they have received such limited attention for many years.

1.2 Research problems

Research on intermediate measures of advertising effectiveness must deal both with *constructs*, which are theoretical concepts that are not directly observable,

and *measures*, which are observable empirical operationalizations of constructs (cf. Cronbach and Meehl, 1955; Peter, 1981). All measures are operationalizations of at least one (more or less explicit) underlying theoretical construct. This means that several measures of the same construct are possible and that they may differ in how well they reflect the underlying construct. It also means that scores on measures that measure more than one construct can be very difficult to interpret, since it is unclear which construct(s) is reflected in the score. Another consequence is that the relationship between measures is highly dependent on the relationship between the underlying constructs. This thesis, addresses three research problems pertaining to both the relationship between measures and constructs, i.e., the operationalization of constructs, and the relationship between the constructs themselves.

The first research problem relates to the relationship between constructs. In an overview of advertising research, Hansen (1998) notes that over the years has there been a search for one single measure of advertising effectiveness. In his view, this search is in vain, and he argues, therefore, that a multitude of measures should be used and that the relationships between variables should be modeled in the analysis of the data. In a similar vein, Mehta (1994) notes that a number of advertising effectiveness measures have been developed as predictors of sales, but that it is unclear how these relate to each other. Mehta also notes that the relationship between different diagnostic advertising measures is unclear as are their relationships to effectiveness measures supposedly related to sales.

The focus on single measures in advertising effectiveness measurement has not resulted in a total lack of empirical studies investigating a number of measures and their interrelationships. For instance, the study by Mehta (1994) is an example of a study aimed at developing methods of advertising effectiveness measurement that included a number of measures. In her study, four well-known

effectiveness measures and their interrelationships were analyzed using structural equation models. A number of other studies focusing on theory development rather than effectiveness measurement have included measures of several theoretical constructs and modeled the relationship between them (e.g., MacKenzie and Lutz, 1989), but there are few studies of this kind and most of those that exist have left out at least one or two measures that have been shown to be important in other studies. For example, in both studies mentioned here *brand beliefs* was left out (cf. Mittal, 1990; Fishbein and Middlestadt, 1995). Another problem is that studies aimed at theory development tend to emphasize other issues than effectiveness measurement. It is also worth noting that the research involving a number of measures is mainly done by academics; practitioners still focus on one or only a few measures at a time.

Given the limited number of studies that have included many measures, and the limitations of these studies, the *nomological networks* to which the effectiveness measures belong are incomplete. A nomological network is defined by Cronbach and Meehl (1955) as the system of "laws" which relates (*a*) observable phenomena, e.g., measures, to each other, (*b*) theoretical constructs to observable phenomena, and (*c*) theoretical constructs to each other. Incomplete nomological networks have at least five practical consequences. First, it becomes difficult to interpret measurement results if it is unclear how they are related to other measures and theoretical constructs (cf. Mehta, 1994). Second, there is a risk that the strength of statistical relationships is inflated or deflated if the statistical model excludes variables that influence one or several of the variables in the model. Third, different effectiveness measures should probably serve different purposes, e.g., diagnostic or evaluative, sometimes dependent on the situation in which they are used. Understanding the context in which the measure exists should simplify the choice of measure in most situations. Fourth, a well-developed nomological network clearly shows if in

fact there are measures that are measuring the same theoretical construct. If there are, these could be considered alternatives and evaluated to find out if one is better than another. Fifth, a nomological network should show whether a measure, in fact, is measuring more than one underlying construct. If a measure taps more than one construct, unequivocal interpretation of scores on the measure becomes impossible since it is not known whether the score reflects one or several of the constructs (cf. Peter, 1981).

For practitioners, a consequence of using too few advertising effectiveness measures is that learning before future advertising campaigns becomes difficult or impossible. One of the goals of both the British and Swedish Advertising Effectiveness Awards is to learn more about how advertising works and how to make advertising more effective (Duckworth, 1997; Sveriges Reklamförbund, 1998). It has also been shown that learning is important to practitioners in a more narrow context than effectiveness competitions. In a British study that investigated why leading ad agencies measured advertising effectiveness, about half of the agencies stated that learning and improving before future campaigns was an important reason for measuring advertising effectiveness (Flandin, Martin and Simkin, 1992). The problem is that if practitioners base their learning on effectiveness measurement that excludes important measures, they risk drawing incorrect conclusions and their learning for the future will be incorrect.

The second research problem also relates to the relationship between constructs, but to the relationship at different points in time. An obvious reason for pretesting advertising before an advertising campaign starts is to predict the outcome of the campaign. Given the complexity in the relationship between advertising and behavioral outcomes, it might be of interest to predict the intermediate outcomes of a campaign, i.e., the intermediate measures used in

advertising posttests. To make this possible, measurement before and after a campaign should be integrated, but in practice they are generally treated as separate entities (Hansen, 1998). In terms of research, only a few attempts have been made to analyze data from pretests and posttests together. Unfortunately, the studies that have combined these two types of data have generally used data aggregated over a group of individuals (e.g., Eastlack and Rao, 1989; Lodish et al., 1995); this makes it difficult to investigate the nature of the relationship between the variables in the study. There is little, if any, published research based on individual data, and therefore, little is known about the ability of different measures in advertising pretests to predict the outcome of campaigns as measured in advertising posttests.

The third, and final, research problem concerns a specific issue related to how measures of constructs should be operationalized, i.e., with one single question or with a number of questions. One of the main reasons for using more than one question, or item, to measure a construct is that measurement error in the questions tends to cancel out, which increases the reliability of the measure (cf. Churchill, 1979; Peter, 1979; DeVellis, 1991). In advertising research it appears that practitioners generally use single-item measures while academics normally use multiple-item measures. In most articles about *liking the ad*, for example, which are typically published by practitioners, measurement of liking the ad is made with a single-item measure (e.g., Biel and Bridgwater, 1990; Haley and Baldinger, 1991), while multiple-item measures generally are used in the mainly academic research on *attitude toward the ad* (e.g., MacKenzie, Lutz and Belch, 1986; MacKenzie and Lutz, 1989; Pieters and de Klerk-Warmerdam, 1996). There are also constructs which are measured in different ways by different academics. For example, in some academic studies brand attitude is measured with one single question (e.g., Patzer, 1991) while in others it is measured with multiple questions (e.g., Gardner, 1985). There also appears to be a difference

between constructs. For example, very few studies have measured *purchase intention* using a multiple-item measure; academics and practitioners alike generally use single-item measures.

If single-item measures are as reliable and valid as multiple-item measures, then it is a waste of respondent time and research money to use multiple-item measures. If, on the other hand, single-item measures are less reliable and valid than multiple-item measures, then the quality of the research which uses single-item measures is impaired. Unfortunately, few, if any, empirical results indicate which is the case. Moreover, little, if any, empirical advertising research exists that has compared single-item and multiple-item measures of commonly used constructs. Most studies include only one operationalization of the constructs included in the research, and the focus is on other issues. Hence, it is unclear whether time and money are wasted or if research quality is lower than it should be.

1.3 Purpose of the thesis

The discussion in the previous section highlights the lack of empirical research with respect to three broad issues in intermediate advertising effectiveness measurement. First, most of the research on intermediate measures of advertising effectiveness tend to exclude relevant measures, which means that relationships between measures might not be correctly understood. Second, little is known about the ability of advertising pretest measures to predict the intermediate measures commonly used in advertising posttests. Third, some constructs are measured with single-item measures by some researchers and multiple-item measures by others, but it is not known whether either way has implications on the quality of the research. In this thesis, empirical research will be carried out that addresses these three areas.

The purpose of this thesis is a) to investigate the relationships between constructs used as intermediate measures of advertising effectiveness in advertising, b) to investigate the relationship between constructs included in advertising pretests and constructs included in advertising posttests and c) to evaluate commonly used alternative operationalizations of constructs used as intermediate advertising effectiveness measures.

1.4 Scope of the thesis

The focus in the thesis is on *intermediate* measures of advertising effectiveness and their underlying constructs. For this reason, there will be no, or limited, discussion of *exposure* (or media) measures, *purchase behavior* measures, *sales* measures or *profit* measures. The focus on intermediate effectiveness measures also means that the research will focus on quantitative measures in advertising pretests and advertising posttests and not qualitative research at other stages of the advertising process. The reasons for these delimitations are outlined in this section.

There are two main types of research problems associated with exposure measures. The first concerns a number of methodological issues, e.g., how and how often data should be collected, which are considered interesting mainly because audience measures are used to set the prices of advertising space in most media vehicles (cf. Bogart, 1976; McKenna, 1988; Rubens, 1989; Ang, 1991; Appel, 1993; Danaher and Beed, 1993; Mallet, 1993). Second, a lot of research effort has been centered on the issue of effective frequency, i.e., how many times an individual should be exposed to an ad for it to be effective (cf. Ephron, 1995; Jones, 1995; Jones, 1997; Rossiter and Percy, 1997; Rossiter and Danaher, 1998). Neither of these two types of research problems is directly

related to the measurement of advertising effectiveness. This is not to say that exposure measures are unimportant in the evaluation of advertising campaigns, but their function is not to evaluate the advertising effectiveness. It should also be noted that the relationship between different exposure variables in most cases is straightforward as they are often mathematical functions of each other (cf. Dickson, 1991; Rossiter and Danaher, 1998).

To include measures of purchase behavior, sales, brand equity and profit in the thesis would probably require data from a number of real products in real markets. The practical difficulties associated with this approach are considerable. Companies are generally reluctant to share proprietary information, even if data are anonymous or disguised. Data on purchase behavior could be collected, and sales and market share data are available from syndicated research, but difficulty of isolating the effects of advertising from other marketing variables would still remain. In this thesis, the relationship between intermediate effectiveness measures and sales or other outcome variables will be discussed briefly in the literature review, but it will not be included in the empirical research.

The thesis will not deal with all types of advertising research. Lucas and Britt (1963) suggested the following four steps in advertising research:

1. Research prior to the development of the advertising
2. Research during the development of the advertising
3. Pretesting of advertisements
4. Posttesting of advertisements

The first two steps fall into what Zaltman and Moorman (1989) have labeled *developmental research*, i.e., mostly qualitative research aimed at discovery and

understanding. There is little or no evaluation of the effectiveness of advertising in these two types of research and, accordingly, they will not be included in this thesis. Instead, the focus will be on the measures in the remaining two steps, i.e., pretesting and posttesting of advertisements, both of which aim to ascertain the effectiveness of advertising.

1.5 Structure of the thesis

This thesis follows a traditional academic format. Chapter 2 reviews theory and earlier research and offers a general theoretical background on advertising and advertising effectiveness measurement. In addition, it contains a detailed discussion of all the measures included in the two studies which constituted the empirical research effort in this thesis. The research questions for the two empirical studies are presented at the end of Chapter 2.

Chapter 3 explains the research approach that was chosen for the two empirical studies in the thesis and gives the background and reasons for the methodological choices. Chapter 3 also offers a detailed description of all the measures that were used in the empirical studies.

Chapter 4 presents and discusses the analyses of the data from the two empirical studies. The chapter is organized according to the research questions presented in Chapter 2.

In the final chapter, Chapter 5, the results of the empirical studies are summarized and the conclusions and their implications for academics are presented. The chapter also includes limitations of the empirical studies, managerial implications and suggestions for further research.

2 Theoretical background and literature review

This chapter discusses the theory and earlier research on advertising effectiveness and the measurement of advertising effectiveness. The first part of the chapter deals with general issues, such as definitions of advertising effectiveness and intermediate measures of advertising effectiveness, how advertising works and some methodological issues pertaining to advertising pretests and posttests. The second part of the chapter offers a detailed discussion of the intermediate measures of advertising included in the empirical studies. This discussion includes theoretical as well as methodological issues.

The word *ad* will be used in this thesis to refer to all types of advertisements irrespective of media. This means that ad also refers to TV ads, which are sometimes referred to as *commercials*. The word *campaign* refers to the total advertising effort made by an advertiser during a certain period of time. A campaign, therefore, can be anything from a one-off ad to numerous ads in different media over a period of several years.

2.1 Introduction

2.1.1 What is advertising effectiveness?

Commercial products are advertised in order to increase the profits of the advertiser. The goal of the advertiser can be short-term profits, e.g., as a result of an immediate sales increase, or long-term profits, e.g., by building a strong brand, or both, but arguably no firm would invest money in advertising if it did not expect a positive return. Generally, advertising is assumed to increase profits by increasing sales, but advertising can also contribute to profits by increasing the price of a brand or lowering the costs associated with it (Rossiter and Percy, 1997). Hence, advertising that increases the profits of an advertiser by

increasing sales of the product, by increasing the consumers' willingness to pay a higher price for the product or by contributing to lower costs associated with the product, or any combination of these three, may be regarded as effective advertising. An ad or campaign that results in a greater increase in sales than another ad or campaign, all else being equal, is more effective. The same logic applies to ads that are relatively better at increasing the willingness to pay a higher price or at lowering costs.

The relationship between advertising and profits is complex. Advertising interacts with many other factors and the effects may be distributed over time. Also, advertising is rarely thought to have a direct effect on sales, price or costs, but is presumed to work indirectly via a number of intermediate variables, e.g., by influencing brand awareness or intention to purchase (for an overview of theories of how advertising works, see e.g., Rossiter and Percy, 1997; Vakratsas and Ambler, 1999). The complexity of the relationship between advertising and profits has led to changes in one or many of the intermediate variables often being referred to as advertising effectiveness (e.g., Aaker, Batra and Myers, 1992; Haugtvedt and Priester, 1997; Tellis, 1998). Strictly speaking, this is not true, but if there is a direct or indirect relationship between the intermediate variables and profits, the intermediate variables can be used as indicators of advertising effectiveness. This means that advertising that causes positive changes in intermediate variables related to sales – or other outcome variables that in turn are related to profit – could be regarded as effective advertising.

In this thesis, advertising effectiveness is regarded as changes in intermediate variables related to the profits of the advertiser. These changes can occur in the short term or in the long term and their impact on profits can be positive or negative. The view is also taken that advertising effectiveness is not

dichotomous, but continuous, i.e., that there are degrees of advertising effectiveness.

The view taken on advertising effectiveness can be extended to include advertising for non-profit organizations or advertising carried out by the state. In these cases, profit has to be defined in other terms, e.g., in terms of donations or reduced costs to society, but the rest of the argument holds.

2.1.2 Why should advertising effectiveness be measured?

What reasons are there for advertisers to measure the effectiveness of their advertising? Given that intermediate measures can be used before, during and after an advertising campaign, it is possible to see a number of different purposes of advertising effectiveness measurement.

One reason for measuring advertising effectiveness is to evaluate whether the investment in advertising yielded a positive financial return. This can only be done after a campaign has finished. Hence, it will not influence the campaign in any way and the evaluation mainly serves accounting purposes. Demonstrating that advertising yields a positive financial return could also serve to justify advertising budgets to stakeholders inside and outside the advertising company. This evaluation is more likely to be based on sales and profit measures than on intermediate measures and measures of the latter type are normally used to estimate the long-term value of advertising.

Another reason for measuring advertising effectiveness is to compare the potential effectiveness of different ads or ad executions (e.g., Bogart et al., 1970; Rossiter and Percy, 1997). Different ads do not always perform equally well in

terms of effectiveness and the advertiser should eliminate ads performing less well in a pretest.

Ads and advertising campaigns do not generally live forever. The decline in effectiveness of ads and campaigns has been referred to as *advertising wearout* (cf. Blair, 1987; Rossiter and Percy, 1997). Advertising effectiveness measurement can be used to check whether ads or campaigns are starting to wear out and whether they should be changed. This measurement must occur during campaigns or before the same ads are used in a new campaign.

Advertising effectiveness measurement can be used to improve parts of advertising campaigns to increase the total effectiveness (Lucas and Britt, 1963). Rossiter and Percy (1997), for example, suggest that continuous tracking studies be used to evaluate different campaign factors, e.g., the media budget and the accuracy and implementation of the media plan. This should be done while the campaign is being carried out.

Yet another important reason for measuring advertising effectiveness is to learn for the future. Effectiveness measurement can contribute to an understanding of why a campaign works or not (Sutherland, 1993; Rossiter and Percy, 1997), and thereby makes it possible to learn what to do and what to avoid when planning future campaigns. Measurement that contributes to the learning process can be carried out before, during and after a campaign.

The reasons for carrying out advertising effectiveness measurement can be compared with the actual reasons given by advertising agencies. A study of leading British agencies (Flandin et al., 1992) showed that less than half (45 %) of the agencies use advertising effectiveness research to learn for future campaigns. A majority of the agencies (68 %) mentioned measuring the

performance of the advertising as a reason, while 45 percent said that demonstrating the effectiveness of the advertising to their clients was a reason for doing effectiveness research. In a comment on the extent of advertising effectiveness research, the authors remarked that it is far from being part of routine advertising management.

2.1.3 What are intermediate measures of advertising effectiveness?

In a recent review of the literature on how advertising works, Vakratsas and Ambler (1999) used the term *intermediate effects* to denote the various mental effects advertising must have on consumers before it influences their behavior. Examples of intermediate effects caused by advertising are effects on memory and cognitive ("thinking") or affective ("feeling") responses. In short, intermediate effects are what happens in the minds of people following exposure to advertising. Based on this notion of intermediate advertising effects, intermediate measures of advertising effectiveness are defined as *measures of the mental effects which advertising creates in consumers*.

The main advantages of intermediate effectiveness measures compared with sales measures (or other behavioral measures) are threefold. First, many intermediate effects can be measured in advertising pretests. This means that it is possible to evaluate the ability of an ad or a campaign to create certain types of intermediate effects before the campaign has started (cf. Tellis, 1998). Second, there is generally less interference in the relationship between advertising and the intermediate effects than in the relationship between advertising and consumer behavior. Sales are influenced by many factors while fewer factors influence the intermediate effects (cf. Grønhaug, Kvitastein and Grønmo, 1991; Tellis, 1998). The number of other factors influencing intermediate effects and the strength of their influence is different for different

types of intermediate effects. (This will be elaborated on later in the thesis.) Third, in most cases, the delay between exposure to the advertising and the intermediate effects is shorter than between advertising and sales. This makes it possible to evaluate the effectiveness of ads and advertising campaigns at a point in time close to exposure, in some cases as soon as the day after exposure.

2.1.4 What is the role of theory in advertising effectiveness measurement?

Observation without theory is impossible. Karl Popper brought this point home to a group of physics students in the 1930s by issuing the following instructions: 'Take pencil and paper; carefully observe, and write down what you have observed!' (Popper, 1989: 46). The students were unable to fulfill this request since the instructions did not include what to observe. To observe something, we need an object to observe and this object emanates from some kind of theory.

Theory is not only necessary for all kinds of observations. A good theory is of great practical use when doing marketing, or other, research. Malhotra (1996: 51) points to the importance of using theory in six situations in the marketing research process. Theory should guide the researcher in the process of:

1. Conceptualizing and identifying key variables
2. Operationalizing key variables
3. Selecting a research design
4. Selecting a sample
5. Analyzing and interpreting data
6. Integrating findings

The situations discussed by Malhotra (1996) show that theory is important and should be used throughout the entire research process. In addition to these

concrete research implementation situations, theory is also necessary for an understanding of how different advertising effectiveness variables relate to each other, which is an important issue in effectiveness measurement (Mehta, 1994). It has also been pointed out that advertising effectiveness measurement should be tailored for different circumstances, but that this must be done within strict theoretical guidelines (Percy and Rossiter, 1997; Rossiter and Percy, 1997).

In the context of advertising effectiveness measurement, there is a need for three types of theoretical guidance. First, it is difficult to know what to measure and when to measure it without an idea of how advertising influences people. This points to a need for some general theory of how advertising works. Second, there must be theory for the nomological networks (Cronbach and Meehl, 1955) of the measures that are used in effectiveness measurement. In this case, theory should cover the relationship between measures and constructs, the relationship between measures and the relationship between constructs. Without theoretical guidance, it is virtually impossible to know how to operationalize measures and what other measures to include in order to correctly understand the relationship between measures. Third, there is a need for theory regarding measurement issues. This should cover how the measurement should be made, the sequence of measures in questionnaires and issues specific for the pretests and posttests. The first two types of theories will be discussed later in this chapter. The third will be discussed to some extent in this chapter, but it will also be discussed and applied in the next chapter (Chapter 3), which deals with the research approach of the empirical studies.

2.2 Theories of how advertising works

General theories of how advertising works tend to come in the form of hierarchy-of-effects models. These models vary with respect to the variables that

are included and the order in which effects take place, but in all models each effect is assumed to be dependent on the preceding effect, i.e., the included variables are related in a specific, hierarchical order (cf. Grønhaug et al., 1991; Tellis, 1998). In the well-known hierarchy-of-effects model *DAGMAR*, for example, it is assumed that advertising works in the sequence *awareness*, *comprehension*, *conviction* and *action* (cf. Scholten, 1996) and in another early model proposed by Lavidge and Steiner (1961) the sequence was *awareness*, *liking*, *preference*, *conviction* and *purchase*, which was summarized as *cognition*, *affect* and *conation*.

Hierarchy-of-effects models have been criticized. According to Scholten (1996) there have been two major criticisms. The first criticism is that the hierarchy-of-effects models have disregarded the marketing situation in which the advertising works; the second criticism is that effective advertising, according to the models, requires complex attitude changes in consumers. Various alternatives to the early hierarchy-of-effects models, in which the order of effects was cognition, affect and conation, have been suggested, e.g., that the order of effects is cognition, conation and affect and that low-involvement advertising works without the affect effect (cf. Smith and Swinyard, 1982). The debate about hierarchy-of-effects models is ongoing and there is no agreed-upon model among advertising researchers or even agreement that advertising works in a hierarchy of effects (cf. Vakratsas and Ambler, 1999).

Despite the criticism directed against them, hierarchy-of-effects models should be useful in the context of advertising effectiveness measurement, e.g., for identifying key variables and interpreting results (see Section 2.1.4), but caution is warranted. Hierarchy-of-effects models are useful as frameworks, but a certain sequence of effects, e.g., awareness, comprehension, conviction and action, should not be applied dogmatically to all situations since the sequence

might vary between different product categories, brands and situations. Further, not all effects are applicable in all situations. In addition, it is important to include the marketing context in effectiveness measurement, e.g., keeping track of changes in price, distribution or other marketing mix variables.

In the present context, a hierarchy-of-effects model from Rossiter and Percy (1997), *the six-step effects sequence*, will be used as theoretical framework. This model is more general than most hierarchy-of-effects models and it includes both individual and advertiser variables. According to the six-step effects sequence, advertising works through five prior effects before affecting the profit of a company. The steps, or effects, are the following:

1. Exposure
2. Processing
3. Communication effects and brand position
4. Target audience action
5. Sales or market share and brand equity
6. Profit

The first four steps are called the *buyer response steps*, which are the effects advertising has on individuals. People are *exposed*, or have the opportunity to see, advertising in different media. People might pay attention to ads, or parts of them, thereby *processing* them to a lesser or greater extent. As a result of processing an ad, there will be *communication effects* on some people, e.g., some people might learn a brand name and others might form an opinion about a brand. The communication effects, in turn, might lead to *target audience action* in the form of some people purchasing the product. The final two steps in the six-step effects sequence are aggregate effects on the market or company levels.

Each of the six steps is associated with different advertising effectiveness measures. Exposure is associated with media measures, processing is mainly associated with ad related measures, communication effects with brand related measures and target audience action with measures of purchase or other behavior. The final two steps are associated with sales and market measures and profit measures.

In the six-step effects sequence, intermediate advertising effects are found in the processing and communication effects steps. Accordingly, the focus in this thesis will be on these two steps, the measures associated with them and the theoretical constructs underlying these measures.

2.2.1 Processing and communication effects

The immediate responses to an ad - attention, learning, emotional responses and acceptance - are referred to as *processing* (Rossiter and Percy, 1997). This means that measures of processing should be focused on the ad, not the brand, and that the measurement should ideally be made directly after exposure to the ad. If measurement is to be made immediately after exposure, then measures of processing should be more suitable for advertising pretests than for advertising posttests.

In contrast to processing, *communication effects* are relatively enduring mental associations and they are connected to a brand (Rossiter and Percy, 1997). The five communication effects are: *Category need*, *brand awareness*, *brand attitude*, *brand purchase intention* and *purchase facilitation*. Four of these are related to the brand (the exception being category need), and, therefore, most measures of communication effects should focus on the brand. Also, as

communication effects can be measured directly after exposure, or after a delay, they could be included in advertising pretests as well as advertising posttests.

It is important to distinguish between processing and communication effects. Communication effects are closer to the behavioral response, i.e., target audience action, in the six-step effects sequence and should have more influence on purchase behavior. They operate on different levels: Processing effects are related to the ad and communication effects are related to the brand (Rossiter and Percy, 1997). Measurement of the two types of effects is also different with regard to when they are best suited to be measured and what the purpose of measurement is.

On the issue of how processing and communication effect measures should be used, Rossiter and Percy (1997) point out that they serve different purposes. Processing measures serve mainly negative-diagnostic purposes and should not be regarded as ends in themselves. Their usefulness lies mainly in helping to explain how and why certain ads work or do not work. As indicators of whether an ad is working or not, measures of communication effects should be considered more important than processing measures.

2.3 Pretesting and posttesting

Pretesting refers to testing ads before circulating them in the media (Lucas and Britt, 1963). Since the cost of producing ads is generally much lower than the cost of media space, it makes sense to pretest ads in order to detect and, if necessary, alter any weaknesses in them. Pretests should also be used to avoid potentially harmful ads, e.g., ads that reduce sales (cf. Bogart et al., 1970) or ads that might harm the image of the advertiser or the brand. Pretests can also be used to choose the most effective ads from a pool of suggested ads. The focus in

pretests is on the creative content of the ad, not on other campaign factors, e.g., media, which have to be evaluated in posttests (Rossiter and Percy, 1997).

According to Lucas and Britt (1963) *posttesting* is effectiveness research carried out after a campaign has run, but this old definition is too narrow with regard to timing; effectiveness measurement can be done while a campaign is running and brand tracking studies usually run continuously. In this thesis, posttesting will be used to refer to advertising effectiveness measurement that is made after the first media insertion of the ad(s) in a campaign. It should also be noted that, technically, pretests can be carried out after a campaign has started, but this does not mean that they should be called posttests. The main difference between pretests and posttests is that pretests evaluate the creative content by trying to control for other factors, while posttests evaluate the effects of the entire campaign, including, e.g., media factors.

Another difference between pretests and posttests lies in their purpose. Pretests are used to choose between alternatives, e.g., to run or not run an ad or to choose the best ads from a pool of ads. Posttests, on the other hand, are used to evaluate the outcome of the campaign implementation that was chosen; different options are, therefore, not compared.

There has been much research published on a number of methodological issues, other than what measures to use, in pretesting. These will be discussed in the next section. There has been considerably less published on methodological issues in posttesting. Given the nature of posttests this is, perhaps, not surprising. Besides choosing the method for collecting data, there are few, if any, methodological choices in posttesting. Since the pros and cons of using telephone, mail, personal interviews or (in a near future) the Internet to collect data are well known and discussed in all textbooks on marketing research (e.g.,

Malhotra, 1996; Kumar et al., 1999), little, if any, research exists that compares different data collection methods of posttesting are compared. Another reason for the lack of research in this area could be that posttest data are considerably more difficult to collect than pretest data. Pretests can be set up relatively easy, in classic laboratory experiment style, while posttests require data from a real campaign or an elaborate research design.

2.3.1 Methodological issues in pretesting

A number of methodological choices need to be made before ads can be pretested. These choices apply to the degree of finish of the ads to be tested, the setting of the pretest, when to perform the measurement, what type of measures to use, and so on. In this section, a number of these methodological issues will be reviewed.

Ads can be tested before the final version of the ad has been produced. TV ads can be tested by showing storyboards, and magazine ads by showing ads with sketches instead of photos, and so on. If pretests with sketches of ads produce valid results, it is possible to save money on production costs. Rossiter and Percy argue that the degree of finish of the ads in pretests depends on the brand attitude strategy (Percy and Rossiter, 1997; Rossiter and Percy, 1997). *Informational* ads require less degree of finish, rough ads can be used, while *transformational* ads need to be almost finished. Rossiter and Percy refer to results from a study by McCollum Spielman in which a high correlation ($r >$ .80) was found between persuasion scores for rough and finished versions of TV ads. Other studies have also found a fairly close relationship between the pretest results for rough and finished ads. Brown and Gatty (1967) obtained mostly comparable results in pretests with either rough or finished TV ads, but it should be noted that they used live-action roughs, not storyboards. A comparison

reported in Appel and Jackson (1975) showed that the rank order of pre-post changes in a measure of persuasion was the same for both rough and finished ads, but that there was a difference in the size of the pre-post differences. In a study by Schlinger and Green (1980) finished TV ads were found to be more attractive, stimulating and less irritating than storyboards, but there were no differences with regard to brand variables, such as persuasion.. The literature review did not reveal any studies which had compared rough and finished versions of print ads.

While it is common to pretest ads in focus groups, this is unsuitable for several reasons. On a general level, it can be noted that the focus group technique is *qualitative*, i.e., it is oriented toward insights and understanding and is of an exploratory character (e.g., Malhotra, 1996). Advertising effectiveness, in contrast, is mainly a *quantitative* research problem: The aim, for example, is to ascertain to what extent something happens or whether one ad performs better than another. Given the quantitative nature of the research problem, quantitative research methods would clearly be more appropriate.

On a more specific level, Rossiter and Percy (Percy and Rossiter, 1997; Rossiter and Percy, 1997) discuss three problems associated with using focus groups to pretest ads. First, the ads are overexposed, which leads to an exaggeration of their positive and negative aspects. Second, the processing of the ads takes place in a group, a setting which is too different from that of normal, individual processing. Third, given the normal size of focus groups, the sample size will be too small for reliable projection of the results.

TV ads can be tested *on-air*, in real commercial breaks, or *off-air*, in some kind of artificial setting. There has been some research on which of these two methods is the best. In on-air pretests, the test ads are inserted in real

commercial breaks and respondents are interviewed after the ads have been aired. In off-air tests, respondents are recruited, e.g., in a shopping-center, exposed to the ad in an artificial setting and then surveyed about their opinions. The result reported in the ARF Copy Research Validity project was that off-air tests were consistently better at predicting sales than on-air tests: The percentage of correct predictions, i.e., to pick the better performing of two ads, was 63 percent for off-air tests and 44 percent for on-air tests (Haley and Baldinger, 1991: 26). In spite of their results, Haley and Baldinger (1991) did not conclude that off-air tests were better than on-air tests. In their view, the difference was due to difficulties in controlling for target group characteristics and program environment in the on-air pretests, problems that were not seen as being inherent in the method.

Another study found that test-retest reliability was lower for on-air tests than for off-air tests (Clancy and Ostlund, 1976). Clancy and Ostlund attribute the difference they found in reliability between the two methods to largely the same reasons. They argue that it was a result of a lack of control over program environment, lack of control over target group characteristics, variation in the time slot and possible interviewer effects that caused the lower reliability in on-air tests. Unlike Haley and Baldinger (1991), Clancy and Ostlund conclude that the lack of control over a number of factors is inherent in on-air tests and that off-air tests are more reliable.

In pretesting, the researcher has to decide on how many times the respondents should be exposed to the ad. According to Percy and Rossiter (1997; Rossiter and Percy, 1997) the appropriate number of exposures depends both on media type and brand attitude strategy (low or high involvement and informational or transformational purchase motivation). Newspaper and magazine ads require one exposure without limit on the exposure time. Broadcast informational ads

require two exposures, while broadcast transformational ads require three exposures. In the ARF study, the results in a comparison of single and multiple exposures were mixed. For some measures, the share of correct predictions was higher for single exposures, and for other measures the share was higher for the multiple exposures (Haley and Baldinger, 1991: 27). Rosenberg and Blair (1994) found that ARS persuasion scores did not differ for a different number of exposures, but that a statistically significant effect on ad recall was evident: Double exposures led to higher levels of recall than a single exposure.

Ads can be tested in a pre/post design, i.e., by measuring pretest measures on the same individuals before and after exposure to the test ad. An alternative procedure is a post-only design in which measurement is only made after exposure to the ad. Any effects caused by the ad are traced by comparing the individuals exposed to the ad with individuals in a control group who are not exposed. Two arguments against pre/post designs are that they require too much time because of the interval required between the measurements, and a great deal of work is needed to prevent the pre-measure from influencing the post-measure (Rossiter and Percy, 1997). Rossiter and Percy conclude that post-only designs are to be preferred when *transformational* ads are tested, but that a pre/post design can be used for *informational* ads. Not everyone has reached the conclusion that pre/post designs are not always appropriate. The ARS persuasion scores developed by *research system corporation*, for example, are based on a pre/post design (see e.g. Blair, 1987; Rosenberg and Blair, 1994). The results in the ARF Study indicate that the post-only design is better than the pre/post design: Post-only tests resulted in 65 percent correct predictions and pre/post in 55 percent correct predictions. Despite this the authors did not conclude that post-only tests are better (Haley and Baldinger, 1991: 26).

A less common method of pretesting ads is to use *continuous measurement*, i.e. reactions to an ad are recorded continuously when it is shown, instead of only afterwards. Fenwick and Rice (1991) tested the reliability of a continuous measurement pretesting method, namely the Program Evaluation Analysis Computer (PEAC), and conclude that the system exhibits a high level of test-retest reliability. Of some interest is the authors' motivation for why continuous measurement should be employed. They argue that emotions can change during a TV ad and that any response measured afterwards could be the result of several emotions experienced during the length of an ad. Hence, it should be possible to identify parts of the commercial which have negative effects on the viewer. It should be noted that the variable that is measured in the PEAC, and in other similar tests, is of the *liking the ad* type. The validity of this type of measure is disputed and will be further discussed below.

Another unusual way to test ad execution is to use physiological measures. An argument in favor of galvanic skin response (GSR) is that it measures an involuntary response that should have higher predictive validity than verbal self-reports, which are influenced by the respondents' rational and cognitive processes (LaBarbera and Tucciarone, 1995). An argument against physiological measures is that they only measure attention (Percy and Rossiter, 1997).

2.3.2 Predictive validity of pretests

One of the main issues concerning pretests is whether they can actually predict the effectiveness of the ads tested. In most cases, this has been taken to mean prediction of the change in sales that the ad will cause or prediction of which ad will increase sales most. Results indicating that advertising pretests can actually predict sales change or pick the best ad have been reported in a number of studies (e.g., Buzzell et al., 1965; Haley and Baldinger, 1991).

Some published results also indicate that the ARS persuasion score used in pretests by *research systems corporation* can predict whether an ad will increase sales or not (Blair, 1987; Adams and Blair, 1992; Blair and Rosenberg, 1994; Rosenberg and Blair, 1994). Some doubts on the predictive validity of the ARS persuasion measure were cast by Lodish et al. (1995) who failed to replicate the results in Blair (1987). Their database of 389 BEHAVIORSCAN® tests contained tests with statistically significant sales differences although the ARS scores were low, as well as tests where the reverse was true, i.e. where sales differences were not statistically significant despite high ARS scores.

A somewhat different view on the validity of advertising pretests is taken by Bloom, Jay and Twyman (1977). They argue that the validity of pretests should be judged on their ability to aid decisions concerning whether to use ads as they are or whether to change them, not on their ability to predict exactly what will happen in the market.

Another way to evaluate at the predictive validity of pretests is to examine their ability to predict the communication effects of from advertising campaigns that are often measured in advertising posttests. It was argued earlier (see section 2.1.3) that intermediate advertising effectiveness measures are often better measures of advertising effectiveness than sales or other outcome variables. If the effectiveness of campaigns is evaluated in terms of measures of communication effects, then it makes sense to evaluate the potential effectiveness of ads on the basis of measures that can predict the post-campaign communication effects. To date, little, if any, research has been done on the ability of pretest measures to predict post-campaign communication effects.

2.3.3 Tracking studies

The invention of tracking studies, or continuous tracking, has been attributed to the British research firm *Millward Brown* (Broadbent, 1999). Characteristically, tracking studies are carried out on a regular basis, e.g., monthly or quarterly, and the main object of measurement is the brand, not the ad or ads. Tracking studies generally include a number of intermediate effectiveness measures, e.g., advertising recall and brand recall, and sometimes also media measures and sales measures. Rossiter and Percy (1997) recommend that tracking studies include measures for all steps in the six-step effects sequence, i.e., not only intermediate measures but also media measures, sales measures, and so on.

According to Rossiter and Percy (1997: 596-7) there are five reasons for tracking:

1. Determining why the campaign is or is not working
2. Adjusting the budget
3. Adjusting the media plan
4. Adjusting the exposure ratio of individual ads in the pool
5. Making minor improvements in ads

It is of some interest to note that the last two reasons require that the tracking study include measures relating to the individual ads in the campaign. This means that tracking studies should ideally include measures of, e.g., ad recall or ad recognition for each individual ad in the campaign and not only overall campaign measures such as recall of any advertising by a brand or brand attitude.

With regard to the relationship between intermediate effectiveness measures in tracking studies and sales, Eastlack and Rao (1989) report some instances in

which tracking measures were not related to sales and recommend therefore that tracking measures such as brand awareness and slogan recall be interpreted with caution. *Millward Brown*, on the other hand, claim that a strong relationship exists between their *Awareness Index*, which is based on a tracking measure, and sales (e.g., Brown, 1985; Hollis, 1995), but this claim is not widely accepted (Rossiter and Percy, 1997: 608).

2.4 How many items does it take to make a measure?

Over the last twenty years, academic researchers in marketing have made increasing use of multiple items for measuring marketing constructs. A starting point for this was the special section on measurement and marketing research in the February issue of the *Journal of Marketing Research* in 1979. Two articles in this section, Churchill (1979) and Peter (1979), argued strongly in favor of developing and using multiple-item measures in marketing following the psychometric measurement tradition, and appear to have influenced much of the development since then. Today, marketing constructs are routinely measured with multiple items and the measures are generally accompanied by equally routinely calculated Cronbach's alphas. The use of multiple items is also reinforced by the use of LISREL, or similar software for structural equation modeling, which works better with multiple-item measures than with single-item measures (cf. Anderson and Gerbing, 1988; Baumgartner and Homburg, 1996).

Some authors have argued that multiple-item measurement might not always be necessary or appropriate for some constructs (e.g., Wanous and Reichers, 1996; Finn and Kayandé, 1997; Wanous, Reichers and Hudy, 1997; Rossiter, 1999). Among practitioners, use of multiple-item measurement is less widespread, and reliance on single-item measurement is therefore greater. In the context of intermediate advertising effectiveness measurement, two measures tend to be

operationalized with multiple items by academics and with single items by practitioners. The first of these is *attitude toward the ad*, which is also referred to as *liking the ad*, and the second is *brand attitude*. Academics tend to use at least three items to measure both these measures while practitioners tend to use one item. Both measures will be discussed in detail later in this thesis.

A procedure for determining whether measurement should be made with multiple items or a single item is offered by Rossiter (1999). This procedure is based on the properties of the *topic*, i.e., the object of measurement, and the properties of the *attribute(s)* measured for the topic. Both topics and attributes can be *concrete* or *abstract*. Coca-Cola and Ford Taurus, for example, are concrete topics, while capitalism and marketing are abstract. Concrete attributes, e.g., trustworthy or likable, can be measured with a single item, while abstract attributes, e.g., service quality or opinion leadership, require multiple items. Topics can be further divided into *singular* or *collective* topics. A singular topic refers to a single object, e.g., President Clinton, or a homogenous set of objects, e.g., Coca-Cola; a collective topic refers to a set of heterogeneous objects, e.g., market researchers or fast food restaurants. Abstract attributes can be further divided into *formative* or *effector* attributes. In the case of abstract formative attributes, the measurement items define the attribute, e.g., socio-economic status, and in the case of abstract effector attributes the attribute causes the response in the items. Together, there are six different combinations of topics and attributes and for one of these combinations, namely a concrete and singular topic with a concrete attribute, Rossiter argues that a single item measure is sufficient. In the other five combinations, various types of indices should be used.

In the present context, it could be argued that a specific ad is a concrete and singular topic and that the likability of the ad, or the attitude toward it, is a concrete attribute. The same can be said for the attitude toward a specific brand.

Very little, if any, research has evaluated whether single-item measures of attitude toward the ad and brand attitude can replace multiple-item measures. This is rather unfortunate as, depending on the results of an evaluation, either practitioners could improve the quality of their measurement or academics could make their measurement more efficient.

2.5 Measures in the empirical studies

The following sections review the theory and earlier research relating to the intermediate advertising effectiveness measures included in the two empirical studies. The starting point for the selection of measures to include in the studies was the discussion of advertising effectiveness measurement in Rossiter and Percy (1997). All measures relevant to the *processing* and *communication effects* steps in the six-step effects sequence were seen as candidates to be included in the study, while the measures for the other steps were not regarded as intermediate effectiveness measures and excluded from the studies.

Rossiter and Percy (1997) discuss six processing measures and seven communication effects measures. One of the processing measures, *learning*, and two of the communication effects measures, *category need* and *purchase facilitation*, were excluded from this work. The measure of *learning* was excluded as this is mainly a diagnostic measure intended primarily for low-involvement advertising and the studies include ads for high-involvement products. The two communication effects category need and purchase facilitation are not relevant to all advertising and, hence, the measures of these

effects are not applicable to all advertising. The other communication effects measures, in contrast, are generally applicable in pretesting and posttesting of all ads.

The advertising effectiveness measurement discussion in Rossiter and Percy (1997) is used as a starting point because it is probably the most comprehensive discussion of measures available. Further, their discussion of advertising measures is also, in contrast to much other work, explicitly related to theories of how advertising works. This facilitates the process of understanding the nomological network of which the measure is part (Cronbach and Meehl, 1955). The literature review revealed only one potentially useful intermediate effectiveness measure, namely ad credibility (MacKenzie and Lutz, 1989), which was not included in the discussion in Rossiter and Percy. This indicated that their discussion was a good starting point. Another indication was that the measurement discussion in Rossiter and Percy is relevant to practitioners. Analyses of the British and the Swedish *Advertising Effectiveness Awards* (Bergkvist, 1999a; Bergkvist, 1999b) show that the advertising effectiveness measurement overview in Rossiter and Percy can be used to classify and analyze measures used by practitioners.

The measures included in the empirical studies are shown in Table 1. The measures are divided into processing and communication effect measures and each measure's suitability, for advertising pretests, posttests, or both is shown (cf. MacKenzie and Lutz, 1989; Rossiter and Percy, 1997).

Measure	Type of test
Processing measures	
Ad recognition	Posttest
Ad recall	Posttest
Adjective Checklist (ACL)	Pretest
Cognitive response measurement (CRM)	Pretest
Ad credibility	Pretest
Attitude toward the ad/liking the ad	Pretest
Communication effects measures	
Brand recall	Pretest and posttest
Brand recognition	Pretest and posttest
Brand benefit beliefs	Pretest and posttest
Brand attitude	Pretest and posttest
Brand purchase intention	Pretest and posttest

Table 1. The measures included in the empirical studies.

In the following sections, each of the measures in Table 1 will be discussed, and in the same order as in the table. The discussion will include the theoretical background of the measures, their relationship to other measures and how they are operationalized. As the focus will be on theory and measurement issues relevant to advertising, the discussion of measures that have an extensive theoretical background, e.g., *brand attitude*, will deal with issues particularly relevant to advertising research and not to the role of brand attitude in a wider context or to attitudes in general.

2.6 Ad awareness: Ad recognition and ad recall

Two commonly used intermediate advertising effectiveness measures are *ad recognition* and *ad recall*. These two measures were introduced more than 50 years ago (Lipstein, 1984) and it has been suggested that their importance as benchmarks for advertising performance is based on their historical impact rather than their ability to predict sales (Tellis, 1998).

Measures of *ad recognition* and *ad recall* belong to the category of *memory tests* (cf. Lucas and Britt, 1963; Mehta, 1994). It has been shown that ad recall and ad

recognition measure the same underlying part of memory for advertising, but that the measures also reflect other mental states, e.g., interest in ads, which can account for differences in results between the two measures (Bagozzi and Silk, 1983). The two measures are also different in that recognition requires simple identification of a previously seen object, whereas recall requires reproduction of an object without any external aid (Strong, 1912).

The response to memory tests of advertising depends on two reactions during exposure: First, whether people note and pay attention to the ad and, second, whether they remember it (Lucas and Britt, 1963). In line with this, Rossiter and Percy (1997) regard ad recognition and ad recall as measures of *processing* that indirectly measure *competitive attention* when included in tracking studies, since they require some learning from the ad.

Some research has been carried out on the relationship between ad recognition and ad recall. Zielske (1982), for example, compared ad recall and ad recognition for both TV ads and magazine ads. He found that the recall of TV ads was, on average, 29 percent lower than the recognition of the same ads, and that the recall of magazine ads was, on average, 62 percent lower than the recognition of the same ads. A smaller difference was found in a study by the Dutch research firm *NIPO* reported in Stapel (1998); ad recognition was five percent lower than ad recall for magazines and six percent lower for newspapers. The *NIPO* study also found a high correlation between ad recognition and ad recall (.96 for magazine ads and .95 for newspaper ads). An earlier study concluded, on the basis of three experiments, that no correlation is to be found between ad recognition and ad recall (du Plessis, 1994) The grounds for this conclusion, however, are very unclear; du Plessis does not report any correlation between the two measures for two of the experiments, and for the third experiment, he reports that $R^2 = .397$ when ad recall is used to explain ad

recognition. Consequently, the correlation between the variables is .63, which must be considered to be fairly high.

Ad recognition and ad recall can be included in both pretests and posttests. If they are included in pretests, the measurement must be made some time after the pretest, otherwise the validity of the measure in this case is dubious (Percy and Rossiter, 1997).

2.6.1 Ad recognition

Ad recognition is a measure of whether people can *recognize* an ad when they are exposed to it. The measurement of ad recognition depends on the medium in which the ad is placed. Ads from visual media, e.g., TV or print ads, should be shown to the respondent. The ad is generally shown during a personal interview, but the Internet will soon offer new ways of displaying visual advertising to respondents. If it is impossible to show the ad to the respondent, ad recognition could be measured by describing the ad, but the results will not be comparable with those from conventional ad recognition measurement (du Plessis, 1994). Radio ads must be played from a recording to the respondent, which can be done either in a telephone interview or a personal interview.

An elaborated version of ad recognition is *masked recognition*, in which the brand and package are deleted from the ad and people are asked to identify the brand (Rossiter and Percy, 1997). The term *proven recognition* is sometimes used to refer to those who correctly identify the brand in masked recognition tests (e.g., Zielske, 1982). If people are asked how many times they have encountered the ad in a given time period, this is a measure of *claimed ad recognition frequency* that can be used to check the effective reach of the campaign (Rossiter and Percy, 1997).

Early research with print ads showed that recognition levels were influenced by variables such as size of the advertisement, product category and gender (Lucas and Britt, 1963). The same variables influenced recognition levels also in later research, in which the use of colors and changes over time also were found to be important (Grønhaug et al., 1991). A study by Rossiter (1981) found that using nouns in the headline, having a personal reference in the headline and picture size were positively related to recognition of magazine ads, and that verbs in the headline were negatively related to ad recognition. The importance of the size of the ad has also been shown with ad recognition figures for 854 black and white magazine and newspaper ads, which showed a strong relationship (r = .94) between ad size and ad recognition (Stapel, 1998). Stapel also reports that ad recognition was 30 percent higher for those who use or buy the product category than for those who do not.

Little, if any, research examines the relationship between ad recognition and other intermediate variables. There appears to have been some interest in research to relate ad recognition to ad recall, but not to any other variables.

2.6.2 Ad recall

Ad recall is a measure of whether the respondent can *recall* seeing advertising or a specific ad for a brand. The wording of ad recall questions determines the degree of aid the respondent gets when trying to recall ads, and for this reason, ad recall measures are only comparable when the questions are worded similarly (Lucas and Britt, 1963). The most important difference in the wording is the cue for the recall. Ad recall can be measured using the brand name as cue, *brand-prompted ad recall*, or the product category as cue, *category-prompted ad recall*, also called *cut-through* (Rossiter and Percy, 1997). Brand-prompted ad

recall can either be *claimed*, i.e., the respondents simply state that they have seen advertising for the brand, or *proven*, i.e., the respondents can correctly describe parts of or the entire ad (Rossiter and Percy, 1997). Sometimes, the term *related recall* is used, but there appears to be some confusion about what this is. According to Tellis (1998), related recall is the same as *claimed recall* while according to Stewart (1986), related recall is the same as *proven recall.*

One of the most widely used intermediate measures of advertising effectiveness is the *Awareness Index* developed by *Millward Brown*. The awareness index is based on brand-prompted ad recall without reference to any particular ad (see e.g., Brown, 1985; Rossiter and Percy, 1997).

According to Rossiter and Percy (1997), brand-prompted ad recall is measured by showing, or reading out, a list of brand names and asking which of the brands the respondent has recently seen, or heard, ads for. This is *claimed* ad recall. The first question can then be followed up with a question asking for a description of the ads seen for the different brands, which makes it possible to verify that the respondent has seen a certain ad. If the respondent correctly describes the ad, or parts of it, the ad recall is *proven*.

Measurement of category-prompted ad recall, or ad cut-through, is made by asking what advertising the respondent remembers for a certain product category (Rossiter and Percy, 1997). The respondent is asked to describe the ads and, if the brand name is left out of the description, what brand the ad was for.

Since ad recall and ad recognition are correlated, they would be expected to share factors of influence. This was the case in the *NIPO* study reported in Stapel (1998), which found a high correlation between ad recall of magazine and newspaper ads and the same factors that influenced brand recognition, i.e., ad

size and product category use or purchase. In a study that analyzed more than 1000 TV ads, it was found that the most important factor in terms of influence on ad recall was that the ad had a brand-differentiating message (Stewart and Furse, 1986). A brand-differentiating message as defined in the study is a message which is unique to the advertised product, i.e., the same message claims could not be made by a competitor. The brand-differentiating message variable accounted for about five percent of the variation in ad recall, but when the study was replicated, the proportion of variance explained decreased to about one percent (Stewart and Koslow, 1989).

Some studies have examined how ad recall is related to other intermediate effectiveness variables. Research, for example, on the relationship between ad recall and *persuasion* (measured with ARS scores) found the correlation between the two variables to be between .2 and .4, and it was suggested that ad recall is mainly important as a variable that moderates persuasiveness (Stewart and Furse, 1984; Stewart, 1986; Stewart and Koslow, 1989). A similar suggestion is made by Plummer (1972), who writes that measures of ad recall indicate how many people have been affected in some way by advertising, but not in what way or the intensity of the effect.

2.6.3 Concluding comments on ad recognition and ad recall

The above discussion shows that ad recognition and ad recall appear to be related variables that probably measure the same underlying memory of advertising. They have been used for a long time and there is no indication in the literature of any current controversy regarding how they should be operationalized or any methodological issues. Measures of ad recognition and ad recall will be included in the empirical studies in this thesis, but mainly as a way of checking that the respondents have been exposed to the ads in the study.

Since ad recognition and ad recall are influenced by media factors, e.g., size of ad and the use of color, it was decided not to try to investigate the two measures in any depth. In this section some more general remarks will be made.

It is quite common to use ad recognition and ad recall as the only measure, or one of a few measures, of advertising effectiveness. Ad recall, for instance, is included in most of the case histories in both the British 1996 and Swedish 1998 *Advertising Effectiveness Awards* (cf., Duckworth, 1997; Sveriges Reklamförbund, 1998). It is also quite common for *Millward Brown* to argue that their *Awareness Index*, based on brand-prompted ad recall, is positively related to sales and can be used as a measure of advertising effectiveness (e.g., Brown, 1985; Hollis, 1995), an argument which has been questioned other authors (cf. Rossiter and Percy, 1997: 608). The question is whether ad recognition or ad recall are suitable measures of advertising effectiveness.

It has been pointed out that ad recall is not an end in itself (Zielske and Henry, 1980); the same can be said for ad recognition. Doubts have also been cast on whether a relationship actually exists between sales and ad recognition or ad recall. In the study by Bogart et al. (1970), for example, the correlation between ad recognition and sales was .13 and the correlation between brand-prompted ad recall and sales was .03.

It might be argued that measures of ad recognition and ad recall should mainly be used for other purposes than to directly evaluate the effectiveness of advertising. Rossiter and Percy (1997) argue that ad recognition should be used to evaluate the media plan, that proven brand-prompted ad recall should be used to evaluate the link between the ad and the brand, and that category-prompted ad recall should be used to evaluate the ability of ads to gain attention in the media. In addition to this, ad recognition and ad recall could be included as independent

variables in statistical analyses of posttest data. For example, ad recall could be used as an independent variable in a regression analysis with brand attitude as dependent variable, which would show whether processing an ad leads to a more positive brand attitude.

Practitioners, and sometimes academics, occasionally refer to various measures of ad recall as *spontaneous* ad recall or *spontaneous* ad awareness (see e.g., Duckworth, 1997; Sveriges Reklamförbund, 1998). The use of the word "spontaneous" is somewhat surprising. Recall of an ad cannot be spontaneous as it has to be elicited by more than a blank stare from an interviewer or silence on the other hand of the phone (cf. Lucas and Britt, 1963). It is equally misleading to talk about unaided recall (e.g., Tellis, 1998), as some degree of aid is needed to elicit recall. For the sake of clarity, and to avoid comparisons of different, not directly comparable ad recall measures, it would be better to use terminology which clearly indicates the type of recall measure , e.g., category-prompted ad recall or brand-prompted ad recall.

2.7 Two measures of acceptance

According to Rossiter and Percy (1997) *acceptance* is when a person agrees with the benefit claims he or she perceives are being made for a brand in an ad (see also Wright, 1973). Another possible response is *rejection*, i.e., that the person does not agree with the benefit claims made in the ad. Rossiter and Percy (1997) discuss two types of measures under the heading of acceptance, namely *adjective checklists* and *cognitive response measurement*. According to Rossiter and Percy, one advantage of adjective checklists is the relative ease of administration, but it has been suggested that the validity in cognitive response measurement is higher since an adjective checklist might prompt responses which did not occur (Rossiter and Percy, 1997).

Neither Rossiter and Percy (1997) nor any other author in the reviewed literature claim that adjective checklists or cognitive response only measure acceptance, i.e., the measures could also measure other underlying constructs. Whether the measures tap into other constructs and, if so, if they tap into the same constructs is not clear from the reviewed literature. Cognitive response measurement, for example, has been used as a measure of *ad perceptions,* which were defined as consumer perceptions of advertising stimuli (MacKenzie and Lutz, 1989: 51). This definition is, at best, a description of what cognitive response measurement captures, but hardly a definition of a theoretical construct. In a similar vein, adjective checklists have been described as measuring subjective reactions to advertising (Wells, Leavitt and McConville, 1971: 11). This thesis regards adjective checklists and cognitive response measurement as measures of acceptance, but it does not rule out that they also might measure other underlying constructs in an advertising processing context.

The following two sections offer detailed discussions of adjective checklists and cognitive response measurement. This is followed by a concluding section in which issues pertaining to each measure and the relationship between them are discussed.

2.7.1 Adjective Checklist (ACL)

An adjective checklist (ACL) is made up of a list of adjectives, or short adjectival descriptions, from which the respondent checks off those that reflect his/her reactions to an ad. ACLs, also known as reaction profiles (Wells, 1964; Wells et al., 1971) or response profiles (Zinkhan and Fornell, 1985; Zinkhan and Burton, 1989), are commonly suggested for use as a diagnostic tool, i.e., to understand why an ad works in one way or another, (cf. Aaker and Bruzzone,

1981; Percy and Rossiter, 1997). It has also been suggested that ACLs can be used to predict the effectiveness of ads (Zinkhan and Fornell, 1985; Aaker and Stayman, 1990; Hollis, 1995).

Measurement with an adjective checklist is done by first showing the respondent an ad and then during or after the exposure asking for her/his reactions to the ad. A typical wording of the question in an ACL is "Please tell how well you think each of these words describes the ad you just have seen" (Wells et al., 1971: 14). The question is followed by a list of adjectives or adjectival statements. Responses have been measured with eight-point semantic differential scales (Wells, 1964), with seven-point semantic differential scales (Mehrotra, Van Auken and Lonial, 1981), five-point scales with endpoints *extremely well* and *not well at all* (Wells et al., 1971) or with a simple, dichotomous, checkbox (Aaker and Bruzzone, 1981; Aaker and Stayman, 1990; Biel and Bridgwater, 1990; Bruzzone and Tallyn, 1997). A study by Crites, Fabrigar and Petty (1994) on a similar type of measurement showed that dichotomous checklists, multiresponse checklists and semantic differentials have equally good and stable psychometric properties. On the basis of that result Rossiter and Percy (1997) recommend the use of dichotomous checkboxes in ACLs.

Much research effort has gone into finding out which adjectives to include in the ACLs. In this work, factor analysis has often been used to identify underlying dimensions in the checklists and to reduce the number of items. Table 2 shows the 26 items in the *Print Ad Reaction Profile*, an adjective checklist for print ads developed by Wells (1964). The items are grouped according to a factor analysis in which Wells found three dimensions in his data. He labeled these dimensions *Attractiveness*, *Meaningfulness* and *Vitality*. (The item "Colorful-Colorless" was, without any explanation, included in both *Attractiveness* and *Vitality* in

Wells' article and for this reason it is also included in both dimensions in Table 2.)

Attractiveness	Meaningfulness	Vitality
Beautiful - - - Ugly	Meaningful - - - Meaningless	New, Different - - - Common, ordinary
Pleasant - - - Unpleasant	Convincing - - - Unconvincing	Fresh - - - Stale
Gentle - - - Harsh	Important to me - - - Unimportant to me	Lively - - - Lifeless
Appealing - - - Unappealing	Strong - - - Weak	Colorful - - - Colorless
Attractive - - - Unattractive	Honest - - - Dishonest	Sharp, bright, clear - - - Washed out
In good taste - - - In poor taste	Easy to remember - - - Hard to remember	
Exciting - - - Unexciting	Easy to understand - - - Hard to understand	
Interesting - - - Uninteresting	Worth remembering - - - Not worth remembering	
Worth looking at - - - Not worth looking at	Simple - - - Complicated	
Comforting - - - Frightening		
Colorful - - - Colorless		
Fascinating - - - Boring		

Table 2. Dimensions and items in the Print Ad Reaction Profile (Wells, 1964).

Dimensions similar to those in the study by Wells (1964) have been found in later research. In a study by Wells et al. (1971) a list of about 1,000 items was reduced to 30 items in six dimensions by means of several data collections and a series of factor analyses. The study was based on TV ads but the items and factors were similar to those in the earlier study. The *attractiveness* dimension was not found with TV ads, but the 1971 study revealed a dimension labeled *personal relevance* that is similar to the *meaningfulness* dimension and a dimension labeled *vigor*, similar to *vitality*. The dimensions and items in the study by Wells et al. (1971) are displayed in Table 3, which also shows dimensions and items from four other studies.

Wells et al. (1971)	Aaker & Bruzzone (1981)	Biel & Bridgwater (1990)	Aaker & Stayman (1990)	Bruzzone & Tallyn (1997)
Humor Jolly, Merry, Playful Humorous, Amusing	Entertaining Clever, Imaginative, Amusing, Original, Uninteresting, Lively, Dull, Easy to forget	Ingenuity Clever, Imaginative, Amusing, Original, Silly, Dull	Amusing/clever Imaginative, Clever, Original, Amusing	Humor Amusing, Clever
Uniqueness Novel, Imaginative, Ingenious, Unique, Original				Uniqueness Imaginative, Original
Personal Relevance Important to me, Meaningful to me, For me, Worth remembering, Valuable	Personal relevance Worth remembering, Convincing, Effective, Informative, Interesting	Meaningful Worth remembering, Effective, Easy-to-forget, Pointless, True-to-life, Believable, Convincing, Informative	Informative/effective Informative, Effective, Convincing, Worth remembering, Well done	Credible Believable, True-to-life Importance Worth remembering, Easy-to-forget
Vigor Vigorous, Exciting, Energetic, Enthusiastic, Exhilarated		Energy Lively, Fast-moving, Appealing, Well done	Lively Fast moving, Lively	Energetic Fast Moving, Lively
Irritation Terrible, Stupid, Ridiculous, Irritating, Phony	Dislike Irritating, Silly, Pointless, Phony	Rubs the wrong way Seen-a-lot, Worn out, Irritating, Familiar, Phony	Irritating/silly Silly, Irritating, Phony, Pointless	Lack of relevance Irritating, Phony, Silly
Sensuousness Tender, Gentle, Soothing, Serene, Lovely	Warm Appealing, Gentle, Well done	Warm Gentle, Warm, Sensitive	Warm Gentle, Warm, Sensitive	Warmth Warm, Sensitive, Gentle
			Familiar Seen-a-lot, Familiar	Familiarity Familiar, Seen-a-lot
			Believable True-to-life, Believable	
			Confusing Confusing	Clarity Confusing, Pointless, Informative
			Dull Forgettable, Dull, Worn-out	Freshness Worn-out, Dull
				Appeal Appealing, Well done
				Persuasive Convincing, Effective

Table 3. Dimensions and items in five adjective checklists.

Table 3 has been organized to show similarities in items and dimensions across studies. It is not surprising to see a high degree of overlap in the items in these studies: The ACLs used in the studies by Aaker and Bruzzone (1981), Biel and Bridgwater (1990), Aaker and Stayman (1990) and Bruzzone and Tallyn (1997) were the standard ACLs offered by the Bruzzone Research Company at the time of the studies. Moreover, ten items were taken from the study by Wells et al. (1971) when the Bruzzone ACL was developed (Aaker and Norris, 1982). The similarities in dimensions are more noteworthy. The studies presented in Table 3 indicate that the dimensions revealed in the factor analyses by Wells et al. (1971) are stable. The factors *humor*, *personal relevance*, *irritation* and *sensuousness* have equivalents in all four of the other studies, *vigor* in three other and *uniqueness* in one other.

Two of the studies in Table 3 used the same data but got different results. Aaker and Stayman (1990) found nine factors and Biel and Bridgwater (1990) found five factors in their analyses of the same data. The reason for this was that Aaker and Stayman used individual data, while Biel and Bridgwater used aggregate data. The input into Biel and Bridgwater's analysis was data aggregated for each ad in the study, i.e., the share of respondents who had checked a specific adjective for each ad, whereas Aaker and Stayman's input was data from each individual respondent. The difference in the results of these two studies illustrates the importance of the type of data used in a study, but it should be noted that the difference was not very large in this case. The five factors in Biel and Bridgwater's study have their equivalents in Aaker and Stayman's study and the richer variation in the individual data appears mainly to have contributed to fine tuning the results obtained from the analysis of aggregate data.

For the sake of completeness, Table 4 shows adjectives and dimensions from two studies by Leavitt (1970; 1975). The dimensions found in these studies are

similar to those found in the research discussed above. It should be noted that in the 1970 study, the adjectives in *amusing* were considered by Leavitt to belong to the *energetic* dimension and were listed separately for descriptive purposes. It should also be noted that *novel*, *authoritative* and *disliked* accounted for very little variance and that Leavitt only included these dimensions for descriptive purposes. The ACLs in Table 4 are of further interest as they have been used in two later studies discussed below.

Leavitt (1970)	**Leavitt (1975)**
Energetic Lively, Exhilarated, Vigorous, Enthusiastic, Energetic, Excited	Stimulating Amusing, Clever, Merry, Playful, Exciting, Fast-moving, Lively, Vigorous, Creative, Imaginative, Novel, Unique, Flat, Dull, Slow, Sluggish, Copy-cat, Old, Repetitious, Worn out
Amusing Merry, Jolly, Playful, Joyful, Amusing, Humorous	Relevant Convincing, Helpful, Important for me, Meaningful for me, Dependable, Frank, Wise, Worth remembering, Believable, Genuine, Natural, Realistic, In poor taste, Phony, Silly, Stupid, Confusing, Unclear, Clear, Informative
Personal relevance Important for me, Helpful, Valuable, Meaningful for me, Worth remembering, Convincing	Gratifying (Sensuous) A good world, Agreeable, Attractive, Dreamy, Sensitive, Soothing, Tender, Warm
Authoritative Confident, Business-like, Consistent-in-style, Responsible, Frank, Dependable	Familiar Familiar, Saw before, Well-known, New
Sensual Lovely, Beautiful, Gentle, Serene, Tender, Sensitive	
Familiar Familiar, Well-known, Saw before	
Novel Original, Unique, Imaginative, Novel, Ingenious, Creative	
Disliked Phony, Terrible, Stupid, Irritating, Unimportant to me, Ridiculous	

Table 4. Dimensions and items in adjective checklists developed by Leavitt (1970; 1975).

The studies that generated the ACLs in Table 2 to Table 4 were oriented toward developing universal adjective checklists applicable to all types of ads within a certain media type. An alternative suggested by Rossiter and Percy (Percy and Rossiter, 1997; 1997) is to use a customized ACL for each ad that is tested. They suggest that the ACL should reflect the creative execution tactics of the ad, i.e., quadrant in the *Rossiter-Percy grid*, and its executional devices, e.g., music or a celebrity endorser.

In the pioneering work by Wells and associates (Wells, 1964; Wells et al., 1971) ACLs were developed to be used in advertising pretests. Rossiter and Percy (Percy and Rossiter, 1997; 1997) also regard ACLs as only appropriate for pretests. A different position is taken by the Bruzzone Research Company, which uses its ACL in posttests. This use was evaluated in a pilot study reported in Aaker and Bruzzone (1981) in which the same 31 ads were evaluated in a pretest[1] study and a posttest mail study. The ACL in the studies consisted of 18 adjectives. The between-test correlation was above .73 for 14 of the adjectives. The adjective *repetitious* had a between-test correlation of .48 and the authors write, without being specific about numbers, that *vigor* is easier to capture immediately after exposure than in a mail survey. Aaker and Bruzzone also note that people tended to check negative words to a higher degree in the mail survey.

A study by Mehrotra et al. (1981) found that some of the adjectives in an ACL had a statistically significant positive correlation with pre/post differences in brand purchase intentions. The ACL comprised 22 adjectives; eight of these had a correlation between .14 and .23 with the pre/post difference in purchase

[1] The design of the study was a traditional mall pretest with forced exposure of the ads. It is not clear, however, whether or not the ads in the study had been aired prior to this pretest. Hence, it cannot be ruled out that the respondents were influenced by previous exposure to the ads.

intention. (The eight adjectives were: *meaningful*, *valuable*, *important*, *helpful*,*appealing*, *pleasing*, *unique* and *clear*.) While, these results indicate that single adjectives could have some predictive value, they are far from conclusive.

The properties of four of the ACLs discussed above have been evaluated in two studies. Zinkhan and Fornell (1985) evaluated the ACLs in Wells (1964) and Leavitt (1970) while Zinkhan and Burton (1989) evaluated the ACLs in Wells (1964), Leavitt (1975)[2] and Schlinger (1979). The purpose of both articles was to investigate whether the dimensional structure of the ACLs found in the earlier studies could be replicated and to evaluate the predictive usefulness of the ACLs. The *Print Ad Reaction Profile* in Wells (1964) was tested with print ads in both studies while the other ACLs were tested with TV ads. The dimensional structure was investigated using target analysis, a type of confirmatory factor analysis. The predictive usefulness was investigated by using logistic regression with brand choice as dependent variable and with multiple regression using either brand attitude or purchase intention as dependent variable.

The target analysis in Zinkhan and Fornell (1985) confirmed the dimensional structure in Wells (1964) but the dimensional structure in Leavitt (1970) was not confirmed. The dimensional structure in Wells (1964) was again confirmed by Zinkhan and Burton (1989) who, in addition, found that the dimensional structure in Leavitt (1975) could almost be confirmed (three out of four factors were supported) while the structure in Schlinger (1979) was not confirmed. Judging from the two studies the dimensional structure of the *Print Ad Reaction Profile* developed in Wells (1964) is stable and the dimensional structure in the other ACLs, developed for and tested on TV ads, is less stable. Whether the difference in dimensional stability depends on the media type, the ACLs in

[2] Zinkhan and Burton state that they had evaluated the ACL in Leavitt's (1970) study, but the ACL in their article is quite clearly taken from Leavitt's (1975) study.

themselves, or both, was not investigated in the two evaluative studies, but the dimensional stability in the Leavitt (1975) ACL indicates that it is also possible to achieve dimensional stability for TV ad ACLs.

The results of the regression analyses in Zinkhan and Fornell (1985) are shown in Table 5. Leavitt's (1970) *energetic* dimension consists of the adjectives from both the *energetic* and *amusing* dimensions (cf. Table 4). *Brand attitude* was operationalized with an eight-point scale measuring how good/bad the brand was and *purchase intention* with the probability of purchasing the brand in the near future. Zinkhan and Fornell ran regression analyses with both *brand attitude* and *purchase intention* as dependent variables. In both models, factor scores from the analysis of the ACLs were used as independent variables. In the models with *purchase intention* as dependent variable, *brand attitude* was included as an additional independent variable.

	Dependent variable: Brand attitude		**Dependent variable: Purchase intention**	
	R^2/Beta	$p <$	R^2/Beta	$p <$
Wells (1964)				
R^2	*.324*		*.596*	
Attractiveness	.546	.001	.292	.001
Meaningfulness	.144	.001	.042	.20
Vitality	.008	.90	.026	.45
Brand attitude	-	-	.565	.001
Leavitt (1970)				
R^2	*.315*		*.265*	
Energetic	.367	.001	.111	.05
Relevance	.128	.01	.106	.05
Familiar	–.235	.001	–.011	.85
Sensual	.311	.001	.026	.60
Brand attitude	-	-	.444	.001

Table 5. Regression models with ACL variables as independent variables (standardized coefficients), from Zinkhan and Fornell (1985).

The results in Table 5 show that factors from ACLs can contribute to explaining the variance in brand attitude and purchase intention. The only factor that was not statistically significant in explaining the variance in *brand attitude* was

vitality from Wells (1964). In the models with *purchase intention* as dependent variable, the contribution of the *ACL factors* was smaller. Only the *attractiveness* factor from Wells (1964) made any substantial contribution to explaining the dependent variable. The bulk of the explanatory power in the *purchase intention* models came from the *brand attitude* variable which had by far the largest beta coefficient in both models.

Table 6 displays the results from Zinkhan and Burton (1989). The authors dropped the *familiar* dimension from Leavitt (1975) as their target analysis indicated it was unstable (cf. Table 4). B*rand attitude* was measured with three seven-point semantic differential scales ("bad-good," "unsatisfactory-satisfactory" and "unfavorable-favorable") that were used to calculate a summed scale and it was not measured together with the Wells (1964) ACL. C*hoice behavior* was a dummy variable for a brand choice situation in which the respondents could buy the target brand, a different brand or refrain from buying at all.

	Dependent variable: Brand attitude		Dependent variable: Choice Behavior	
	R^2/Beta	$p <$	R^2/Beta	$p <$
Wells (1964)				
R^2	-		*.102*	
Attractiveness	-	-	.77	.001
Meaningfulness	-	-	.35	.05
Vitality	-	-	.25	$p > .15$
Leavitt (1975)				
R^2	*.216*		*.207*	
Stimulating	.342	.001	.127	.05
Relevant	.221	.001	.017	.90
Gratifying	.201	.001	.086	.10
Brand attitude	-	-	.399	.001
Schlinger (1979)				
R^2	*.198*		*.174*	
Entertainment	.263	.001	.166	.001
Confusion	–.272	.001	–.010	.85
Relevant news	.168	.001	.073	.15
Brand reinforcement	.138	.01	.053	.30
Brand attitude	-	-	.350	.001

Table 6. Regression models with ACL variables as independent variables (standardized coefficients), from Zinkhan and Burton (1989).

The results for the models with *brand attitude* as dependent variable in Table 6 are similar to those in Table 5. In particular, all the *ACL factors* contribute to the explanation of the variance in *brand attitude*, and the contribution of the *ACL factors* in the models with *choice behavior* as dependent variable is limited. The *attractiveness* and *meaningfulness* factors in Wells (1964) had rather large beta coefficients, but it should be kept in mind that the model did not include *brand attitude* as an independent variable as did Zinkhan and Fornell's (1985) model in Table 5. In all, variables from ACLs would appear to be of some value in predicting variables higher up in the hierarchy-of-effects, but variables such as *brand attitude* appear to make a more substantial contribution.

2.7.2 Cognitive response measurement

Cognitive response measurement, or thought listing, has mainly been used in *persuasion research*, including advertising research, and *problem-solving research* (Wright, 1980). Cognitive response measurement (CRM) was first

applied to advertising by Wright (1973) and since then many academic advertising researchers have followed suit. Today, commercial applications of CRM can be found amongst practitioners. For example, *Research International* in the US has developed an advertising pretest method called *Cognitive Response Analysis*SM (CRASM) (Lebenson and Blackston, 1997).

Cognitive response measurement is believed to mirror the thoughts people have in response to messages (Wansink et al., 1994). These thoughts are believed to be prominent in both short-term and long-term preference judgments. In the short term, they are believed to be available in memory and in the long term they are believed to have a high likelihood of retrieval because of a rehearsal effect (Wright, 1980). Within advertising research, CRM has been said to measure what people do to advertising, whether they internalize the advertising message or hold it at a distance (Biel, 1996). CRM has also been regarded as useful for advertising research as it provides information about individuals' idiosyncratic responses to messages (Nelson, Shavitt, Schennum and Barkmeier, 1997).

Biel (1996) suggests that cognitive response measurement is suitable for packaged goods as well as durables, retailers and services. This means that CRM could be useful for both low- and high-involvement products. Another suggestion is that CRM is mainly relevant for high-involvement products (Percy and Rossiter, 1997). Since the literature review did not reveal any empirical tests of whether CRM is more or less useful for different types of products, the open question of whether any differences exist, remains unanswered.

Cognitive response measurement is performed by asking the respondent to write down, or report orally, her or his thoughts during and immediately after exposure to an advertising message. The instructions to the respondent can vary. Wright (1980) distinguishes between *general* instructions, which probe for all

sorts of thoughts, and *directive* instructions, which probe for certain types of thoughts, e.g., counterarguments. It is also possible to use directive instructions which probe for thoughts *and* other types of mental activities. Certain studies include instructions aimed at *thoughts* and *feelings* (e.g., Crites et al., 1994; Wansink et al., 1994) and others use instructions aimed at *thoughts, ideas* and *reactions* (e.g., Shavitt and Brock, 1990; Biel, 1996; Lebenson and Blackston, 1997). Rossiter and Percy (1997) suggest that the instructions should probe for *thoughts, feelings* and *mental images*. The instructions in cognitive response measurement can also be more or less directed with respect to the object of the thoughts and other mental activities. Examples include *general* instructions without any specific object, (e.g., Shavitt and Brock, 1990; Lebenson and Blackston, 1997), having the *product* and *advertising message* as specified objects (e.g., Wright, 1973) and specifying the *brand* and *eating the brand* (Wansink et al., 1994).

Both general and directive instructions carry a number of possible disadvantages. General instructions could lead respondents to directing themselves, e.g., reporting only thoughts that they think the researcher is interested in, while directive instructions can be subject to misinterpretation or influence the respondents' mental activities (Wright, 1980). Whether directive instructions bias the responses in CRM, i.e., whether they are *reactive* or *non-reactive*, was investigated by Wansink et al. (1994). In their study, they did not find any differences in measures of *brand attitude* and *attitude toward the ad* between a group that had received directed instructions and a control group that had received general instructions. The study also showed that the directive instructions had the advantage of generating a larger number of responses on the target issue, i.e., brand usage, than the general instructions did.

It is common to limit the time the respondents have available to write down their cognitive responses. The idea behind limiting the time is to get more message-evoked thoughts, which are fairly quick to retrieve or complete, and fewer thoughts made up from scratch, a process which takes more time (Wright, 1980). A common time-limit is three minutes (e.g., Wright, 1973; Nelson et al., 1997). The responses are normally captured on a sheet of paper with boxes and the respondents are instructed to write one thought in each box (cf. Cacioppo and Petty, 1981). An alternative way is to capture the cognitive responses as a block of text which is divided into separate thoughts according to certain criteria, and by someone other than the respondent (e.g., Stephens and Russo, 1997). The measurement of cognitive responses should be carried out immediately after exposure to an advertising message (cf. Wright, 1980; Cacioppo and Petty, 1981), which means that CRM is applicable only to advertising pretests.

A central part of cognitive response measurement is the coding of the responses. The responses can be coded into content categories, i.e., based on what the thought is about, and/or based on the evaluative nature of the thought, i.e., whether the thought is positive, negative or neutral (cf. Shavitt and Brock, 1990). Content categories that have been used in studies include *counterargument, support argument or source derogation* (e.g., Wright, 1973), *self-related or not self-related* and *product-, communicator- or execution-related* (e.g., Biel, 1996; Lebenson and Blackston, 1997), *thought or feeling* (e.g., Crites et al., 1994), *utilitarian or social identity* (e.g., Nelson et al., 1997) and *single-fact interpretations, abstractions, or global evaluations* (e.g., Chattopadhyay and Alba, 1988). It has also been suggested that mention of the correct brand name should be coded within advertising research (Percy and Rossiter, 1997).

The evaluative nature of the responses, sometimes referred to as their *valence*, has been coded as positive or negative (e.g., Biel, 1996), positive, negative or neutral (e.g., Crites et al., 1994; Lebenson and Blackston, 1997), or on 5- or 7-point scales with endpoints such as "very unfavorable/very favorable" and "very negative/very positive" (e.g., Nelson et al., 1997; Stephens and Russo, 1997). The coding of the evaluative nature of the responses can be done by one or several judges or, as is common, by the respondents themselves.

Coding responses to CRM involves making a subjective judgment of how to classify each response. One way to reduce bias in the coding is to use several judges and measure the reliability of the coding in terms of inter-judge reliabilities (cf. Wright, 1980; Rust and Cooil, 1994). This need, of course, not be done if the respondents code their own responses.

Two different coding schemes were compared in two experiments by Stephens and Russo (1997). The responses in their study were coded by two judges. The first scheme, taken from Batra and Ray (1986), was rather complex and included 14 content categories, e.g., support arguments and counterarguments, various execution-related categories and a number of personal association categories. The judges agreed on 64 and 69 percent of the responses in the two experiments. In the second scheme, the judges coded the evaluative nature of the responses on a 7-point scale with end-points -3 ("very negative") and +3 ("very positive"). For exactly equal scale values, the judges agreed on 59 and 48 percent of the responses. When the judgments were treated as equal if the scale values were within one unit, agreement was 93 and 94 percent. The low inter-judge agreement (cf. Rust and Cooil, 1994) for the exact coding of the two schemes indicates that reliability in the coding of CRM responses could constitute a problem.

Stephens and Russo (1997) also investigated the explanatory power of the two coding schemes in their study. They ran regression analyses with *attitude toward the ad* and *brand attitude* as dependent variables and responses coded according to the two schemes as independent variables. The analyses showed that the R^2s were the same for the two coding schemes, i.e., the two coding schemes had the same explanatory power. Stephens and Russo also used data from a third experiment to compare the explanatory power of self-coded evaluative responses and evaluative responses coded by judges. In the experiment both the respondents and two judges coded the responses on a 7-point evaluative scale. Regression analyses with *attitude toward the ad* as dependent variable resulted in an R^2 of .69 for self-coded responses and an R^2 of .49 for judge-coded responses. With *brand attitude* as dependent variable, the R^2 was .65 for self-coded responses and .53 for judge-coded responses.

The CRM responses can be transformed into quantitative variables in a number of ways. Lebenson and Blackston (1997), for example, used an aggregate (for each ad) database with the average number of responses per respondent and the percentage of responses in each response category in their analyses. An alternative that yields individual data is to calculate the *net cognitive response* for each individual respondent (cf. Crites et al., 1994; Rossiter and Percy, 1997). The net cognitive response is calculated by subtracting the number of negative responses from the number of positive responses.

The relationship between CRM responses and other measures have been investigated in a number of studies. The results from three regression analyses in Wright (1973) are shown in Table 7. The dependent variables were a question measuring whether the product arguments in the ad were convincing, brand attitude and buying intention. The independent variables were three content

categories of CRM responses (counterargument, support argument and source derogating) and two measures of message reception.

	Dependent variable: Convincing arguments	Dependent variable: Brand attitude	Dependent variable: Buying intention
	R^2/Beta	*R^2/Beta*	*R^2/Beta*
R^2	*.373*	*.301*	*.232*
Counterargument	–.452	–.422	–.387
Support argument	.234	.231	.167
Source derogating	–.197	–.141	–.065
Reception (free-response)	–.024	–.030	–.115
Reception (multiple-choice)	.058	.042	.056

Table 7. Regression models with CRM variables as independent variables (standardized coefficients), from Wright (1973).

The results in Table 7 show that quantified CRM responses are related to other measures, but that the relationship is different for the content categories. Wright (1973) concluded that counterargument was by far the best predictor of the three CRM response types.

Lebenson and Blackston (1997) ran correlations between CRM responses and *liking the ad*, *brand attitude* and *purchase intentions*. In their analysis, they used data aggregated over ads from a commercial pretest database and analyzed ads for packaged and non-packaged goods (durables and services) separately. Tests of 139 ads for packaged goods and 30 ads for non-packaged goods were conducted. The correlations are shown in Table 8.

	Percent positive response		Percent self-relevance	
	Packaged	*Non-packaged*	*Packaged*	*Non-packaged*
Liking	.40	.77	.54	.19
Brand attitude	.47	.52	.52	.33
Purchase intention	.57	–.36	.40	–.03

Table 8. Correlations between cognitive response measures and three intermediate effectiveness measures for two types of products (from Lebenson and Blackston, 1997).

The correlations in Table 8 indicate a positive relationship between the two CRM measures and *liking the ad*, *brand attitude* and *purchase intention*. They also give a weak indication of differences in the relationships between packaged and non-packaged goods; the correlations between the CRM measures and purchase intention differed for the two product groups. The relationship was negative or non-existent for non-packaged goods while it was positive for packaged goods. The authors do not provide any detailed information on how the purchase intention question was worded, but they write that the question was adapted to suit the type of product. Differences can be seen in the relationship between the other measures for the two product groups, but these were smaller and the coefficients all have the same sign. Given the sample size of 30 ads for non-packaged goods, these other differences seem less interesting.

The temporal stability of cognitive responses was investigated in a study by Shavitt and Brock (1990). They recontacted respondents one week after they had participated in an ad pretest and asked them to recall as many of their thoughts during the test the previous week as possible. For the seven ads in the study, the respondents recalled slightly more than half (53 %) of the CRM responses. There was some variation between the ads, from 37 to 63 percent and also some variation in the recall rate of different content categories. Self-thoughts had the highest recall rate while product thoughts had the lowest; evaluative thoughts were recalled to a greater extent than neutral thoughts.

Several possible problems are associated with cognitive responses measurement. One of these is that CRM requires the respondents to be articulate (cf. Schlinger, 1979). That this really is a problem is not indicated in any of the reviewed studies, whose respondents have included "female heads-of-household" (Shavitt and Brock, 1990), "adult women" (Wright, 1973) and "members of the Parent Teacher Association" (Wansink et al., 1994). Another possible problem is that

the data can be burdensome to analyze (Wright, 1980), the burden being the coding of the responses. Yet, as suggested by the results in Stephens and Russo (1997), this needn't be the case. In their study self-coding of the evaluative nature of the responses had the best explanatory power (cf. Percy and Rossiter, 1997; Rossiter and Percy, 1997). If the qualitative properties of the CRM responses are put aside, self-coded CRM data could be treated as quantitative data, and this would entail less burdensome analysis. The self-coding in itself, in the context of an advertising pretest, is not complicated to perform, and consequently, CRM could be used as a rather handy market research tool.

2.7.3 ACL or CRM, or both?

The literature review showed that a considerable amount of research has been done on adjective checklists as well as cognitive response measurement. The literature review indicates that both types of measures might be important in pretesting advertising, but gives no indication as to whether one of the measures is more useful than the other.

The results in studies which related ACLs factors or quantified CRM responses to other advertising measures (e.g., Wright, 1973; Zinkhan and Fornell, 1985; Zinkhan and Burton, 1989; Lebenson and Blackston, 1997) showed that variables from ACLs or CRM contributed to explain the variance in, e.g., brand attitude in regression analyses. This indicates that both ACLs and CRM measure some underlying construct which is related to constructs later in the hierarchy-of-effects. Whether this is the same construct, acceptance or a different construct, is not entirely clear, since the theoretical background of the two measures was disregarded or discussed briefly in the studies reviewed. The articles that discussed why cognitive responses were important (e.g., Wright, 1980; Wansink et al., 1994) mentioned that they mirror the responses people

have to advertising, which are believed to be important since they will be easily available from memory when attitude or preferences are formed. The articles on ACLs generally had little, if anything, to say about why the responses measured by ACLs were important.

Given that the objective of the two measures is to capture responses to advertising, it could be argued that adjective checklists and cognitive response measurement measure the same underlying construct(s). Both measures are generally administered immediately after exposure to an ad and the instructions for the two measures can be fairly similar. The great difference lies in the format for capturing the responses: Adjective checklists have set response alternatives and cognitive response measurement is open-ended. In all, it appears that ACLs and CRM could measure the same underlying construct(s), provided that the instructions preceding the two measures are similar. It also seems likely that ACLs and CRM measure other underlying constructs in addition to acceptance, since both measures typically probe for general responses to advertising.

All the ACLs included in the research reviewed above were designed as standard ACLs intended to be applicable to a large number of ads, and the items in them were all related to the ad, not the product. The use of exclusively ad-related items, i.e., no brand-related items, distinguishes ACLs from CRM. CRM has the potential to capture both ad- and brand-related thoughts or feelings, unless the instructions are explicitly directed to only one of the two types of responses. According to Rossiter and Percy (1997) ACLs should be tailor-made for each ad. This would increase their potential to capture a greater variety of responses, but also increase the workload associated with using ACLs and decrease the comparability across studies.

Both ACLs and CRM provide qualitative information on individual reactions to advertising (cf. Aaker and Bruzzone, 1981; Nelson et al., 1997), albeit in a different format. The responses captured by ACLs is structured and can easily be quantified, while the responses from CRM are unstructured, making them potentially richer in informative content, but also more difficult to quantify. The qualitative information can enable an understanding why an ad works in a certain way, or, perhaps, why it is not working in the way it was intended. The qualitative information can also be used in developing new advertising campaigns.

The literature review did not reveal any research in which a comparison has been made of adjective checklists and cognitive response measurement as advertising pretest measures. The two types of measures could be compared in empirical research. If the same ads are included in pretests administered to two groups, one with a pretest questionnaire with an ACL and the other with a pretest questionnaire with CRM, the results in the two groups could be compared. An alternative procedure would be to include both an ACL and CRM in a pretest questionnaire, which would allow for an individual level comparison of the two measures.

2.8 Ad credibility

The *ad credibility* construct is not an established intermediate measure of advertising effectiveness, but it is included in the present context because of the results in a study by MacKenzie and Lutz (1989). These authors defined ad credibility as the extent to which claims about a brand in an ad are perceived by consumers as truthful and believable, and they included the construct as a possible antecedent to attitude toward the ad in their study. MacKenzie and Lutz measured ad credibility using three semantic differential scales: "convincing -

unconvincing," "believable - unbelievable," and "biased - unbiased." The first two of these items are common in adjective checklists (e.g., Leavitt, 1975; Biel and Bridgwater, 1990; Bruzzone and Tallyn, 1997).

The results of the study by MacKenzie and Lutz (1989) showed that ad credibility influenced not only attitude toward the ad but also brand attitude. This result indicates that ad credibility could be useful as an intermediate measure of advertising effectiveness, and hence its inclusion in the present study.

2.9 Attitude toward the ad and liking the ad

Attitude toward the ad was introduced as an important construct in the academic world in the early 1980s (Brown and Stayman, 1992). Among practitioners *liking the ad* gained momentum following the results from the ARF Copy Research Validity Project (Haley and Baldinger, 1991) presented in the early 1990s, but it had been used for some time before that. The measurement of attitude toward the ad and liking the ad is similar: In both cases the aim is to elicit an evaluative judgment of the ad, but there are also some differences which will be discussed in this section. In the six-step effects sequence, measures of liking the ad are related to the *processing* step (Rossiter and Percy, 1997).

Attitude toward the ad, or A_{Ad}, has been defined as an evaluative or affective response to a particular ad at a particular point in time (MacKenzie and Lutz, 1989). In advertising theory, attitude toward the ad is seen as a causal mediator of brand-related responses to advertising (MacKenzie et al., 1986). The interest in measures of liking the ad can be explained by its possible predictive abilities. Theories of why liking the ad is important, are generally in line with the theories

of attitude toward the ad. Articles about liking the ad (e.g., Greyser, 1973; Biel and Bridgwater, 1990; Haley and Baldinger, 1991; Hollis, 1995) indicate that the underlying construct is the same as for attitude toward the ad. The only difference between liking the ad and attitude toward the ad, apart from the name, lies in how they are operationalized in measurement.

The theoretical proposition that liking an ad, or attitude toward an ad, mediates brand-related responses might not be true for all types of ads. According to Rossiter and Percy (1997), liking the ad is only essential for *low-involvement transformational* brand strategies. Liking the ad is desirable for *high-involvement transformational* brand strategies, but in this case there must be brand-focused liking as well. For low-involvement and high-involvement *informational* brand strategies, liking of the ad is not necessary since brand attitude is thought to be based mainly on information and not on (positive) emotions related to the ad or the brand, as is the case for transformational ads.

Among academics, attitude toward the ad is usually measured with a question probing for the respondent's overall reaction to the ad. Responses are measured on a number of semantic differential items. It is probably most common to use three items to measure responses (e.g., MacKenzie and Lutz, 1989; Lord, Lee and Sauer, 1995), but some studies use two items (e.g., MacKenzie et al., 1986; Pieters and de Klerk-Warmerdam, 1996), and others use four (e.g., Gardner, 1985) or even ten items (e.g., Madden, Allen and Twible, 1988). Typical items include "good - bad," "pleasant - unpleasant," "favorable - unfavorable," and "uninteresting - interesting." The responses on the semantic differential items are usually used to form a summary index. Among practitioners, on the other hand, measurement is typically carried out using one single question about how much the respondent liked or disliked the ad. Response is measured with a single item with end-points such as "I liked it very much" and "I disliked it very

much" (cf. Biel and Bridgwater, 1990; Haley and Baldinger, 1991). Further, academics usually measure attitude toward the ad in experimental settings, similar to pretests, whereas practitioners mainly measures liking the ad in pretests, but to some extent, also in posttests.

A meta-analysis of attitude toward the ad in Brown and Stayman (1992) revealed a number of statistically significant relationships with other intermediate variables. Brown and Stayman used results from 47 samples in 43 articles to calculate correlations between attitude toward the ad and *feelings* (r = .54), *ad-related cognitions* (r = .55), *brand-related cognitions* (r = .33), *brand attitude* (r = .67) and *purchase intentions* (r = .43). These correlations, the authors concluded, suggest that attitude toward the ad is important in explaining the effects of advertising.

In a study by Lebenson and Blackston (1997) correlations between liking the ad and *brand attitude* and liking the ad and *purchase intention* were calculated for packaged goods and for durables and services. The correlation between liking the ad and brand attitude was .74 for packaged goods and .57 for durables and services. The correlation between liking the ad and purchase intention was .63 for packaged goods and –.28 for durables and services. These results indicate that liking the ad could be more influential for certain types of ads, e.g., ads for low-involvement products, than for others.

One model that has been suggested to describe the causal relationship between attitude toward the ad and brand attitude is the *dual mediation hypothesis* (MacKenzie et al., 1986). According to this hypothesis, attitude toward the ad has both a direct influence on brand attitude and an indirect influence via brand cognitions (also referred to as brand beliefs or brand benefit beliefs, see Section 2.11). There was some support for the dual mediation hypothesis in Mackenzie

et al. (1986), but a later study by MacKenzie and Lutz (1989) found support for a causal relationship between attitude toward the ad and brand cognitions and attitude toward the ad and brand attitude, but not between brand cognitions and brand attitude. Based on these studies and a number of others, Brown and Stayman (1992) argue that there is support for the dual mediation hypothesis. According to Brown and Stayman, the dual mediation model holds, but the relationship between brand cognitions and brand attitude is modest, and the direct effect of attitude toward the ad on brand attitude is probably not as strong as previously suggested.

The ARF Copy Research Validity Project (Haley and Baldinger, 1991) concluded that liking the ad was better at prediction of sales than any other pretest measure included in the study. The basis for this conclusion was that the liking the ad measure correctly predicted which ad in a pair of ads would increase sales significantly more than the other in 87 percent of the cases in the study.

The results and conclusions of the ARF study have been challenged. According to Rossiter and Eagleson (1994), the results of the ARF Study are extremely difficult to interpret, and they contend that most people have made superficial interpretations of the results. Rossiter and Eagleson presented their own analysis of the results from the ARF Study and they showed that if random variation in the data was properly accounted for, it was not possible to conclude that liking the ad was the single best measure. They also added results from two later studies and found that measures of liking the ad performed worse than measures of pre/post persuasion. The measures of pre/post persuasion performed well enough to reach an accuracy of 82 percent, while liking the ad reached only 50 percent accuracy (i.e., equal to chance). Similar results were presented by Blair and Rosenberg (1994), who combined results from studies by *research system*

corporation and the ARF Study. They found that liking the ad predicted 41 percent correct, recall 47 percent, brand name recall 44 percent and the ARS Persuasion score 100 percent (Blair and Rosenberg, 1994: 37).

2.9.1 One theoretical construct - two measures

The literature review on *attitude toward the ad* and *liking the ad* indicates that these two terms refer to the same underlying construct. The main difference appears to be that attitude toward the ad tends to be measured with multiple items and liking the ad with single items. Attitude toward the ad is theoretically important as a mediator of brand responses to advertising (MacKenzie et al., 1986), but it might not be important for all types of ads (Rossiter and Percy, 1997). In empirical studies, attitude toward the ad and liking the ad have been found to be related to other intermediate effectiveness variables. There is some dispute regarding the ability of liking the ad to predict the sales effectiveness of ads, with some authors arguing that liking the ad is the single best predictor of sales (Haley and Baldinger, 1991) and others arguing better measures exist (Blair and Rosenberg, 1994; Rossiter and Eagleson, 1994). Liking the ad is a processing measure and as such mainly relevant in pretests (cf. Rossiter and Percy, 1997).

One question is whether measures of attitude toward the ad and liking the ad are interchangeable. The literature review did not reveal any studies in which this has been tested with individual level data. Brown and Stayman (1992) analyzed whether using single- or multiple-item measures of attitude toward the ad had any significant effect on correlations between attitude toward the ad and other variables. They found a small, but not statistically significant effect in their meta-analysis on the correlation between attitude toward the ad and brand attitude, but not on any other correlation in their study. It would be possible to

investigate the interchangeability of measures of attitude toward the ad and liking the ad by including measures of both in the same study. This would make it possible to use individual level data to calculate correlations between the measures and to compare how they behave in their relationships with other variables.

2.10 Brand awareness: Brand recognition and brand recall

The discussion of the communication effects measures will begin with measures of *brand awareness*. According to Rossiter and Percy (1997) a buyer is aware of a brand if she or he can recognize or recall it, within a product category, in enough detail to make a purchase. Brand awareness is a prerequisite of purchase of a brand and necessary for the formation of brand attitude (Rossiter and Percy, 1997). Measures of brand awareness are in many respects similar to the two measures of ad awareness, i.e., ad recognition and ad recall, which were discussed previously (Section 2.6). Measures of brand awareness are, in the same way as ad awareness, memory tests (cf. Lucas and Britt, 1963; Mehta, 1994) and a distinction is made between brand recognition and brand recall.

Measures of brand awareness are included in many standardized tracking studies, e.g., the tracking offered by *Millward Brown* (e.g., Brown, 1985), but little has been published in recent years on using measures of brand awareness as measures of advertising effectiveness. During the last ten years, the focus has been on brand awareness as a measure of *brand equity* (e.g., Keller, 1993; Aaker, 1996) rather than as a measure of advertising effectiveness. Given the importance of brand awareness, which Rossiter and Percy (1997) regard as a universal communication objective, it feels relevant to include measures of brand awareness among intermediate measures of advertising effectiveness.

An important distinction regarding brand awareness is that between *brand recognition* and *brand recall*. Brand recognition is relevant for products that are chosen at the point of purchase, whereas brand recall is relevant for products chosen prior to the point of purchase (Rossiter and Percy, 1997). Consequently, measures of brand recognition and measures of brand recall are not interchangeable and the brand awareness measure must be chosen to fit the communication objectives of the brand.

Measures of *brand recall*, sometimes referred to as *top-of-mind awareness* or *unaided recall*, seek to ascertain whether respondents can *recall* the brand name, package or logo. The measurement is made by asking respondents what brand name(s) comes to their minds when the product category or category need is mentioned (Reynolds, Olson and Rochon, 1997; Rossiter and Percy, 1997).

When measuring *brand recognition*, the aim is to measure whether respondents can *recognize* the brand and/or package. This is done by showing pictures of packages or logotypes and asking the respondents which pictures they recognize. A somewhat simpler way is to present a list of brand names and ask the respondents which brands they recognize. Rossiter and Percy (1997) recommend that brand recognition measurement should strive to simulate as closely as possible recognition situations that occur in the real world. This means that brands that are normally displayed on their own, e.g., a fast-food restaurant situated alongside a road, should also be exposed alone in brand recognition measurement, and that brands that are normally displayed in a competitive environment, e.g., products in a supermarket, should be exposed alongside other brands. Rossiter and Percy also discuss the less common *auditory brand recognition*, which occurs, for example, when a waiter reads out a list of beer brands. Auditory brand recognition should be measured by reading

out brand names, e.g., in a telephone interview, and asking respondents which names they recognize.

Factors that have been shown to influence the level of brand awareness include media exposure, media frequency and the creative execution of the ad. An analysis of tracking data for 29 advertising campaigns found a statistically significant positive relationship between GRP and brand awareness (Batra et al., 1995). Blair (1987), too, reports a positive relationship between GRP and brand awareness, and a relationship that is stronger for ads with higher ARS persuasion scores in pretests. (Interestingly, neither Batra et al. nor Blair specifies whether they measured brand recognition or brand recall.) Singh and Cole (1993) found that the length of television ads influences brand name recall; brand name recall was statistically significantly higher for 30-second ads than for 15-second ads. Singh and Cole also found that repetition had a positive effect of on category-prompted brand recall and that the repetition effect was stronger for short ads.

Little research appears to have been done on the relationship between brand recognition or brand recall and other intermediate variables. One reason for this could be that it is difficult to include measures of brand awareness in pretests or laboratory experiments with pretest-like conditions. The measurement of brand awareness must be delayed (Rossiter and Percy, 1997), and therefore adds to the workload of a study.

2.10.1 Concluding remarks on measures of brand awareness

The literature review indicated that measures of brand awareness are discussed more often in the context of brand equity than in the context of advertising

effectiveness. There seems to be little current controversy, if any at all, concerning how brand recognition or brand recall should be measured.

In the studies by Batra et al. (1995) and Blair (1987), the authors reported that they had measured brand awareness without specifying whether they had measured brand recognition or brand recall. The same tendency can be found in the case histories of the 1996 British *Advertising Effectiveness Awards* (Duckworth, 1997). In four case histories, the authors wrote about brand awareness without specifying whether it was brand recall or brand recognition. This is unfortunate given that brand recognition and brand recall are related but different communication objectives and that measures of one cannot replace the other (Rossiter and Percy, 1997). To avoid comparisons of two measures that cannot be compared, authors, academic and practitioners alike, should specify the type of brand awareness measured in a study.

In the empirical studies in this thesis, measures of brand awareness will be included, but due to the nature of the studies, they are not of major importance.

2.11 Brand benefit beliefs

Beliefs about an object are generally regarded as central for the formation of the attitude toward that object (cf. Eagly and Chaiken, 1993), and in most attitude research during the 1960s and 1970s, beliefs were seen as the only or the most important antecedent of attitudes (Fishbein and Middlestadt, 1995). In marketing, the beliefs people hold about brands, generally referred to as *brand beliefs* or *brand benefit beliefs*, are of central importance as an antecedent of brand attitude (Mittal, 1990; Rossiter and Percy, 1997).

Beliefs about a brand, or other objects, are generally believed to have two components (Eagly and Chaiken, 1993; Rossiter and Percy, 1997). The first is the subjective probability that the brand has a certain attribute, or delivers a certain benefit, and the second is the evaluation of that attribute or benefit. For example, a benefit of a toaster could be that it toasts quickly. The attitude toward a toaster brand would then partly be a function of the probability that the toaster is quick and an evaluation of how good or bad this is.

It is generally expected that a limited number of attributes, or benefits, are important, or salient, for the attitude toward a brand. The question is: Which benefits are important for a certain product category or a certain brand? Fishbein and Middlestadt (1995) suggest that the attributes that are salient in a population should be identified by asking a sample to list characteristics, qualities and attributes of the object under study. Attributes mentioned by more than ten percent of the sample should then be included in measurement of beliefs about an object.

On the issue of measurement, Fishbein and Middlestadt (1995) argue that measurement should capture both the subjective probability that an object has an attribute and the evaluation of the attribute. This should be done by measuring the *belief strength*, i.e., the strength of the belief that links an object to an attribute, and the evaluation of each salient attribute. Belief strength can be measured with, e.g., 7-point scales with end-points *likely* (+3) to *unlikely* (-3). Measurement of the evaluation can also be made on a 7-point scale, but with end-points *good* (+3) and *bad* (-3).

Rossiter and Percy (1997) suggest that it is sufficient to measure only the belief strength, but they argue that measurement of brand benefit beliefs should be adapted to the quadrants in the *Rossiter-Percy grid* that the brand attitude

strategy belongs to. For *low-involvement/informational* ads, the measurement can be made dichotomously, simply by asking respondents whether they perceive the brand to possess a certain benefit or not. For *high-involvement/-informational* ads, the degree of benefit delivery is important and, as a consequence, the measurement should be made with graded scales measuring the extent to which the respondent perceives the brand to deliver on the benefits. For ads in *the low-involvement/transformational* quadrant, the measurement should be made with several semantic differential scales for each benefit. For the *high-involvement/transformational* quadrant, the measures should be adapted to the specific product, and either likert-type scales or semantic differential scales would be appropriate.

Little research has been done that identify which factors influence brand benefit beliefs in an advertising context. It has been suggested that attitude toward the ad could influence brand beliefs (e.g., Mittal, 1990; Fishbein and Middlestadt, 1997), but generally brand benefit beliefs are simply assumed to be an outcome of exposure to an ad, without any explanation of how this happens.

Results from empirical studies have shown that brand benefit beliefs, following exposure to advertising, have had an influence on brand attitude. Gardner (1985), for example, found a positive relationship between brand beliefs and brand attitude. This result was found under both a brand processing set, in which the processing of the ad was oriented toward the brand advertised in the ad, and a non-brand processing set, in which the processing of the ad was oriented toward other factors. Mittal (1990) divided brand beliefs into "utilitarian" and "image" beliefs, and his analyses found that both types of beliefs were positively related to brand attitude.

2.11.1 A forgotten measure?

Surprisingly little research has been conducted on brand beliefs in an advertising context during the last 20 years. A likely reason is that the focus has shifted to other factors such as attitude toward the ad or peripheral cues, that have been considered as important in attitude formation. It has been argued that the importance given to these other factors has been exaggerated as brand beliefs have been completely left out or measured in an inappropriate way (Fishbein and Middlestadt, 1995; 1997). This argument is supported by empirical results. Results presented in Gardner (1985) and Fishbein and Middlestadt (1995) show that the influence of other factors, e.g., brand attitude, is inflated if brand beliefs are left out. This highlights the importance of including measures of brand beliefs when measuring the effectiveness of advertising.

Given their importance in the formation of brand attitude, it is somewhat surprising that little appears to be known about which factors influence, mediate or moderate brand beliefs in an advertising context. If, for example, a positive attitude toward the ad, or perceived credibility in the ad, facilitates the formation of brand benefit beliefs that are in line with the advertiser's intentions, then the effectiveness of advertising, from an advertiser viewpoint, could be increased by adjusting ads accordingly.

2.12 Brand attitude

Brand attitude is a central construct in most areas of marketing. The construct is important since brand attitude is generally seen as an antecedent of both intention to use or purchase a brand and actual usage or purchase of a brand (e.g., Howard, 1989; Engel et al., 1995). Rossiter and Percy (1997) regard brand attitude, along with brand awareness, as a universal communication objective of all advertising campaigns.

Attitudes are generally regarded as an overall evaluation of an object that can be positive or negative (cf. Eagly and Chaiken, 1993). In line with this, brand attitude is generally defined as an overall, positive or negative, evaluation of a brand (e.g. Engel et al., 1995). In the theories of attitude formation that were dominant in the 1960s and 1970s, attitudes were seen as a function only of cognitive beliefs and the values associated with those beliefs, but in the 1980s, theories emerged that rejected purely cognitive attitude formation (Fishbein and Middlestadt, 1995).

A well-known example of a purely cognitive model is the *expectancy-value model* developed by Martin Fishbein (cf. Eagly and Chaiken, 1993; Fishbein and Middlestadt, 1995). According to this model, attitudes are a function of beliefs about an object and the evaluative aspects of the beliefs. The beliefs are the subjective probabilities that an object has certain attributes and the evaluative aspects are the evaluations, positive or negative, of each attribute. Attitudes are the result of summing up the beliefs weighted by their evaluations.

A model of attitude formation that includes non-cognitive factors, and that has become popular within advertising research, is the *elaboration likelihood model* (e.g., Petty et al., 1983; Petty and Cacioppo, 1984). According to this model, attitude change following exposure to a persuasive message follows the *peripheral route* under low product involvement and the *central route* under high product involvement. When persuasion follows the central route, attitude change is determined by substantive arguments related to the product, e.g., information about product benefits such as speed or safety, while persuasion following the peripheral route is determined by immaterial information, e.g., celebrity endorsers or elements of the creative execution of the ad. In this model,

the central route represents the cognitive part and the peripheral route the non-cognitive part.

According to Rossiter and Percy (1997) brand attitudes are the result of *brand benefit beliefs*, which are given different emotional weights, *freestanding emotions* and a *choice rule*, which states how brands should be evaluated. The relative importance of brand benefit beliefs and freestanding emotions depends on whether the product is *high-* or *low-involvement*, but also on whether the motive for purchase is *informational* or *transformational*. The four possible combinations of involvement level and type of purchase motive make up the brand attitude strategy quadrants in the *Rossiter-Percy grid* (Rossiter and Percy, 1987; Rossiter, Percy and Donovan, 1991; Rossiter and Percy, 1997). The way brand attitudes are formed or changed is different for each quadrant of the grid and the recommendations for the creative execution of ads offered by Rossiter and Percy are different for each of the four quadrants.

The measurement of brand attitude is similar to the measurement of attitude toward the ad. Certain researchers, mainly academics, measure brand attitude with multiple-item measures (e.g., Gardner, 1985; Zinkhan and Burton, 1989; Laroche, Kim and Zhou, 1996; Kokkinaki and Lunt, 1999), while other researchers use single-item measures (e.g., Zinkhan and Fornell, 1985; Haley and Baldinger, 1991; Patzer, 1991). In multiple-item measures, responses are typically measured using semantic differential items such as "good - bad," "dislike very much - like very much," "appealing - unappealing," "attractive - unattractive," while typical end-points on single-item measures are "excellent" and "poor." Rossiter and Percy (1997) recommend that brand attitude should be measured with a *single item*, *relative* to the brand's competitors and with reference to *a particular usage situation* (and thereby to a particular purchase motivation).

The discussion of attitude toward the ad (section 2.9) showed that some theories exists which suggest that attitude toward the ad has a direct causal influence on brand attitude and that empirical results exist to supporting this proposition (e.g., MacKenzie et al., 1986; MacKenzie and Lutz, 1989; Brown and Stayman, 1992). It has been suggested that the notion of a direct causal influence between attitude toward the ad and brand attitude is wrong. On the basis of theory and empirical results, Fishbein and Middlestadt (1995; 1997) argue that any causal influence goes via brand beliefs and that the empirical results indicating a direct influence are flawed by non-existent or inadequate measurement of brand beliefs. Results in Mittal (1990) show that the strength of the relationship between attitude toward the ad and brand attitude is inflated if brand beliefs are excluded from the analysis, but Mittal does not rule out a direct causal influence.

Some empirical support can be found in the advertising literature for a positive relationship between *brand attitude* and *brand purchase intention*. Lebenson and Blackston (1997) report a correlation of .74 between the two variables for packaged goods and a correlation of .14 for non-packaged goods. An interesting result is provided by Laroche et al. (1996), who in a structural equation model found a positive relationship between brand attitude and brand purchase intentions for the focal brand, and a negative relationship between brand attitudes for competitor brands and the purchase intention for the focal brand. It was also shown earlier (see Table 5 in section 2.7.1) how an earlier study (Zinkhan and Fornell, 1985) had found a strong influence of brand attitude on purchase intention in two regression analyses.

Some doubts have been raised on the use of brand attitude as a measure of advertising effectiveness. Wansink and Ray (1992) point to the risk of ceiling effects, i.e., that brand attitudes for established brands are on such a high level

that there is no room for a positive response to advertising. This would entail that measures of brand attitude cannot be used to predict or evaluate campaign effectiveness.

2.12.1 Concluding comment on brand attitude

Brand attitude is an important construct in advertising. Brand attitudes are thought to influence purchase or usage intentions as well as purchase and usage. In purely cognitive models of attitude formation, attitudes are seen as the result of cognitive beliefs, while other models also include non-cognitive factors, e.g., attitude toward the ad. Empirical results indicate that non-cognitive factors influence brand attitude, but it has been argued that the influence could be indirect, via brand beliefs (Fishbein and Middlestadt, 1995; 1997). There is no conclusive empirical evidence as to whether the influence of non-cognitive factors on brand attitude is direct or indirect, but results generally indicate the existence of some kind of relationship.

Brand attitude has been measured with multiple-item measures and single-item measures. The literature review did not reveal any empirical studies which have included both alternative ways of measurement. As was suggested for attitude toward the ad, both types of measures could be included in an empirical study and compared with individual level data.

2.13 Brand purchase intention

Measures of purchase intentions were originally developed to supplement observable financial data, e.g., income or assets, in the prediction of consumer purchases (Juster, 1966), but they subsequently spread into a number of areas. Today, measures of *brand purchase intention* are well established as measures of advertising effectiveness. In fact, they are so well established that many

researchers who test the impact of various factors on advertising effectiveness use brand purchase intention as dependent variable, i.e., as a measure of advertising effectiveness, without offering any justification for this choice (e.g., Tripp, Jensen and Carlson, 1994; O'Mahoney and Meenaghan, 1997). Brand purchase intention, like brand attitude, is relevant in many marketing contexts and many authors note that purchase intentions are routinely used in marketing research to predict purchase of products (e.g., Kalwani and Silk, 1982; Jamieson and Bass, 1989; Morwitz, 1997). It has been noted that behavioral intentions, which include brand purchase intention, are often used as surrogates for behavior in studies, that unquestioningly assume that measures of intention and behavior are virtually the same (Belk, 1985).

Behavioral intentions are generally seen as a self-instruction or decision to act in a certain way (cf. Belk, 1985; Eagly and Chaiken, 1993). Accordingly, a brand purchase intention is a self-instruction to purchase a certain brand, but some authors also include other purchase-related behavior (cf. Rossiter and Percy, 1997). In theories of consumer behavior, brand attitude is seen as an antecedent of brand purchase intention, which in turn precedes purchase behavior (e.g., Howard, 1989; Engel et al., 1995).

There is no agreement among academics on whether purchase intention should be measured with multiple-item measures, as there appears to be regarding *attitude toward the ad* and *brand attitude.* There are articles in which a single-item was used (e.g., Morwitz, Johnson and Schmittlein, 1993; Laroche et al., 1996) and there are articles in which multiple-item measures were used (e.g., Singh and Cole, 1993; Tripp et al., 1994). Single-item measures are used and accepted probably because measures of purchase intention were widely used and evaluated (e.g., Juster, 1966; Axelrod, 1968; Clancy and Garsen, 1970) long

before the psychometric measurement tradition gained acceptance in marketing following Churchill's (1979) and Peter's (1979) influential articles.

To predict behavior as well as possible, measures of behavioral intentions should correspond with the behavior of interest with regard to target, action, context and time (cf. Eagly and Chaiken, 1993; Fishbein and Middlestadt, 1995). Suppose, for example, that the aim is to predict consumption of coffee at home during the next week. The behavioral intention question, then, should specifically ask about *consumption*, not purchase or another behavior, *at home*, not at work, at Starbucks, or elsewhere, and during the *next week*, not during an unspecified period or a period other than one week. It has also been suggested that screening respondents to make sure they actually have intentions regarding the object under study, and measuring intentions as close in time to behavior as possible, will improve measures of behavioral intentions (Belk, 1985).

Rossiter and Percy (1997) recommend that brand purchase intention be measured by asking the respondent whether he/she intends to try, buy or use (the wording is dependent on the advertiser's action objective) a certain brand to satisfy a certain category need within a certain period of time. In some cases, it can also be relevant to inquire about the amount of the brand that the respondent intends to try, buy or purchase. Rossiter and Percy recommend a four-point scale to measure purchase intentions for low-involvement products and Juster's (1966) 11-point probability scale for high-involvement products.

The theoretical importance of *brand attitude* as an antecedent to brand purchase intention was discussed earlier (section 2.12), along with the empirical results supporting this relationship. Earlier in the thesis, reference was made to studies that have found positive correlations between brand purchase intention and

cognitive responses and attitude toward the ad (e.g., Brown and Stayman, 1992; Lebenson and Blackston, 1997).

A careful analysis of the relationship between purchase intentions and actual purchase was done by Kalwani and Silk (1982). Kalwani and Silk analyzed two data sets for durable products and five data sets for packaged goods. For durable products, measurement was made with an 11-point scale; in fact, one of the data sets came from Juster's (1966) study. For packaged goods, the measurement was made with 5-point or 7-point scales. The results revealed a relationship between purchase intention and actual purchase. For durable products, about 50 percent of the variation in the purchase variable could be explained in a linear model with purchase intention as independent variable. The results for packaged goods also showed a relationship between purchase intention and purchase in this case, but a non-linear model described it better. Only the respondents who had chosen the highest scale step had a purchase probability substantially different from zero.

A relationship between purchase intention and actual purchase has also been established in later studies. A study by Morwitz and Schmittlein (1992) found a relationship between purchase intention and purchase, and that the predictive power of purchase intention can be improved by segmenting respondents on the basis of demographic variables. In another study, intention to consume was shown to contribute to the prediction of actual consumption of three foodproducts when the analysis was performed separately for light and heavy users (Wansink and Ray, 1992).

Some doubts have also been cast on the predictive abilities of behavioral intention measures. Belk (1985) reviewed 13 studies published between 1954 and 1981 that had investigated the relationship between behavioral intentions

and behavior. In most of them, the predictive accuracy (in terms of correlations or shares of correct predictions) was low and Belk concluded that measures of behavioral intentions frequently fail to predict behavior. In some of the studies, the correlation between intention and behavior was below .20, but in other studies correlations were greater than .80. Belk suggests that the problems involved in using intentions to predict behavior are the result of situational factors, context factors and by people not having relevant intentions at the time of measurement. According to Belk, some of these problems can be dealt with by improving measurement (see discussion above), but he also argues for models of consumer behavior that include factors other than behavioral intentions as antecedents to behavior.

In sum, brand purchase intention is widely used not only in advertising but also in marketing in general. The main function of brand purchase intention is as a predictor of sales. Much research was done on the measurement of purchase intention in the 1960s (e.g., Juster, 1966; Axelrod, 1968), which still is considered as being of value. It is common in academic research, for example, to use Juster's 11-point scale, but there are also examples of researchers using multiple-item measures. Empirical results support that purchase intentions can be used to predict purchase behavior, but also that the relationship is different for durables and packaged goods (Kalwani and Silk, 1982). Other results indicate that the relationship between behavioral intentions and behavior is not always strong and that other factors should be included in attempts to predict purchase behavior. The literature review did not reveal any substantial disagreements over how the measurement of brand purchase intention should be made or how the construct is related to other constructs.

2.14 Summary and measure interrelationships

This chapter started with a broad definition of advertising effectiveness as changes in variables directly or indirectly related to the profits of an advertiser. The reasons for an advertiser to measure advertising effectiveness include evaluation of the financial returns on advertising, choosing the most effective ad execution, altering elements in an ongoing campaign and to learn before future campaigns. The next step was to define intermediate measures of advertising effectiveness as measures of the mental effects advertising creates in consumers. Depending on the purpose of measuring advertising effectiveness, measurement with intermediate effectiveness measures can be made before, during and after advertising campaigns.

A brief discussion of the role of theory in advertising effectiveness measurement showed that theory, explicit or implicit, is a prerequisite for measuring advertising effectiveness and that good theoretical guidance can improve measurement. A general theory of how advertising works is the *six-step effects sequence* in Rossiter and Percy (1997), which is a hierarchy-of-effects model. The model makes an important distinction between *processing*, the immediate responses to advertising, and *communication effects*, the enduring mental associations connected to a brand.

Intermediate advertising effectiveness can be measured in *pretests*, before an ad is circulated in media, and in *posttests*, after a campaign has started. A number of methodological issues related to pretests were discussed. The first of these concerned the degree of finish of the ads in pretests. Results in earlier research with TV ads indicate that pretests of rough sketch versions of ads and pretests of finished ads can yield similar results as, particularly with regard to brand variables (Brown and Gatty, 1967; Appel and Jackson, 1975; Schlinger and Green, 1980). Little, if any, research on this issue have been done with ads for

other media types, and it is not clear whether the results would be similar with, for example, print ads.

Other methodological issues regarding pretests were also discussed. There are good reasons not to test ads in focus groups or to use continuous measurement or galvanic skin response (Percy and Rossiter, 1997; Rossiter and Percy, 1997). Earlier research also shows that when pretesting ads it is advisable to control for effects of media environment and other factors not related to the ad (Clancy and Ostlund, 1976). There are also good reasons not to pretest *transformational* ads in a pre/post design (Rossiter and Percy, 1997). A better alternative for these ads is to use an experimental design with a control group.

Few methodological issues are related to posttests in general; most of them are connected to various measures included in posttests rather than to posttests in general.

A general measurement issue concerns how many items should be in a measure for it to be valid and reliable. The trend among academics in marketing during the last twenty years has been to use multiple items, generally three or more. The influential articles by Churchill (1979) and Peter (1979) are largely responsible for this. Among practitioners, however, single-item measures are more common than multiple-item measures. A growing debate among academics is questioning whether it is always better to use multiple-item measure (e.g., Wanous and Reichers, 1996; Finn and Kayandé, 1997; Wanous et al., 1997; Rossiter, 1999).

Using the discussion of measures of advertising effectiveness in Rossiter and Percy (1997) as a starting point, 11 intermediate measures of advertising effectiveness were discussed in some detail. Six of the measures were measures

of processing (*ad recognition*, *ad recall*, *adjective checklists*, *cognitive response measurement*, *ad credibility* and *attitude toward the ad*) and five of the measures were measures of communication effects (*brand recall*, *brand recognition*, *brand benefit beliefs*, *brand attitude* and *brand purchase intention*). All five communication effects measures can be included in both pretests and posttests, ad recognition and ad recall are mainly suitable for posttests and the remaining processing measures are mainly suitable for pretests. For each of the eleven measures theory and earlier research were reviewed.

The literature review also discovered an additional processing measure, *ad credibility*, which had been found to be related to attitude toward the ad and brand attitude (MacKenzie and Lutz, 1989). It was decided to include ad credibility in the empirical studies.

The review of the intermediate effectiveness measures indicated that the research on measures of advertising effectiveness is somewhat fragmented. Most articles include only a few measures of advertising effectiveness and articles relating many different measures to each other are rare. It is also interesting to note that little research exists on which measures from a pretest are related to measures in a posttest. The literature review did not discover one single article with individual level data from the two types of tests.

An important part of the literature review was to discuss the relationship between the constructs underlying the different measures. For each construct, theory was reviewed to find out (1) which constructs could be expected to influence the construct, and (2) which constructs could be expected to be influenced by the construct. For most of the ten measures, a considerable amount of theory and earlier research was found that examined how the underlying construct is related to other constructs. However, for some of the

measures, littler earlier research existed and less theory than expected was to be had.

The expected relationships between the constructs underlying the pretest measures are summarized in Figure 1. The five pretest constructs from Rossiter and Percy (1997) are included, as well as the ad credibility construct, taken from MacKenzie and Lutz (1989). (In the model, adjective checklists and cognitive response measurement are included, although they are not constructs. Given the somewhat weak theoretical background of these measures, it was decided to include the measures and not any underlying construct.) The relationships suggested in the model are briefly outlined below.

The arrows in the figure indicate an influence of adjective checklist (ACL) or cognitive response measurement (CRM) on ad credibility ($Cred_{Ad}$), attitude toward the ad (A_{Ad}), brand benefit beliefs ($Beliefs_{Brand}$) and brand attitude (A_{Brand}). In the literature review, there were studies that found support for a relationship between CRM and A_{Ad} (e.g., MacKenzie and Lutz, 1989; Lebenson and Blackston, 1997) and other studies that found support for a relationship between ACLs or CRM and A_{Brand} (e.g., Wright, 1973; Zinkhan and Fornell, 1985; Zinkhan and Burton, 1989).

No direct indications emerged from the literature review that ACL or CRM have a direct influence on $Cred_{Ad}$ or $Beliefs_{Brand}$, but these relationships are nevertheless suggested here. Given that both ACL and CRM capture a wide range of responses to ads, it seems reasonable to expect a relationship between ACL/CRM and $Cred_{Ad}$. If ACL/CRM do have an influence on A_{Brand}, it might be mediated by $Beliefs_{Brand}$ in the same way as A_{Ad} has been suggested to influence $Beliefs_{Brand}$ (Fishbein and Middlestadt, 1995; 1997). Neither of these two relationships was tested in the reviewed literature, and it seems reasonable,

therefore, to include them tentatively in order to investigate the relationships empirically.

Some indications were found in the literature that a relationship might exist between ACL/CRM and brand purchase intention (PI_{Brand}) (e.g., Wright, 1973; Zinkhan and Fornell, 1985; Lebenson and Blackston, 1997). This relationship is not included in the model as the statistical relationships in the studies reviewed appeared tenuous: None of the studies included brand beliefs, and there are no strong theoretical reasons exist to suggest a direct relationship.

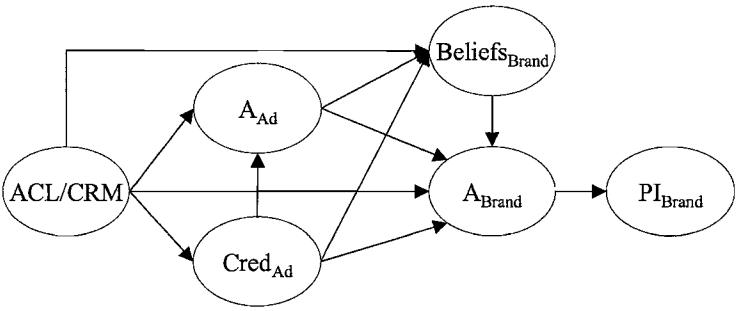

Figure 1. Expected relationships between constructs in advertising pretests.

In Figure 1, ad credibility ($Cred_{Ad}$) is expected to influence attitude toward the ad and brand attitude, as found by MacKenzie and Lutz (1989). The model also includes a relationship between $Cred_{Ad}$ and $Beliefs_{Brand}$. This relationship was included since it is reasonable to expect that the influence of $Cred_{Ad}$ on A_{Brand} might be mediated by $Beliefs_{Brand}$, but this has not been tested in earlier studies. Several studies found a relationship between A_{Ad} and A_{Brand} (e.g., MacKenzie et al., 1986; MacKenzie and Lutz, 1989; Brown and Stayman, 1992; Lebenson and Blackston, 1997), a relationship which is included in this model. Some of the reviewed studies indicates a direct influence from A_{Ad} to PI_{Brand} (e.g., Brown and Stayman, 1992; Lebenson and Blackston, 1997), but this relationship is not

included in the figure for the same reasons as the relationship between ACL/CRM and PI_{Brand} was left out, i.e., due to lack of theoretical support and the tenuous nature of the statistical relationships.

It is reasonable to expect a relationship between A_{Ad} and $Beliefs_{Brand}$ (cf. Fishbein and Middlestadt, 1995; 1997) and the relationship is included in the figure. The final relationship in the model is between A_{Brand} and PI_{Brand}. This is a well-established relationship in terms of both theory and empirical results (cf. Eagly and Chaiken, 1993; Fishbein and Middlestadt, 1995; Rossiter and Percy, 1997).

The relationships between the posttest constructs are summarized in Figure 2. The relationships on the right side of the figure, between $Beliefs_{Brand}$, A_{Brand} and PI_{Brand} are included for the same reason as the relationships between the pretest variables in Figure 1, i.e., these relationships are well established in earlier theory and research, and are not discussed any further here.

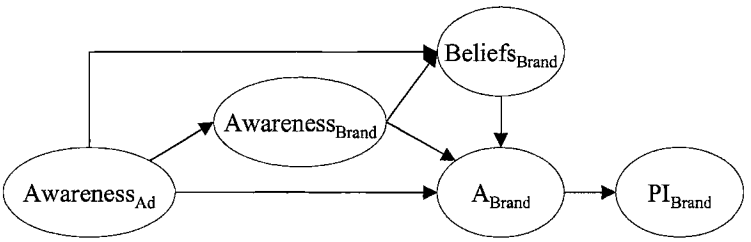

Figure 2. Expected relationships between constructs in advertising posttests.

Figure 2 suggests a relationship between ad awareness, i.e., ad recognition and ad recall, and three other constructs: brand awareness, i.e., brand recognition and brand recall, brand beliefs and brand attitude. These relationships are included since ad awareness is thought to be a posttest measure of processing, and processing of the ad is thought to lead to brand beliefs and brand attitude if the

ad works as intended (cf. Rossiter and Percy, 1997). The relationships between $Awareness_{Brand}$ and $Beliefs_{Brand}$ and $Awareness_{Brand}$ and A_{Brand} are included as brand awareness is a prerequisite for brand beliefs and brand attitude, but the relationship is not seen as causal (cf. Rossiter and Percy, 1997).

It should be noted that the relationships suggested in Figure 1 and Figure 2 are mainly relevant in an advertising measurement context. In a wider context, other variables besides those in these figures would be expected to influence, for example, brand attitude or brand purchase intention (cf. Howard, 1989; Engel et al., 1995).

The literature review did not reveal any earlier research on the ability of measures in pretests to predict the communication effects of advertising campaigns as measured in advertising posttests. One reasonable expectation is that the best predictors of the various communication effects measured in posttests are measures of the same effects in the pretest. If the communication effect measures are identical in both pretest and posttest, as recommended by Rossiter and Percy (1997), then the correspondence with respect to target, action, context and time, between the pairs of measures is more or less perfect (cf. Eagly and Chaiken, 1993; Fishbein and Middlestadt, 1995). The near perfect correspondence between the measures should contribute to an increase in the predictive ability of measures of communication effects in pretests. A factor which could contribute to decrease the predictive ability would be that communication effects measured in a pretest to a great extent are dependent on the exposure to the ad that preceded the measurement, while the communication effects measured in a posttest are more or less independent of the advertising. This would mean that communication effects measured in a pretest are related to both the ad and the brand, while the communication effects measured in a

posttest should be entirely or mainly brand-related (cf. Rossiter and Percy, 1997).

An alternative to using communication effects measures in a pretest to predict communication effects would be to use processing measures in the pretest for the same purpose. An argument for this approach would be that the formation of communication effects takes more time than the interval between exposure and measurement in a pretest. Another argument would be that multiple exposures at different points in time are required for the formation of communication effects. This could mean that information from processing measures, e.g., cognitive response measurement or attitude toward the ad, is better at predicting the communication effects that are the expected result of advertising campaigns.

2.15 Research questions for the empirical studies

In the discussion of the research problems for this thesis (Section 1.2), it was suggested that the nomological networks (Cronbach and Meehl, 1955) of the intermediate measures of advertising effectiveness were incomplete and that the interrelationships between constructs were not fully understood. The review of theory and earlier research confirms that most advertising effectiveness research has focused on one or a few measures and that very little research includes larger parts of the relevant nomological networks. The literature review also revealed some unresolved issues concerning the relationship between constructs and measures, both with regard to operationalizations of constructs and to what constructs are, actually, measured by certain measures.

The relationships in Figure 1 and Figure 2 (Section 2.14) that were suggested on the basis of theory and earlier research can be regarded as a first step toward more complete nomological networks of intermediate advertising effectiveness

measures, as the relationships between a fairly large number of relevant constructs are included. (Note, however, that the nomological network is incomplete in the sense that relationships between constructs and measures have been left out for the sake of simplicity.) The relationships in these two suggested nomological networks can be tested empirically with statistical analysis and hence provide the basis for two of the research questions that were formulated for the two empirical studies. The remaining research questions pertain to the relationship between constructs and measures, except for question which was formulated specifically for the second empirical study and concerns the degree of finish of ads used in advertising pretests.

The research questions for the first empirical study were the following:

1. *Are the alternative ways of operationalizing* attitude toward the ad *and* brand attitude, *i.e., with single-item or multiple-item measures, interchangeable?*
2. *Are the two suggested alternative measures of* acceptance, adjective checklists *and* cognitive response measurement, *interchangeable or are they complementary?*
3. *Are the intermediate effectiveness measures commonly used in advertising pretests related to each other as expected from theory and earlier research?*
4. *Are the intermediate effectiveness measures commonly used in advertising posttests related to each other as expected from theory and earlier research?*
5. *What is the ability of commonly used advertising pretest measures to predict the communication effects of advertising campaigns?*

In the second empirical study, research questions 1, 2, 3 and 5 from the first study were kept and two new research questions were added. The first of these, research question 6, concerned the number of items in the operationalization of ad credibility and is an extension of research question 1. The second question

concerned the degree of finish of print ads included in advertising pretests and draws on earlier research done with TV ads (Brown and Gatty, 1967; Schlinger and Green, 1980). The new research questions were formulated in the following way:

6. *Is a single-item operationalization of* ad credibility *interchangeable with a multiple-item measure of the same construct?*
7. *Are results from advertising pretests based on rough-sketch versions of print ads similar to the results from pretests with finished versions of the ads?*

3 Research Approach

This chapter describes the research approach of the two empirical studies. The chapter starts with an introduction to the general approach chosen in both empirical studies. The introduction is followed by a detailed discussion of the research approach in the first study, which covers the overall research design, the subjects and data collection procedure, the order of questions in the questionnaires, wording and scales of the questions in the questionnaires, the reliability of multiple-item measures, and other methodological issues. The description of the first study is followed by a detailed discussion of those parts of the second empirical study which differ from the first.

3.1 Introduction

Of the two empirical studies that were carried out, the first was the major empirical effort. The second empirical study was mainly done to supplement Study 1. Both studies were similar in design to laboratory experiments, since the aim was to control for as many relevant factors as possible, but the studies were not experiments in the real sense of the word. No treatment was varied between experimental groups, the treatment was the same for all subjects, but there was some variation in the measurement that was made between groups. The subjects in both studies were students.

An issue which is often discussed in connection with laboratory research is the *mundane realism* of the research, i.e., whether the experimental setting is similar to the world outside the laboratory (Swieringa and Weick, 1982). Lack of mundane realism is often seen as a threat to the external validity of the research, which, in turn, might limit the generalizability of the research (cf. Calder, Phillips and Tybout, 1981; 1982). Of particular interest in this context appears to

be the use of students as subjects. It has been argued, for example, that when students are used as subjects, age, education and variables related to these two, which are constant with student samples, are disregarded, while other variables receive too much attention (Wells, 1993).

The issue of realism and external validity in marketing research is discussed at length in two articles by Calder, Phillips and Tybout (1981; 1982). These authors maintain that there are two types of generalizability, *effects application* and *theory application*, which put different demands on the design of research. In effects application research, the aim is to obtain findings which can be directly generalized to a real-word situation of interest. In an advertising context, this could be to investigate the impact on brand attitude of various executional elements in ads, e.g., background color or number of words in the headline. In theory application research, the aim is to obtain theory which can be generalized, on the basis of theoretical expectations and empirical tests, to many real-world situations. An advertising example of theory application research could be how advertising influences brand attitude.

The two types of generalizability have methodological implications, which are discussed mainly by Calder et al. (1981), but also to some extent by Calder et al. (1982). In effects application research, the aim of directly generalizable results requires the research methodology to correspond very closely to the situation to which the findings are to be directly generalized. This means, for example, that respondents should be randomly sampled from the population of interest and that the research setting should be similar to the real-world situation. In the development of theory, the first step in theory application research, the aim should be to use methodology which supports rigorous tests of the theories, mainly by reducing or removing variance caused by variables extraneous to the theory which is tested. This means, for example, that respondents should be

homogenous with respect to variables not included in the theory and that no variation should be caused by extraneous variables in the research setting. In the second step, when theory is applied in real-world tests, which Calder et al. (1981) label *intervention*, the research methodology should allow for variation in variables which are extraneous to the theory, but which might influence the variables in the theory.

The focus in the present studies is on the relationship between different constructs and the relationship between constructs and measures. This focus means that the goal of the present research, in the terminology of Calder et al. (1981; 1982), is theory application and, given that many of the relationships between constructs have never been empirically tested, the theory-application research is in the theory development phase. Consequently, the research methodology of the studies should be directed towards rigorous tests of propositions derived from theory, i.e., extraneous variance should be reduced as much as possible.

The methodology chosen for the two empirical studies was similar to that used in laboratory experiments. This is a common approach in advertising research (e.g., Strong, 1912; Axelrod, 1963; Aaker and Brown, 1972; Patzer, 1991; Singh and Cole, 1993; Tavassoli, Shultz and Fitzsimons, 1995; Unnava, Agarwal and Haugtvedt, 1996; Nelson et al., 1997). The main advantage of this research approach is that the influence on the intermediate measures of many factors are controlled for, i.e., the extraneous variance is reduced. The ads in the study, for example, were for brands that are unknown to the subjects, and any communication effects will therefore be caused entirely by the ads in the studies, not by previous experience of the brands, word-of-mouth, advertising done by the brand, or other, similar factors. Further, the subjects were exposed to the ads

in a controlled context, which means that media factors and the number of exposures were controlled for.

The extraneous variance was also reduced by using students as subjects. Students are homogenous with respect to background such as age and education and, since individual variables were excluded as explanatory variables, this reduced the extraneous variance. It should also be added that using students as subjects in research is convenient, as they are generally available to researchers and they have the time needed to participate in studies. The use of students as subjects in advertising research is common (e.g., Murry and Dacin, 1996; Broach et al., 1997; Stephens and Russo, 1997). In the present context, one of the expected advantages of using students as subjects was that it would be easier to keep them as participants for the more than six weeks it took from start to finish in Study 1, than it would be for most other groups.

In one important respect, the methodology of the two studies was very realistic. The pretesting was carried out in much the same fashion as research companies do pretests, i.e., people were exposed to the ads under artificial circumstances and measurement made immediately following exposure. This means that any methodological shortcomings in pretests should be a problem also to real-world practitioners.

A number of steps were taken to increase the realism of the study in areas that were believed to require as much realism as possible. In the premeasure survey carried out before the pretest in both empirical studies, the relevance of the product categories in the studies was measured and only those subjects that satisfied the relevance criteria were included in the analyses. The ads that were included in both studies were real ads for real brands, but both ads and brands were unknown to the subjects. Using real ads for real brands reduces the risk

that the subjects would feel that the ads in the study were fake and not worth considering in the same way as real ads. This risk was further reduced by telling the subjects that the brands in the ads would be introduced on the Swedish market in the near future. Another step toward increased realism was taken in the first empirical study. Two exposures were inserted between the pretest and the posttest, so that the degree of realism would be higher than in studies that have measured some time after a pretest without any exposures in between (e.g., Shavitt and Brock, 1990; Lebenson and Blackston, 1997).

All parts of the empirical studies were handled by the author, from administering mail surveys to punching questionnaire responses into SPSS. The questionnaires in both studies were pretested in small convenience samples before they were administered in the actual studies. It was felt that convenience samples were sufficient since almost all the measures included in the questionnaires were well-established measures that have been used in advertising research for many years.

3.2 Study 1

The first study was carried out between late April and early June 1999. The subjects were first and second year students at the Stockholm School of Economics. In the study, four ads were measured in a pretest, exposed to the subjects in two mock questionnaires and measured in a posttest. In addition, the subjects' product category involvement and usage were measured in a pre-measure questionnaire administered before the ad test. Except for the ad test, which was administered in a classroom, all questionnaires in the study were administered by mail. An overview of the study is provided in Figure 3.

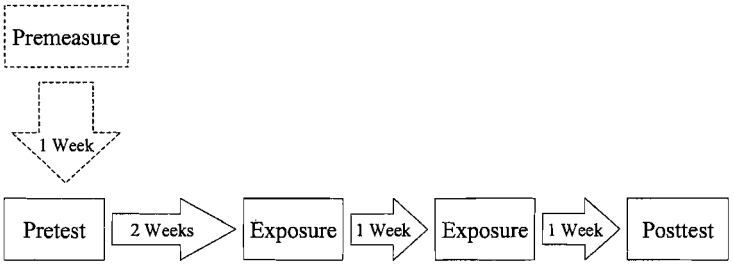

Figure 3. Overview of Study 1.

There were two weeks between the pretest and the first separate ad exposure, while the interval between the other parts of the study was about one week. The timing of the different parts of Study 1 is summarized in Table 9. The date in the table refers to when an action was taken or when a questionnaire was received by the subjects.

Part of Study	Timing
Recruitment	April 19-21
Pre-measure	April 20-22
Pretest	April 26-29
Exposure 1	May 11
Exposure 2	May 18
Posttest	May 26 - June 17
Letter of thanks	June 19

Table 9. Parts and timing of Study 1.

The following sections offer a detailed discussion of the design and implementation of Study 1.

3.2.1 Subjects

The subjects in the study were students who participated voluntarily, i.e., they did not receive any course credits for their participation. They were given brief information right before class and were asked to participate in "a study about marketing." All students attending were handed a brief questionnaire with questions about their name, age, gender and when it suited them to attend a

session to complete a survey. Those who wished to participate in the study handed back the questionnaire when the class they were attending was over. The interest among the students to participate in the study was so great that the recruitment took only three days, instead of the anticipated five days.

As incentives to participate, the subjects were offered a Red Cross lottery ticket (value SEK 25) and the chance to win cinema tickets or gift vouchers for the student book shop if they completed all parts of the study. The subjects who participated in the pretest sessions were offered a light lunch, as the pretest took place during the students' lunch break.

The subjects were randomly divided into three groups. The first group was a control group which did not take part in the pretest. This group was included to make it possible to evaluate to what extent the pretest influenced the posttest. The other two groups participated in the pretest, but answered partly different questionnaires. One group answered a pretest questionnaire with cognitive response measurement (CRM), which was replaced by an adjective checklist (ACL) in the other group. The total sample size for Study 1 was 155; 121 subjects completed all parts of the study resulting in an overall response rate of 77 percent. The sample sizes and the response rates for the three groups for each step in the study are displayed in Table 10.

	Sample Size	Pre-measure	Pretest	Exposure 1	Exposure 2	Posttest	Final Sample
Control	51	45 (88%)	-	37 (73%)	33 (65%)	33 (65%)	39
CRM	52	50 (96%)	45 (87%)	48 (92%)	47 (90%)	44 (85%)	42
ACL	54	53 (98%)	47 (87%)	48 (89%)	50 (93%)	44 (81%)	40
Total	157	148 (94%)		133 (85%)	130 (83%)	121 (77%)	

Table 10. Sample sizes and response rates in Study 1.

Seven subjects in the CRM group and seven in the ACL group did not attend the pretest session. These subjects were sent the exposure questionnaires and if they

returned these they received the posttest questionnaire. Of these, two from the CRM group and four from the ACL group returned all of the other questionnaires. These subjects were added to the control group, as they had answered all the questionnaires except the pretest. The final, adjusted, sample sizes are given in the far right column of Table 10.

The non-response in the study was mainly caused by unwillingness to participate, i.e., the non-response caused by wrong addresses or other technical difficulties was almost negligible. One pre-measure questionnaire and one exposure 1 questionnaire were returned due to wrong addresses. It should also be noted that the exposure 2 questionnaire was sent out before it was certain whether the exposure 1 questionnaire had been returned since the speed of the return mail was low. The posttest questionnaire was not sent until both the exposure 1 and exposure 2 questionnaires had been returned.

The average age of the subjects who participated in all steps of the study was just below 22 years. The youngest subject was 19 and the oldest was 31. Of the 121 subjects, 56 (46%) were women and 65 (54%) were men.

The cognitive response measurement and the adjective checklist, which were included in different pretest questionnaires, could have had different influences on how the subjects processed the ads and what communication effects were formed. In order to check for treatment effects, the mean scores on the processing and communication effects measures for the two pretest groups were compared. The scores for all four ads that were included in the study were compared.

There were between 26 and 42 valid observations for the four ads in the CRM group and between 29 and 40 valid observations in the ACL group. The

comparison showed that the difference between the two groups was statistically significant ($p < .10$) in only three instances, out of a total of 28. This result indicates that the two types of acceptance measures did not have different influences on the responses to the rest of the questions in the pretest questionnaire.

To check for treatment effects in the posttest, the results for the ACL group (n varied between 26 and 50), the CRM group (n varied between 26 and 48) and the control group (n varied between 14 and 43) were compared. The comparison showed that in only two instances, out of a total of 16, were the communication effect measures included in both the pretest and the posttest statistically significantly ($p < .10$) different. This indicates that the extra exposure and, presumably, the thorough processing of the ads in the pretest groups, did not influence the formation of communication effects to such an extent that they differed from those in the control group.

The CRM, ACL and control groups were also compared with regard to their scores on the four measures of brand and ad awareness in the posttest questionnaire. This comparison showed that the control group had consistently lower scores than the ACL and CRM groups on the four measures of brand and ad awareness. This difference was probably due to the extra exposure to the ads, i.e., a repetition effect, and to the active processing of the ads that the questions in the pretest questionnaire are likely to have caused.

3.2.2 Ads

Four ads were used in the study. They were real ads for real brands, taken from British and American magazines. The brands are not, and have not been, available on the Swedish market and should be unknown to the subjects. They

were translated from English into Swedish by the author and the Swedish text was inserted into the ads by an ad agency. They were printed on A4 size paper, and both the paper and the printing was of magazine quality.

To ensure that the ads in the study differed from each other as much as possible, they were selected to have creative execution tactics that correspond to different cells in the *Rossiter-Percy grid* (Rossiter and Percy, 1997). In addition at least one ads should be for a service. It was also important that the products in the ads were of relevance to the subjects. Table 11 shows the product categories in the ads and the location of the ad in the Rossiter-Percy grid.

	Informational	**Transformational**
Low involvement	Painkillers	Coffee for home use
High involvement	Pension Plan	Jeans

Table 11. The product categories in the ads in Study 1.

The place of the ads in the grid was verified by two expert judges, both Ph.D. candidates in marketing with teaching experience in the field. In addition, the subjects' product category involvement was measured in the pre-measure questionnaire, which also included some questions designed to measure the relevance of the product categories to the subjects (see Section 3.2.3).

Through every part of the study, the order of the ads in the questionnaires was rotated so as to minimize order effects. It has been suggested that if the same individuals evaluate more than one ad in the same session, only the evaluation of the first ad is valid (Rossiter and Percy, 1997). To test whether the evaluations of the ads differed depending on where in the questionnaire they were placed, one-way ANOVAs with rotation order as independent variable were run for each of the measures in both the pretests. There were between 11 and 22 valid observations for the four rotations in Study 1. There were seven measures,

except the ACL or CRM, for each ad, which resulted in a total of 28 one-way ANOVAs. There were eight models in which a statistically significant ($p < .10$) F-test indicated that differences occurred due to the order in which the ads were evaluated. Hence, 20 models did not show any differences in measurement values due to the rotation order of the ads.

The main source of the statistically significant differences due to order effects was the pension plan ad. Six out of the eight statistically significant differences came from the responses to this ad. The pattern in these significant differences was that Rotation 2, which evaluated the pension plan ad last of the four ads, had the highest score, while Rotation 4, for which the pension plan ad was the second ad, had the lowest score. These results do not completely rule out any effects on the results caused by the inclusion of several ads in the pretest, or the order in which they were evaluated, but neither do they indicate any serious problems.

3.2.3 Pre-measure

The pre-measure questionnaire was mailed to the subjects immediately following recruitment and it was completed and returned by the subjects before the pretest. The aim of the pre-measure survey was to ascertain the relevance of the product categories in the four ads to the subjects and to measure the product category involvement. To achieve this aim, the questionnaire measured four constructs for each product category. These four constructs are shown in Table 12.

Constructs
1. Usage of product category
2. Purchase of product category
3. Product category purchase intention
4. Product category involvement

Table 12. The constructs measured in the pre-measure questionnaire.

Usage and purchase of the product categories were measured using the following two questions:

> "Have you, on at least one occasion, had/used PRODUCT CATEGORY during the TIME FRAME?"
>
> "Have you, on at least one occasion, bought PRODUCT CATEGORY during the TIME FRAME?"

Responses to both questions were measured with a simple "Yes/No."

The purchase intention was measured with one question:

> "How likely is it that you will buy PRODUCT CATEGORY on at least one occasion in the TIME FRAME?"

Responses were measured on a five-point scale ("Definitely will not buy;" "Probably will not buy;" "Might/might not buy;" "Probably will buy;" "Definitely will buy") taken from Jamieson and Bass (1989).

To obtain comparison values and to avoid sensitizing the subjects to the product categories in the rest of the study, there were questions about eight product categories in the questionnaire. The product categories and the time frames in the usage, purchase, purchase intention and brand loyalty questions are shown in Table 13.

Product category	Usage	Purchase	Purchase Intention	Brand Loyalty
Coffee	1 month	2 months	1 month	1 year
Soft drinks	1 month	2 months	1 month	1 year
Jeans	1 year	1 year	1 year	3 years
Mutual funds	(have/have not)	-	10 years	-
Painkillers	1 year	2 years	1 year	3 years
Sugar-free chewing-gum	1 month	2 months	1 month	1 year
Pension plan	(have/have not)	-	10 years	-
Holiday travel	1 year	1 year	1 year	3 years

Table 13. Product categories and time frames in pre-measure relevance questions.

The questionnaire did not ask purchase questions for mutual funds or pension plans, and there were two versions of the purchase intention question for these two product categories: One for those who were already using the product category, and one for those who were not. The order of the product categories in the questionnaire was rotated to avoid order effects.

Table 14 shows the frequencies of positive responses on the three relevance questions for Study 1. For the purchase intention question, all the responses in the two highest scale steps (4 or 5) were coded as positive.

	Painkillers	Coffee	Pension Plans	Jeans
Category usage	124 (84%)	93 (63%)	8 (5%)	138 (93%)
Category purchase	92 (62%)	54 (36%)	-	73 (49%)
Category purchase intention	91 (62%)	42 (28%)	91 (62%)	80 (54%)
At least one of three	127 (86%)	96 (65%)	93 (63%)	141 (95%)

Table 14. Study 1: Product category usage, purchase and purchase intention.

The results in Table 14 show that the product categories were relevant to most of the subjects in Study 1. More than half of the subjects had used painkillers, coffee or jeans and almost two thirds intended to purchase a pension plan. Almost one-third of the subjects had purchased coffee, while about half of the subjects had purchased painkillers and jeans. Purchase intention was the highest for painkillers and pension plans (62 %) and the lowest for coffee (28 %). The high score on purchase intention was, probably, the result of the rather long time frame (ten years) in the question. The rather low purchase intention for coffee, on the other hand, probably was due to both the time frame (one month) and to the fact that the subjects drink coffee in other places than their homes.

Table 14 also shows the number of subjects who answered positively on at least one of the three relevance questions. A positive answer on at least one of the questions can be considered as the minimum for the product category to be regarded as being relevant to the subject. Consequently, all analyses in the thesis, which are presented in Chapter 4, are based only on the replies from those who answered positively to at least one of the three relevance questions.

Product category involvement was measured using the *Revised Product Involvement Inventory* (RPII) developed by McQuarrie and Munson (1991; see also Bearden, Netemeyer and Mobley, 1993). This is a ten-item scale and responses are measured using seven-point semantic differential scales. The instructions preceding the items were the following:

> "Below you will find ten pairs of adjectives. Indicate how well one or the other adjective in each pair describes how you perceive PRODUCT CATEGORY."

The ten items in the RPII scale are shown in Table 15.

Item
Important - Unimportant
Irrelevant - Relevant*
Means a lot to me - Means nothing to me
Unexciting - Exciting*
Dull - Neat*
Matters to me - Doesn't matter
Fun - Not fun
Appealing - Unappealing
Boring - Interesting*
Of no concern - Of concern to me*

*Indicates item is reverse scored.

Table 15. Items in the RPII (McQuarrie and Munson, 1991) involvement scale.

The minimum score on the RPII is 10 and the maximum score is 70. Descriptive statistics for the RPII scores for painkillers, coffee, pension plans and jeans for Study 1 are shown in Table 16. A one-way ANOVA was performed which

showed that the means were not equal over the product categories. The highest mean score was for pension plans and jeans (47 for both), with coffee only marginally lower (45) and painkillers markedly lower (34). The responses for all four product categories were well dispersed over the scale. There were responses from 10 to 70 over the four categories, but painkillers and pension plans did not get any of the highest scores and pension plans and jeans did not get any of the lowest scores.

	Painkillers (n=125)	**Coffee** (n=96)	**Pension plan** (n=93)	**Jeans** (n=140)
Mean	34	45	47	47
Std. Deviation	10.03	12.85	8.69	10.97
Minimum	10	10	19	19
Maximum	58	70	61	70

ANOVA:
Study 1: F=39.938, p < .000
Study 2: F=3.875, p < .011

Table 16. Study 1: RPII Involvement scores for the product categories in the ads.

For comparison, the mean RPII scores for soft drinks, mutual funds, sugarfree chewing gum and holiday travel are presented in Table 17. The mean score in both studies was similar, with holiday travel getting the highest scores and sugarfree chewing gum the lowest. The mean scores for holiday travel (65) was markedly higher than the mean scores for the other product categories. Also, the mean score for mutual funds (54) was markedly higher than the mean scores for soft drinks and chewing gum, which were on about the same level as coffee.

	Soft drinks (n=145)	**Mutual funds** (n=130)	**Sugarfree chewing gum** (n=128)	**Holiday travel** (n=137)
Mean	45	54	41	65
Std. Deviation	10.04	9.41	11.17	5.46
Minimum	10	22	10	38
Maximum	70	70	63	70

ANOVA:
Study 1: F=173.351, p < .000
Study 2: F=39.660, p < .000

Table 17. Study 1: RPII Involvement scores for the other product categories.

The relative size of the *RPII* scores in Table 16 does not clearly support that painkillers and coffee are low-involvement products and that pension plans and jeans are high-involvement products. It is clear that involvement with painkillers is lower than involvement with the other products, but the slight difference between the other three categories does not indicate any difference in involvement. A comparison between the product categories in the ads (Table 16) and the other categories (Table 17) shows that some products are clearly more involving than pension plans and jeans, but it also shows that involvement with painkillers is low relative to many products.

The involvement scores for the four product categories indicate that the difference between ads, in terms of involvement, was smaller than what was originally desired. Ideally, the product categories should have ranged in *RPII* scores from, say, the painkiller level (34) to the holiday travel level (65). The narrower range for the products in the ads did not have any direct implications for the rest of the study, except that the intermediate measures in the study were not tested on ads for products with very high involvement.

3.2.4 Pretest

The pretest was carried out during four separate sessions held in a classroom at the Stockholm School of Economics. Each subject had been assigned to a group and instructed to come to the classroom during the lunch break. Some 25 subjects attended each session. Upon arrival, the subjects were instructed to sit down, wait and not to look through the questionnaire in front of them.

When the test started, the subjects were told that they were going to see four ads for brands that were not yet available on the Swedish market, but that the brands

would be available in the near future. They were also told that there were no right or wrong answers to the questions that would follow each ad and that it was their opinions as consumers, not as students at a business school, that mattered. The importance of answering all the questions in the questionnaire was also emphasized.

The ads were inserted in the questionnaire and each one was followed by all the questions relating to it. The subjects could take as much time as they liked to look at each ad. (Pretests of the questionnaire indicated that two to three minutes was sufficient for most people.) After looking at the ad, the subjects could take the time they needed to answer the questions, except for the cognitive response measurement (discussed in detail below). An overview of the constructs measured in the pretest is offered in Table 18.

Measures
1. Cognitive response measurement *or* Adjective checklist
2. Ad credibility
3. Liking the ad
4. Attitude toward the ad
5. Brand purchase intention
6. Brand attitude - single item
7. Brand attitude - multiple items
8. Brand benefit beliefs

Table 18. The constructs measured in the pretest questionnaire.

The instructions for the cognitive response measurement (CRM) followed immediately after each ad. The instructions drew heavily upon Cacioppo and Petty (1981) and to some extent on Wright (1973) and Wansink et al. (1994). The following instructions were identical for all four ads in the study.

> "We now would like you to write down the thoughts and feelings you had while looking at the ad. We are interested in your thoughts and feelings about the ad in itself, as well

as the product in the ad. On the next page you will find a number of boxes. Write down your first thought/feeling in the first box, your second thought/feeling in the second box, and so on. Write only one thought/feeling in each box. It is not important whether your spelling or grammar is correct or not. There is no right or wrong, we just want you to write down your thoughts/feelings as honestly as possible.

You have got three minutes. Don't worry if you don't write in all boxes. There is more space to write in than is usually needed, in case some person would like very much space."

The subjects were asked to read the instructions, but to wait before they started writing down their thoughts or feelings. When all the subjects had read the instructions, they were told to turn the page and to start writing. On the page following the instructions were ten large boxes with one smaller box attached to it. On that page, the subjects were also reminded to write only one thought or feeling in each box and to not to worry about spelling and grammar. After three minutes, the subjects were instructed to stop writing.

The three minutes the subjects were given to write down the thoughts and feelings appeared to be more than enough for most of them. A large majority finished writing before the three minutes were up.

When all the subjects had answered all the questions in the entire pretest questionnaire, they received oral instructions to go through all their CRM responses and indicate whether the response was positive, negative or neutral.

This was done by writing a "+," "-" or "0" in the small box attached to the larger box in which the CRM response was written.

The input into the quantitative analyses from the cognitive response measurement was the *net cognitive response* (cf. Crites et al., 1994; Rossiter and Percy, 1997). In the data entry, the subjects' coding of the valence of responses was entered as 1, 0 or -1. To calculate the net cognitive response, these values were then summed for each individual. The cognitive responses were coded in this manner because of the results reported in Stephens and Russo (1997), who reported two results of interest in the present context. First, when the explanatory power of a complex coding scheme with 14 content categories was compared with the explanatory power of a coding scheme that only included the valence of the responses, R^2 did not differ in regressions with attitude toward the ad and brand attitude as dependent variables. Second, when Stephens and Russo compared the explanatory power of self-coded valence judgments of the responses and judge-coded valence judgments of the responses, the self-coded judgments outperformed the judge-coded judgments. In light of these results, it was not considered worthwhile to employ a more elaborate coding scheme and involve several judges.

The adjective checklist (ACL), the alternative to cognitive response measurement, also followed immediately after the ad in the questionnaire. The subjects filled out the ACL immediately after seeing the ad and could take as much or as little time as they desired. The instructions were printed on the same page as the ACL.

> "Think about the ad you have just seen and indicate whether the following words/phrases describe or don't describe the ad or your reactions to the ad."

There were 35 items in the ACL. Following Crites et al. (1994), response was measured using dichotomous checkboxes. The items in the ACL and the sources from which they were taken are shown in Table 19.

Items		
Attractiveness[1]	*Meaningfulness*[1]	*Dislike or Irritating/Silly*[2]
Beautiful	Meaningful	Irritating
Pleasant	Convincing	Silly
Gentle	Important to me	Pointless
Appealing	Strong	Phony
Attractive	Honest	
In good taste	Easy to remember	*Familiar*[3]
Exciting	Easy to understand	Seen-a-lot
Interesting	Worth remembering	Familiar
Worth looking at	Simple	
Comforting		*Dull*[3]
Colorful	*Vitality*[1]	Forgettable
Fascinating	New, Different	Dull
	Fresh	Worn-out
	Lively	
	Colorful	
	Sharp, bright, clear	

[1]From Wells (1964)
[2]From Aaker and Bruzzone (1981) and Aaker and Stayman (1990)
[3]From Aaker and Stayman(1990)

Table 19. Items in the Adjective Checklist in Study 1.

In the analyses of the data from the study, the ACL items were factor analyzed and the factor scores were used as independent variables in regression analyses. These analyses will be discussed in detail in Chapter 4.

The part of the questionnaire following the CRM or the ACL was identical in both versions of the pretest, both with regard to the constructs that were measured and how the questionnaire was designed.

The single-item measure of *attitude toward the ad* was written following Haley and Baldinger (1991) and Walker and Dubitsky (1994). The question was worded:

"Thinking about the ad for BRAND, which of the following statements best describes your feelings about the ad?"

Responses were measured on a five-point scale ("I liked it very much;" "I liked it;" "I neither liked it nor disliked it;" "I disliked it;" "I disliked it very much") taken from Haley and Baldinger (1991).

There were three items in the multiple-item measures of *attitude toward the ad* and *ad credibility*. The six items shared the same instruction:

"Below you will find six pairs of adjectives. Indicate how well one or the other adjective in each pair describes how you perceived the ad for BRAND."

The responses to the items were measured using seven-point semantic differential scales. The items for both attitude toward the ad and ad credibility were taken from MacKenzie and Lutz (1989). The items are displayed in Table 20.

Attitude toward the ad	Ad credibility
Good - Bad	Convincing - Unconvincing
Pleasant - Unpleasant	Believable - Unbelievable
Favorable - Unfavorable	Biased - Unbiased

Table 20. Items in the measures of attitude toward the ad and ad credibility in Study 1.

When the items in the multiple-item measure of attitude toward the ad were analyzed with principal components analysis, the results clearly showed that there was one factor in the data. The coefficient alpha for the scale was .92, which is very good according to commonly used recommendations for group data (cf. DeVellis, 1991; Nunnally and Bernstein, 1994).

The principal components analysis of the three ad credibility items indicated that there was a problem with the third item ("Biased - Unbiased"). The analysis

indicated that there could be two dimensions in the data, even if there was only one factor with an eigenvalue greater than one and the loading for the third item on the first factor was low (.33). The coefficient alpha was .64, which is barely acceptable. A frequency table for the third variable revealed that the responses were heavily skewed toward the upper end of the scale. More than 90 percent of the responses were on scale step 4 or higher. For the other ad credibility items, responses were dispersed over the entire scale. Based on these results, it was decided to drop the third item from the ad credibility measure. A principal components analysis of the remaining two items revealed one clear factor and the coefficient alpha increased to .87.

The measurement of brand purchase intention differed between the ads. Both the wording and the scale differed between the low involvement products and the high involvement products following Rossiter and Percy (1997). In all questions the word "try" was used, as all the brands were previously unknown to the subjects and a time frame was provided by making the intention conditional on category need by using "if" or "the next time" (cf. Rossiter and Percy, 1997). Two high involvement questions were worded as follows:

> "If you were going to take out a private pension plan, how likely would you be to try BRAND W?"
>
> "If you were going to buy a pair of jeans, how likely would you be to try BRAND X?"

Responses to these two questions were measured on Juster's (1966) 11-point scale (see Table 21).

Scale steps
Certain or practically certain (99 percent chance)
Almost sure (90 percent)
Very probable (80 percent)
Probable (70 percent)
Good possibility (60 percent)
Fairly good possibility (50 percent)
Fair possibility (40 percent)
Some possibility (30 percent)
Slight possibility (20 percent)
Very slight possibility (10 percent)
No chance or almost no chance (0 percent)

Table 21. The steps on Juster's (1966) 11-point scale.

The two low involvement questions was worded as follows:

> "The next time you buy painkillers, how likely would you be to try BRAND Y?"
>
> "The next time you buy coffee, how likely would you be to try BRAND Z?"

The responses were measured on the same five-point scale, taken from Jamieson and Bass (1989), as the product category purchase intentions in the pre-measure questionnaire ("Definitely will not try;" "Probably will not try;" "Might/might not try;" "Probably will try;" "Definitely will try").

The single-item measure of brand attitude followed the recommendations in Rossiter and Percy (1997) both with regard to the wording of the question and to the scale. Following Haley and Baldinger (1991), an opening line which referred to the ad was added to the question. This was done to resolve the difficulty of judging previously unknown brands which were identified in the pretesting of the questionnaire. The question included in the pretest was worded in the following way:

"Based on the ad you just saw, what would you say about BRAND?

Of all brands of PRODUCT CATEGORY, BRAND is the..."

Response was measured with a four-point scale ("the single best brand;" "one of several top brands;" "an average brand;" "below-average brand"). The single-item brand attitude measure for the pension plan brand differed somewhat from the format described here. The question asked about "company" instead of "brand" and the response scale was altered accordingly.

The multiple-item measurement of brand attitude was made using three items. The instructions preceding the items were worded as follows:

"Below you will find three pairs of adjectives. Indicate how well one or the other adjective in each pair describes your overall feeling of BRAND PRODUCT CATEGORY."

Responses were measured with three seven-point semantic differentials ("good - bad;" "pleasant - unpleasant;" "like - dislike"). The first two of these were found in MacKenzie and Lutz (1989) and Engel et al. (1995), while the third drew upon Gardner (1985) and Wansink and Ray (1992). In the study, the multiple-item measure of brand attitude worked well. Principal components analysis clearly showed one factor in the data and the coefficient alpha was .93.

The measurement of brand benefit beliefs was made using different items for the different ads. The items were preceded by the following instruction:

"Indicate the extent to which you think that BRAND PRODUCT CATEGORY..."

The responses were measured on a seven-point scale ranging from "to a very large extent" to "to a very small extent". There were between two and four items for each ad, items which were chosen on the basis of judgment and pretested before Study 1 was carried out. The items, the number of factors found in

principal components analysis and coefficient alpha for each measure are displayed in Table 22.

Items	
Pension plans ...are flexible ...are individually adjusted ...offer personalized advice ...offer skilled advisers (1 factor, coefficient alpha: .79)	*Jeans* ...are good looking ...are trendy ...are exclusive (1 factor, coefficient alpha: .80)
Painkillers ...give effective pain relief ...is effective against many different kinds of pain (1 factor, coefficient alpha: .86)	*Coffee* ...tastes good ...is exclusive ...has a good aroma ...is a coffee for discriminating people (1 factor, coefficient alpha: .88)

Table 22. Items in the brand benefit belief measures in Study 1.

The analyses of the measures showed that they worked well as scales; they were all unidimensional and the coefficient alphas were high, i.e., between .79 and .88.

3.2.5 Exposure 1 and 2

The first exposure questionnaire was sent to the subjects two weeks after the pretest. The second exposure questionnaire was sent one week after the first, i.e., three weeks after the pretest. The decision to expose the subjects to the ads twice was based on a belief that most advertising requires more than one exposure to work (cf. Krugman, 1972; Rossiter and Percy, 1997), but also on an assumption that the subjects' processing of the ads in the study would be more thorough than under normal circumstances. It was also felt necessary to limit the time period that the study would run, and any additional exposures would have required at least a week each.

The two exposure questionnaires were similar in their design. The four ads were inserted, without any explanation, between pages with questions. The inserted

ads were in all respects identical to those employed in the pretest. The questions in the two questionnaires were entirely unrelated to the four ads and the products in them. The questions in the first exposure questionnaire were about the subject's favorite product, which they were asked to specify themselves. The questions in the second exposure questionnaire were about different brands of candy bars.

3.2.6 Posttest

The posttest questionnaire was sent to the subjects when both their exposure questionnaires had been returned. The first posttest questionnaires were mailed about a week after the second exposure questionnaire had been mailed.

The three communication effects constructs that were measured in the pretest, i.e., brand benefit beliefs, brand attitude and brand purchase intention, were also measured in the posttest. The recommendation in Rossiter and Percy (1997) is that measurement of communication effects should be made in exactly the same way in the posttest as in the pretest, and this recommendation was followed, with one minor exception. The introduction to the single-item brand attitude question in the pretest ("Based on the ad you just saw...") was dropped in the posttest.

The three items in the multiple-item measure of brand attitude worked almost as well as a scale as they did in the pretest. The principal components analysis clearly identified one factor and the coefficient alpha was .87. The same was true for the different brand benefit beliefs measures. The principal component analyses clearly showed that there was only one factor in each of the measures and all the coefficient alphas were greater than .80.

The order of the sets of communication effects measures for the different brands was rotated to avoid order effects. The order of the measures in the questionnaire followed the recommendations in Rossiter and Percy (1997), except for the order of brand recall and category advertising cut-through. The measures and their order are shown in Table 23.

Measures
1. Brand recall
2. Category advertising cut-through
3. Brand recognition
4. Brand-prompted ad recall
5. Brand purchase intention[1]
6. Brand attitude - single item[1]
7. Brand attitude - multiple item[1]
8. Brand benefit beliefs[1]
9. Ad recognition

[1]Measure identical in pretest and posttest questionnaires

Table 23. Measures in the posttest questionnaire.

In the cover letter and in the instructions on the first page of the questionnaire, the subjects were carefully instructed to answer the questions in the order they appeared in the questionnaire and to never go back to add to or change any answers. They were reminded not to go back on all the pages that were sensitive to such changes.

The brand recall measure followed an example offered in Rossiter and Percy (1997), with one exception. Rossiter and Percy used the phrase "people like yourself" in the brand recall question. As this expression cannot be translated word for word into Swedish, the phrase "yourself or somebody like you" was used. The brand recall question was worded as follows:

> "When you think of PRODUCT CATEGORY intended for yourself or somebody like you, what brands come to mind?"

The question, which was asked for the four product categories in the ads, was followed by six empty lines on which the subjects could write the brand names they recalled.

Many different brands were named in response to the category-prompted brand recall question. In all, the subjects named 19 painkiller brands, 24 different coffee brands, 20 pension plan brands and 39 jeans brands. The large number of different brands indicates that the subjects took the questionnaire seriously and that they were willing to put some effort into responding to it, i.e., that the experimental realism was high (cf. Swieringa and Weick, 1982). However, the number of subjects who mentioned the four target brands were low. This outcome will be discussed in detail in the results chapter.

The category advertising cut-through (CPAR), which is a measure of ad recall, followed the recommendations in Rossiter and Percy (1997). The question was worded as follows:

> "Which printed ads for PRODUCT CATEGORY have you seen recently?
>
> a. If you can remember one or more ads, write a short description of each ad in the large boxes below. (How did it look? What was the message? etc.)
>
> b. If you can remember the brand that was advertised, write the name of the brand in the smaller box."

The question was followed by four large boxes, for the subjects' descriptions of the ads. Each of the boxes contained a smaller box that was meant for filling in the brand name. Category advertising cut-through was measured for the four product categories and each question appeared on a page of its own. The order of the product categories was rotated to avoid order effects.

The mean and the maximum number of recalled ads are shown in Table 24, together with the share of subjects who recalled the ads in the study. The coding of correct recall of the ads in the study was fairly lenient; all ad descriptions which indicated that the subject was describing the target ad, as well as references to "the ad in the study about marketing" or similar, were coded as recall of the target ad.

	Painkillers	Coffee	Pension Plan	Jeans
Mean no. of ads recalled	.58	1.01	.61	.81
Max. no. of ads recalled	2	3	2	4
Share recall of target ad	55%	70%	55%	59%

Table 24. Responses to the category advertising cut-through measure.

Brand recognition was measured for five brands in each of the four product categories in the ads. In addition to the brand name in the ad, the listed brand names included one fake brand that was made up for the study and three real brands available on the Swedish market. The brand recognition question was worded in the following manner:

> "Which of the following brands do you recognize?
> Indicate for each brand whether you recognize it or not."

The brands were listed under product category headings. The subjects indicated in checkboxes for each brand whether or not they recognized it. All the brand recognition questions were on the same page and neither the order of the product categories nor the brand names were rotated.

The measure of brand-prompted ad recall drew on the Millward Brown *Advertising Awareness Index* (see e.g., Brown, 1985; Hollis, 1995) and Rossiter and Percy (1997).

> "For which of the following brands have you seen printed
> ads (not TV or radio) recently?

Indicate for each brand whether you have seen it advertised or not."

Care was taken to design the brand-prompted ad recall measure differently from the brand recognition measure, to prevent the subjects from following the same response pattern for the two questions. The order of the product categories and the brands was different and brackets were used instead of checkboxes. For each brand the subjects indicated, by checking a "yes" or "no," whether they had seen the brand advertised in print recently. All the brand-prompted ad recall questions were on the same page in the questionnaire and neither the order of the product categories nor the brand names were rotated.

Ad recognition was measured using the original ads which were included as the last pages of the questionnaire. To prevent the subjects from accidentally seeing the ads and thereby being influenced by them, they were held together with a paper-clip. The ads were preceded by an instruction page asking the subjects to look at the ads and then to indicate on the last page whether they had seen the ad before or not. The question on the final page read:

"Have you seen any of these ads before?"

The question was followed by a listing of the four brands with checkboxes for answering "yes," "no" or "uncertain."

3.3 Study 2

The second study was carried out between late October and early November 1999. It was in many respects identical to the first parts of Study 1. The subjects were first year students at the Stockholm School of Economics, the subjects' product category involvement and usage were measured in a pre-measure questionnaire and an pretest was carried out. After the pretest, the data collection was finished. The timing of the different parts of Study 2 is summarized in

Table 25. The date in the table refers to when an action was taken or when a questionnaire was received by the subjects.

Part of Study	Timing
Recruitment	October 25-29
Pre-measure	October 26 - November1
Pretest	November 1-4

Table 25. Parts and timing of Study 2.

The research design of the two studies differed in three main ways. First, the degree of finish of the ads varied; half of the subjects was shown the same finished versions of the ads as in Study 1, while the other half was shown rough sketches of the ads. Second, there was only one version of the pretest questionnaire in Study 2 and it contained both cognitive response measurement and an adjective checklist. This meant that all subjects responded to both an ACL and CRM. The order of these two measures was varied among the subjects. Third, in Study 2 the pre-measure data collection was done via the Internet, instead of in a mail survey. The following sections discuss the design and implementation of Study 2, with an emphasis on the differences between the studies.

3.3.1 Subjects

The subjects were first year students at the Stockholm School of Economics, which means that they could not have participated in the first study. The recruitment procedure was similar to that used in Study 1. Students were invited to participate in a study about marketing. A letter was handed out and a brief presentation held just before class started. After the first invitation, all first year students were also sent an e-mail, reminding them about the study, which was then followed by another reminder, also by e-mail. Those who were interested in participating were asked to go to a web site and fill out a questionnaire. As

incentives to participate, the subjects were offered the chance to win cinema tickets or gift vouchers for the student bookshop if they completed both parts of the study and a light lunch in connection with the pretest. Unfortunately, the interest in participating in Study 2 was lower than for Study 1 and the final number of subjects was lower than expected.

The subjects were randomly divided into two groups based on the degree of finish of the ads. One group was exposed to the rough sketch versions of all four ads and one group was exposed to the finished versions. One of the rough sketch groups and one of the finished ad groups completed the CRM before the ACL, while the other two groups completed the ACL before the CRM.

A total of 71 students completed the pre-measure questionnaire. Of these, 54 participated in the pretest. The number of subjects for each type of questionnaire is shown in Table 26.

	Pretest
Rough sketch: CRM → ACL	10 (30%)
Rough sketch: ACL → CRM	11 (31%)
Finished ad: CRM → ACL	16 (19%)
Finished ad: ACL → CRM	17 (20%)
Total	54 (100%)

Table 26. Sample sizes in Study 2.

The demographics of the subjects in Study 2 were similar to those in Study 1. The average age was just below 22 and the age range was between 19 and 45. Out of the 54 subjects, 25 (46%) were women and 29 (54%) were men.

The low willingness to participate in Study 2, in comparison with Study 1, meant that the possibility of an in-depth exploration of research question 7, which concerns the use of rough sketch versions of ads in pretests, was limited.

The number of subjects who had judged the rough sketch versions of the ads and fulfilled the relevance criteria was in most cases too low to enable their responses to be analyzed separately.

3.3.2 Ads

The ad agency that had made the Swedish versions of the ads for Study 1 was hired to make rough sketches of the same ads for Study 2 . The ad agency was instructed to produce sketches that were similar to those used in their work with clients, with pictures that were hand drawn and copy that was legible. The printing of the finished ads was of exactly the same quality as in Study 1. The rough sketches were printed on A4 size paper with a color copying machine, of high, but not quite magazine, quality. The order of the ads in the questionnaires was rotated to minimize order effects. When one-way ANOVAs were used to test for order effects, out of a total of 36, only three, statistically significant ($p <$ .10) order effects were found.

3.3.3 Pre-measure

The constructs that were measured in the pre-measure were the same for both studies, as were the measures used. The only difference was how the data was collected: In Study 2, the data was collected via the Internet instead of in a mail survey.

The data was collected in co-operation with a marketing research firm specializing in data collection via the Internet. A web site was set up on which the pre-measure questionnaire was available. In the letter inviting the students to participate in the study, they were given the URL to the questionnaire and asked to complete it before the pretest. All subjects were given the same URL, since

rotation of the product categories in the questionnaire was part of the technological solution for the web site.

Table 27 shows the frequencies of positive responses on the relevance questions in Study 2. For the purchase intention question, all responses on the two highest scale steps (4 or 5) were coded as positive.

	Painkillers	Coffee	Pension Plan	Jeans
Category usage	40 (78%)	30 (61%)	11 (22%)	44 (85%)
Category purchase	26 (50%)	17 (34%)	-	22 (42%)
Category purchase intention	23 (44%)	17 (34%)	38 (70%)	17 (32%)
At least one of three	43 (80%)	34 (63%)	36 (67%)	47 (87%)

Table 27. Study 2: Product category usage, purchase and purchase intention.

The results in Table 27 show that the product categories were also relevant to most of the subjects in Study 2. The most notable difference between the studies was that the purchase intentions for both painkillers and jeans were almost 20 percentage points lower in Study 2 than Study 1. Also, for Study 2, all subsequent analyses were based only on the replies from those who answered positively to at least one of the three relevance questions.

Table 28 shows the descriptive statistics for the involvement (*RPII*) scores, which could vary between 10 and 70, for the products in the ads in Study 2. A one-way ANOVA showed that the means were not equal over the product categories. The highest mean score (47) was for pension plans, but the mean score for coffee was almost as high. Painkillers had the lowest mean score and the mean score for jeans was somewhat higher.

	Painkillers (n=35)	Coffee (n=27)	Pension plans (n=31)	Jeans (n=43)
Mean	39	46	47	42
Std. Deviation	9.23	12.18	9.63	11.60
Minimum	17	10	20	10
Maximum	57	63	61	61

ANOVA:
Study 1: F=39.938, p < .000
Study 2: F=3.875, p < .011

Table 28. Study 2: RPII Involvement scores for the product categories in the ads.

The descriptive statistics for the RPII scores for the other products included in the pre-measure are shown in Table 29. The mean score for holiday travel was the highest; the second highest score, but markedly lower, was for mutual funds; and soft drinks and sugarfree chewing gum had the lowest scores.

	Soft drinks (n=45)	Mutual funds (n=42)	Sugarfree chewing gum (n=42)	Holiday travel (n=44)
Mean	42	53	42	64
Std. Deviation	11.13	11.96	12.32	7.61
Minimum	20	18	12	41
Maximum	64	70	65	70

ANOVA:
Study 1: F=173.351, p < .000
Study 2: F=39.660, p < .000

Table 29. Study 2: RPII Involvement scores for the other product categories.

The *RPII* results for Study 2 were similar to those for Study 1, except in two respects. The mean score for jeans in Study 2 was lower (-5) than in Study 1 and the mean score for painkillers was higher (+5).

3.3.4 Pretest

The pretest was carried out in the same manner as in Study 1; the time, the place and all other practicalities were the same in both studies. Only minor changes were made in the instructions the subjects received before the pretest started; the subjects were told to complete the questionnaire in their own time and the

subjects that were shown rough sketches were told that the sketches would eventually become ads.

The major difference in the pretest in Study 2 was that all subjects answered both cognitive response measurement and an adjective checklist. The order of the two measures was rotated and some changes in the CRM and the ACL: The three-minute time limit on the CRM was removed and the number of items in the ACL was reduced. Further, some changes in the measurement of ad credibility were made. One item from Study 1 was dropped and a single-item measure was added.

The removal of the time limit on the CRM meant that no oral instructions were given during the pretest and that the subjects completed it entirely in their own time. The design of the questionnaire was unchanged. In the written instructions before each CRM, the only change was that "You have got three minutes" in the last paragraph was replaced with "Take the time you need." As the subjects could complete the questionnaire in their own time, the instructions for the coding of the CRM responses were enclosed in writing at the end of the questionnaire. To prevent the subjects from seeing them in advance, the instructions were placed on the final page of the questionnaire, which was folded and held together with paper-clips. The instructions were worded as follows:

> "We now want you to go back to the thoughts and feelings you wrote down for each ad earlier and judge whether they are positive, neutral or negative. Read each thought/feeling you have written and indicate in the small box whether it is positive (write a +), neutral (write a 0) or negative (write a -). You must not change your answers when you do this coding."

The instructions for the ACL were unchanged. The major change was that the number of items was reduced from 35 to 16. This was done by taking the two items with the highest loadings from each of the eight factors found in the factor analysis of the ACL in Study 1. These items are shown in Table 30.

Items	
Factor 1	*Factor 5*
Comforting	Familiar
Beautiful	Worn-out
Factor 2	*Factor 6*
Colorful	Convincing
Sharp, bright, clear	Meaningful
Factor 3	*Factor 7*
Silly	Easy to understand
Pointless	Simple
Factor 4	*Factor 8*
Exciting	Worth remembering
Interesting	Easy to remember

Table 30. Items in the Adjective Checklist in Study 2.

Two changes were made in the measurement of the *ad credibility* construct. First, the translation into Swedish of the third item, "Biased - Unbiased," which did not work in Study 1, was changed to an alternative translation. Second, a single-item measure of ad credibility was added. This question was worded in the same way as the ad liking question:

> "Thinking about the ad for BRAND, which of the following statements best describes your feelings about the ad?"

Responses were measured on a five-point scale ("I thought it was very credible;" "I thought it was credible;" "I thought it was neither credible nor not credible;" "I thought it wasn't credible;" "I thought it wasn't credible at all").

The new third item in the multiple-item measure of ad credibility worked better than the one tried in Study 1. The principal components analysis indicated that

the three-item scale was unidimensional and the coefficient alpha was .75. When coefficient alpha was calculated for the two first items, as a comparison with Study 1, it was found to be .86.

For the remaining measures in the pretest in Study 2, the principal components analyses clearly revealed only one factor for each of the multiple-item measures. All coefficient alphas were greater than .80 and, in some cases, greater than .90.

4 Results

In this chapter, the results of the analyses of the data collected in Study 1 and Study 2 will be presented and discussed. The purpose of the analyses was to elucidate the research questions presented at the end of Chapter 2 and, given the nature of the research questions, the analyses almost entirely are in the form of statistical analyses, most often regression analyses. Throughout the chapter the results will be related to the results in earlier studies, which were discussed in the literature review in Chapter 2. In the introduction to the chapter, an overview of the analyses is offered. This is followed by descriptive results from both studies, which in turn are followed by five sections with analyses directly related to the research questions of this thesis.

4.1 Introduction

The analyses presented and discussed in this chapter were divided into six parts (Section 4.2 to Section 4.7). With the exception of the first part of the analyses, the analyses were divided into parts according to which research question or research questions they addressed. The order of the analyses follows the order in which the research questions were presented at the end of Chapter 2, with the exception of the research questions that were only related to Study 2.

The first part of the analyses (Section 4.2) is not directly related to any of the research questions for Study 1, but one of the research questions for Study 2 will be addressed. In this part of the chapter, descriptive statistics for the pretests in Study 1 and Study 2 and the posttest in Study 1 are presented and discussed. This provides an overview of the data and leads on to the rest of the analyses. In connection with the descriptive results for the pretest in Study 2, research

question 7 (i.e., whether the degree of finish of ads influences advertising pretest results) will be addressed.

Following the descriptive statistics, the issue of the number of items in the operationalization of constructs will be addressed (Section 4.3). This relates to the first research question for both studies and research question 6, pertaining only to Study 2. The aim of these analyses was to investigate whether single-item and multiple-item measures of ad credibility, attitude toward the ad and brand attitude were interchangeable. Since both single-item and multiple-item measures of attitude toward the ad and brand attitude were used in the questionnaires in both studies, and ad credibility was operationalized in both ways in the questionnaire in Study 2, it was possible to make within-group comparisons of the two measure types.

The third part of the analyses (Section 4.4) deals with the second research question in both studies, i.e., the question of whether adjective checklists and cognitive response measurement are interchangeable or complementary. The analyses addressed whether adjective checklists overdiagnose ads (cf. Rossiter and Percy, 1997) and whether differences exist in the explanatory power of the two measure types. Differences in explanatory power were investigated using regression analyses. This made it necessary to perform principal components analysis on the items in the adjective checklists, so as to obtain factors which could be used as variables in the analyses. With the data from Study 1, the comparison was made between-groups, since two different pretest questionnaires were used in Study 1, while it was made within-group for Study 2, which had only one pretest questionnaire.

In the fourth part of the analyses (Section 4.5) the focus was on the relationships between the effectiveness measures included in the pretest in both studies. This

relates to the third research question and will be addressed with regression analyses. The analyses included all constructs measured in the pretests, except the ACL and CRM variables, and the investigation was carried out with a number of regression analyses. The analyses were performed separately for the pretests in Study 1 and Study 2.

The fourth research question, which concerned the relationships between effectiveness measures in advertising posttests, was dealt with in the fifth part of the analyses (Section 4.6). The analyses were similar to those carried out to investigate the relationships between variables in pretests, i.e., a number of regression analyses were made. The analyses in this part were based on data from Study 1, since Study 2 did not include a posttest.

The sixth, and final, part of the analyses (Section 4.7) is aimed at elucidating the fifth research question, i.e., how good commonly used advertising pretest measures are at predicting the communication effects of advertising campaigns. This issue was analyzed by using pretest effectiveness measures as independent variables in regression analyses with posttest effectiveness measures as dependent variables. This meant that only the subjects who participated in both the pretest and posttest in Study 1 could be included in the analyses.

Most of the analyses in this chapter were done separately for each of the four ads included in the two studies. The reason for this is that theory and earlier research lead us to expect that type of ad moderates the relationship between the variables included in the study. It has been suggested that some constructs, e.g., attitude toward the ad, are only important for certain types of ads (Rossiter and Percy, 1997). Theory and empirical results also indicate that the processing of information in ads varies with the level of involvement (e.g., Petty et al., 1983; Petty and Cacioppo, 1984). Further, earlier studies have found a better fit for

models with data for one brand than for models based on data pooled for several brands (Kalwani and Silk, 1982). In all, there appear to be grounds for treating type of ad as a moderating variable, since it is likely to influence whether a relationship exists between certain constructs, and can effect the strength of the relationships between constructs (cf. Baron and Kenny, 1986).

Although analyses were done separately for each ad, they were also performed on data pooled for all four ads. This was done as it is quite common to include data for many ads or brands in the same analysis (e.g., Gardner, 1985). The results of these analyses will in most cases not be discussed in the text, but are included to provide a reference for interested readers.

Throughout the analyses, a level of $p < 0.10$ has been used as a cut-off point for statistical significance. This rather lenient level was chosen since it was felt to be less serious to make *Type I errors*, i.e., rejecting a true null hypothesis, than to make *Type II errors*, i.e., accepting a false null hypothesis. Type I errors were felt to be less serious than Type II errors, particularly for the relationships tested with regression analyses, since a Type II error might lead to that a construct being excluded from further research, which probably has worse consequences than including a construct on the basis of a Type I error. For the convenience of readers who would prefer a more stringent cut-off point for statistical significance, all tables that present results provide information on whether the results were significant on the 10, 5 or 1 percent level.

When the regression analyses were done, some of the models were found to have fairly high R^2s, but few or no statistically significant coefficients, which is often an indication of multicollinearity. As a precautionary measure, all the regression models presented and discussed in this chapter were checked for multicollinearity. This was done following recommendations in Hair, Anderson,

Tatham and Black (1998). None of the models were found to have condition indices greater than 15 in combination with two or more variance proportions greater than .90. On the basis of this, it was concluded that multicollinearity did not constitute a problem in any of the models.

4.1.1 A note on notation

Throughout the results chapter, the measures used in the pretest and posttest will be referred to by variable names. These names, which are presented in Table 31, follow to a large extent the notation in earlier studies.

Measure	Variable name
Pretest only	
Number of Cognitive Responses	CRM_{Count}
Net Cognitive Response	CRM_{Net}
Adjective checklist factors	$ACL_{FactorX}$, $ACL_{FactorY}$, etc.
Ad credibility	$Cred_{Ad}$
Attitude toward the ad	A_{Ad}
Both pretest and posttest	
Brand Benefit Beliefs	$Beliefs_{Brand}$
Brand Attitude	A_{Brand}
Brand Purchase Intention	PI_{Brand}
Posttest only	
Category advertising cut-through	$Cut\text{-}through_{Ad}$
Category-prompted ad recall	$Recall_{Ad}$
Brand-prompted ad recall	$BPAR$
Ad recognition	$Recognition_{Ad}$
Brand recall	$Recall_{Brand}$
Brand recognition	$Recognition_{Brand}$

Table 31. Measures and variable names.

For the three constructs that had more than one operationalization, i.e., ad credibility, attitude toward the ad and brand attitude, the operationalization will be shown by a number at the end of the subscript that indicates the number of items used for the measure. For example, A_{Brand1} refers to the single-item measure of brand attitude and A_{Brand3} refers to the multiple-item measure of the same construct.

In the analyses with data from both the pretest and the posttest, a further addition to the subscript indicates in which of the two tests the measure was included. For example, $A_{Brand1(Pre)}$ refers to the single-item measure of brand attitude that was included in the pretest and $A_{Brand1(Post)}$ refers to the single-item measure of brand attitude included in the posttest.

4.2 Descriptive results

In the following sections, frequencies, shares, means, standard deviations and minimum and maximum values will be presented and discussed for the measures in the pretest in Study 1, the pretest in Study 2 and the Posttest in Study 1. The results will be presented in three separate sections, but differences and similarities in the three sections will be discussed.

4.2.1 Study 1: Pretest

In the first study, the two pretest questionnaires were identical in all respects, except for the cognitive response measurement (CRM) and adjective checklist (ACL). About half of the subjects who participated in the pretest listed their thoughts or feelings (responses) for each ad in CRM, while the other half filled out an ACL.

Descriptive statistics of two quantitative measures from the cognitive response measurement are presented in Table 32. The first measure is a simple count of the number of responses written for each of the ads (CRM_{Count}). The mean number of responses (around six) was similar for the four ads, and there was not much difference in the range of number of responses. The second measure is net cognitive response (CRM_{Net}), i.e., the sum of the subjects' coding of whether a response was positive (+1), neutral (0) or negative (-1) for each of the ads (cf. Rossiter and Percy, 1997). Given the number of responses, CRM_{Net} could vary

between -10 and + 10 for all ads, except the pension plan ad which could vary between -9 and +9. As can be seen in Table 32, CRM_{Net} did in fact vary between -9 and +8, with a mean score between -3.3 (for the painkiller ad) and +1.3. (for the coffee ad).

		All (n=136)	Painkillers (n=40)	Coffee (n=26)	Pension plan (n=28)	Jeans (n=42)
CRM_{Count}	Mean	5.9	5.8	6.5	5.8	5.8
	Std. Deviation	1.94	1.96	1.82	1.97	1.98
	Minimum	2	3	3	2	3
	Maximum	10	10	10	9	10
CRM_{Net}	Mean	–.4	–3.3	1.3	.8	.6
	Std. Deviation	3.92	2.67	4.10	3.81	3.44
	Minimum	–9	–9	–5	–6	–8
	Maximum	8	3	8	8	7

One-way ANOVA with *ad* as independent variable:
CRM_{Count}: F=.890, p < .448
CRM_{Net}: F=13.924, p < .000

Table 32. Study 1: Number of cognitive responses (CRM_{Count}) and net cognitive response (CRM_{Net}).

Descriptive statistics for the number of items checked in the adjective checklist are presented in Table 33. The ACL consisted of 35 items; the maximum number of items checked was 25, and the minimum was three. The mean number of items checked differed somewhat, with an average of about 12 for the painkiller ad and 16 for the jeans ad.

		All (n=145)	Painkillers (n=40)	Coffee (n=29)	Pension plan (n=31)	Jeans (n=45)
ACL_{Count}	Mean	14.2	12.3	14.9	13.2	16.1
	Std. Deviation	4.38	3.80	3.90	4.66	4.20
	Minimum	3	3	8	5	6
	Maximum	25	23	24	25	23

One-way ANOVA with *ad* as independent variable:
F=6.768, p < .000

Table 33. Study 1: Number of items checked in the adjective checklist (ACL_{Count}).

The share of subjects who checked an item are shown in Table 34. The main difference in what items were checked for each of the ads was between the

painkiller ad and the other three ads. The coffee, pension plan and jeans ads all had *pleasant* as the item with the highest share. For these ads, more than 80 percent of the subjects had checked *pleasant*, while less than three percent had checked *pleasant* for the painkiller ad. The most commonly checked item for the painkiller ad was *colorful* (which by mistake was included twice in the ACL), which was checked by 85 percent of the subjects; the share of subjects who checked *colorful* for the other ads was less than half of that. Another difference was that 65 percent of the subjects indicated that the painkiller ad was *irritating*, while the corresponding share for the other ads was below 15 percent. Another interesting result was that the pension plan ad was felt by many to be *honest* (71%) and *meaningful* (71%), while less than 30 percent checked these items on the other three ads. The mean scores for *fascinating*, *dull* and *exciting* were fairly similar for the four ads, with less than 20 percentage points difference between the ads with the highest and lowest shares. No item had either a very small share or a very large share on all ads, which indicates that the items were relevant for describing the ads, but the differences between the items also indicate that they discriminated between the ads.

	Painkillers (n=40)	**Coffee** (n=29)	**Pension plan** (n=31)	**Jeans** (n=45)
Beautiful	3%	76%	45%	73%
Comforting	0%	69%	55%	76%
Colorful	85%	21%	6%	36%
Fascinating	33%	24%	13%	16%
Meaningful	23%	14%	71%	18%
Convincing	10%	28%	55%	16%
Important to me	13%	14%	39%	16%
Strong	63%	28%	32%	11%
Honest	15%	24%	71%	27%
Easy to remember	73%	31%	32%	33%
Easy to understand	68%	79%	45%	60%
Pleasant	3%	90%	81%	93%
Worth remembering	28%	17%	39%	22%
Simple	48%	86%	42%	84%
New, Different	50%	21%	19%	4%
Fresh	13%	62%	39%	58%
Lively	63%	3%	19%	16%
Colorful	85%	17%	13%	38%
Sharp, bright, clear	45%	10%	13%	53%
Irritating	65%	10%	13%	7%
Silly	55%	24%	13%	33%
Pointless	35%	24%	6%	47%
Gentle	0%	69%	65%	91%
Phony	40%	21%	13%	22%
Seen-a-lot	20%	72%	55%	89%
Familiar	28%	55%	52%	80%
Forgettable	55%	38%	26%	42%
Dull	48%	34%	35%	36%
Worn-out	28%	45%	26%	69%
Appealing	10%	76%	58%	71%
Attractive	13%	69%	16%	67%
In good taste	3%	76%	58%	62%
Exciting	40%	41%	23%	36%
Interesting	38%	48%	65%	40%
Worth looking at	43%	76%	68%	71%

Table 34. Study 1: Percent of subjects who checked items on ACL.

Descriptive statistics for the remaining processing measures and the communication effects measures in the pretest in Study 1 are shown in Table 35. For all measures the mean, standard deviation, minimum and maximum values are shown. For each measure, except the two types of brand purchase intention measures, a one-way ANOVA was carried out. The F statistic from these

analyses were statistically significant for all measures, which means that the means were not equal over the four ads for any of the measures. The equality of the means of the two brand purchase intention measures was tested with *t*-tests as each measure was only measured for two of the ads. The difference in means between the two high-involvement ads and the difference between the two low-involvement ads were statistically significant.

		Painkillers (*n*=80)	**Coffee** (*n*=55)	**Pension plan** (*n*=59)	**Jeans** (*n*=85)	**ANOVA**
$Cred_{Ad2}$	Mean	4.9	7.3	9.2	6.8	*F*=24.155***
	Std. Deviation	2.98	3.21	2.78	2.89	
	Minimum	2	2	3	2	
	Maximum	14	14	14	14	
A_{Ad1}	Mean	2.1	3.5	3.4	3.3	*F*=42.147***
	Std. Deviation	0.85	0.92	0.95	0.89	
	Minimum	1	2	1	1	
	Maximum	4	5	5	5	
A_{Ad3}	Mean	7.4	14.3	14.3	14.2	*F*=59.889***
	Std. Deviation	3.45	4.4	3.88	3.92	
	Minimum	3	5	6	6	
	Maximum	18	21	21	21	
$Beliefs_{Brand}$	Mean	4.0	4.9	4.9	3.3	*F*=26.265***
	Std. Deviation	1.58	1.29	.90	1.23	
	Minimum	1	1	3	1	
	Maximum	7	7	7	6	
A_{Brand1}	Mean	1.5	2.5	2.3	1.9	*F*=30.314***
	Std. Deviation	.71	.74	.62	.59	
	Minimum	1	1	1	1	
	Maximum	4	4	4	3	
A_{Brand3}	Mean	7.9	14.4	14.1	12.0	*F*=46.403***
	Std. Deviation	4.09	4.05	3.41	3.24	
	Minimum	3	3	7	3	
	Maximum	21	21	21	21	
PI_{Brand} (LI)	Mean	2.0	2.6	-	-	*t*=-3.380** (indep. samples)
	Std. Deviation	1.02	.96	-	-	
	Minimum	1	1	-	-	
	Maximum	5	5	-	-	
PI_{Brand} (HI)	Mean	-	-	4.5	3.0	*t*=4.395*** (indep. samples)
	Std. Deviation	-	-	1.93	2.05	
	Minimum	-	-	1	1	
	Maximum	-	-	9	10	

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 35. Study 1: Descriptive statistics for measures in the pretest. (The n *given for each ad is the lowest over all measures in the table. In some cases, the* n *was 1 or 2 higher.)*

The first measure in Table 35 is *ad credibility* ($Cred_{Ad2}$). In Study 1, one of the three items that made up the measure of ad credibility did not work out as intended. As a consequence, ad credibility was measured with the remaining two

items only. This means that the minimum score was 2 and that the maximum score was 14. The descriptive statistics presented in Table 35 show that the range of $Cred_{Ad2}$ was between 2 or 3 and 14 for all four ads. The mean credibility scores varied between 4.9 for the painkiller ad and 9.2 for the pension plan ad, and the scores for the coffee and jeans ads were in between.

Responses on the single-item measure of attitude toward the ad (A_{Ad1}) were measured using a five-point scale. Table 35 shows that the mean scores on A_{Ad1} in Study 1 were around 3.5, with the exception of the painkiller ad (2.6). The range of responses varied somewhat between the ads. There were responses on the entire scale for the pension plan and jeans ads, while the coffee ad had responses between 3 and 5. The painkiller ad did not have any responses on scale step 5, but there were responses on the other four steps of the scale.

The multiple-item *ad attitude* (A_{Ad3}) measure scores could vary between 3 and 21, since responses were measured on three seven-point items. In Study 1, a marked difference could be seen between the painkiller ad and the other three ads; Table 35 shows that the mean score for the painkiller ad was about half of the mean score for the other ads. The range of scores was wide for all ads, with no responses on the three highest scale steps for the painkiller ad and no responses on the two or three lowest scale steps for the three other ads.

Brand benefit beliefs were measured using a different number of items for each of the four ads. To achieve comparability, the total score was calculated as an average of the items. As the individual questions had seven scale steps, the minimum is 1 and the maximum is 7. The results from Study 1 (in Table 35) show that the jeans ad had the lowest mean score with 3.3 and that the coffee and the pension plan ads had the highest with 4.9.

The single-item measure of *brand attitude* had four scale steps. The results from Study 1 (in Table 35) show that there were responses on all four steps for all ads except the jeans ad. The mean score for the painkiller ad was the lowest (1.5), and the mean score for the coffee ad the highest (2.5).

The mean scores on the multiple-item measure of *brand attitude* exhibited the same pattern as with single-item measurement, i.e., the painkiller ad had the lowest score and the coffee ad had the highest. There were three seven-point items and the scores ranged from 3 to 21 for three of the ads. For the pension plan ad the range went from 7 to 21.

The final construct measured in the pretest is *brand purchase intention*. This construct was measured with a single-item, but with different scales for the low-involvement and high-involvement ads. For the low-involvement ads, PI_{Brand} was measured with a five-point scale and for the high-involvement ads PI_{Brand} was measured with an eleven-point scale. The results for the two low-involvement ads, i.e., the painkiller and coffee ads, were similar in terms of the range of the scores, which went from 1 to 5 for both ads, but the mean score was statistically significantly lower for the painkiller ad. For the two high-involvement ads, the mean score was significantly higher for the pension plan ad than for the jeans ad. There were no responses on the highest scale step for either of the ads and there was no response on the scale step 10 for the pension plan ad.

The results in this section indicate that the subjects took an interest in their task during the pretest in Study 1. The average number of thoughts or feelings listed in the CRM was about six, which means that the average subject in the CRM group wrote down a total of 24 responses during the pretest session. The high average number of ACL items checked, more than 12 for all the ads, could also

be interpreted as an indication of interest, provided that they were chosen carefully and not just checked at random. Other indications of the subjects interest was the low internal non-response: the maximum was 2 observations for some of the measures for the jeans ad, and no internal non-response at all for the other ads. In all, these results indicate that the *experimental realism* of the pretest, i.e., whether subjects believed in, attended to and took the pretest seriously (Swieringa and Weick, 1982), was high.

Statistically significant differences were found between the four ads in the mean scores on all the different measures included in the pretest. Generally, the mean score for the painkiller ad was lower than the mean scores for the other ads, but some of the measures differed fairly much between one or two of the other ads. It is not unreasonable to assume that the "true" mean scores for at least some of the measures were different for different ads, given that they were chosen to represent different product categories and differed in their creative execution (place in the *Rossiter-Percy grid*). Assuming that the "true" mean scores were different for the four ads, the differences in the mean scores for the measures in the pretest provide an indication of construct validity (cf. Cronbach and Meehl, 1955; DeVellis, 1991).

4.2.2 Study 2: Pretest

The main differences in the pretest between the two studies was that both CRM and an ACL were included in the pretest questionnaire in Study 2, and that some of the subjects evaluated rough drafts of the ads, while the others judged the same finished ads as in Study 1. Descriptive statistics will be shown for both the rough and finished versions of the ads for all measures, except the CRM and the ACL. For the CRM and ACL, only descriptive statistics for each of the ads will be given. Barring these differences, the presentation will follow the format used

for the results from Study 1. The two studies will be compared at the end of the section.

Descriptive statistics for the CRM are shown in Table 36. The mean number of thoughts or feelings was around 4.5 for each of the ads. The net cognitive response (CRM_{Net}) was negative for the painkiller ad and slightly positive or positive for the other three ads. The possible range of CRM_{Net} was between -9 and +9, but the actual range was between -7 and +7.

		All (*n*=156)	**Painkiller** (*n*=41)	**Coffee** (*n*=34)	**Pension plan** (*n*=35)	**Jeans** (*n*=46)
CRM_{Count}	Mean	4.3	4.3	4.4	4.4	4.3
	Std. Deviation	1.63	1.65	1.54	1.77	1.64
	Minimum	2	2	2	2	2
	Maximum	9	9	8	9	9
CRM_{Net}	Mean	.1	–1.4	1.6	.1	.3
	Std. Deviation	3.13	2.78	3.19	2.96	2.99
	Minimum	–7	–7	–7	–6	–5
	Maximum	7	6	7	5	7

One-way ANOVA with *ad* as independent variable:
CRM_{Count}: F=.038, p < .990
CRM_{Net}: F=6.522, p < .000

Table 36. Study 2: Number of cognitive responses (CRM_{Count}) and net cognitive response (CRM_{Net}).

There were 35 items in the ACL in Study 1, which is quite a large number both from the perspective of respondent fatigue and analytical complexity. To reduce the number of items before Study 2 was carried out, the results from the principal components analysis (discussed in Section 4.4.2, below) of the ACL in Study 1 were used. In Study 2, only the two items with the highest loadings on each factor were kept, which almost halved the total number of items. The mean number of ACL items checked varied between 6.0 and 8.2 for the four ads, as shown in Table 37. The maximum number of items checked was 13, which is almost all of the 16 items included in the ACL, and the minimum was 0.

		All (n=159)	Painkillers (n=43)	Coffee (n=33)	Pension plan (n=36)	Jeans (n=47)
ACL_{Count}	Mean	7.1	6.0	7.4	6.7	8.2
	Std. Deviation	2.74	2.75	2.46	2.86	2.42
	Minimum	0	1	1	0	3
	Maximum	13	12	13	11	12

One-way ANOVA with ad as independent variable:
F=5.866, p < .001

Table 37. Study 2: Number of items checked in the adjective checklist (ACL_{Count}).

No items in the ACL were left unchecked, and no item was checked by all subjects. The share of subjects who checked the different items is shown in Table 38. There were five items for which the differences between the four ads were rather small, i.e., less than 20 percentage points, and where the share of responses was about 50 percent. These items were: *Easy to remember*, *easy to understand*, *worth remembering*, *silly* and *exciting*. The items *beautiful*, *comforting* and *familiar* had a much lower share of responses for the painkiller ad than for the other three ads. The item *colorful* had large shares on the painkiller and jeans ads (70% on both) and smaller on the coffee (48%) and pension plan (33%) ads. Some differences were also found for the remaining items, but no clear pattern in the responses emerged.

	Painkillers (*n*=43)	Coffee (*n*=33)	Pension plan (*n*=36)	Jeans (*n*=47)
Beautiful	5%	79%	19%	70%
Comforting	9%	67%	44%	57%
Colorful	70%	48%	33%	70%
Meaningful	21%	21%	64%	22%
Convincing	30%	42%	50%	23%
Easy to remember	53%	52%	53%	36%
Easy to understand	77%	64%	58%	68%
Worth remembering	49%	45%	50%	32%
Simple	58%	76%	64%	94%
Sharp, bright, clear	44%	27%	39%	70%
Silly	28%	12%	14%	26%
Pointless	33%	30%	22%	45%
Familiar	14%	38%	28%	60%
Worn-out	26%	27%	25%	63%
Exciting	47%	52%	39%	49%
Interesting	37%	61%	67%	40%

Table 38. Study 2: Percent of subjects who indicated items on ACL.

Descriptive statistics for the remaining processing measures and the communication effects measures included in Study 2 are shown in Table 39. For each measure, a two-way ANOVA with *ad* and *degree of finish of the ad* as independent variables was carried out. For five of the measures ($Cred_{Ad1}$, $Cred_{Ad3}$, A_{Ad1}, $Beliefs_{Brand}$ and A_{Brand1}) there was a statistically significant main effect of *ad*, i.e., the mean score on the measure was not equal for all the ads. One measure ($Cred_{Ad1}$) had a statistically significant main effect for the degree of finish of the ad, and three measures (A_{Ad3}, A_{Brand3} and PI_{Brand}, HI) had statistically significant interaction effects between *ad* and *degree of finish*, i.e., the mean scores were not equal for all combinations of ads and degree of finish over the measures.

There appears to be a pattern in the difference in the mean scores between rough and finished ads in the sense that the difference was consistent for all measures over the ads, although the direction of the difference differed between ads. For the painkillers and jeans ads, the mean scores for the finished ads were lower for all measures, except for one measure for the jeans ad. For the coffee and pension

plan ads the mean scores were almost consistently higher for the finished ads. The consistency over the ads and the difference between the ads meant that no measure differed consistently between finished and rough ad versions over all the ads.

		Painkillers		Coffee		Pension plan		Jeans		ANOVA
		Finished (n=27)	Rough (n=15)	Finished (n=23)	Rough (n=9)	Finished (n=18)	Rough (n=16)	Finished (n=27)	Rough (n=18)	
$Cred_{Ad1}$	Mean	2.5	2.7	2.8	3.3	3.4	3.4	2.6	2.8	Ad: F=11.912**
	Std. Deviation	1.11	.96	.88	.82	.96	.86	.83	.86	Finish: F=5.996*
	Minimum	1	1	1	2	2	2	1	1	Interact.: F=.434
	Maximum	5	4	4	4	5	5	4	4	
$Cred_{Ad2}$	Mean	6.8	7.4	8.0	7.9	8.9	8.0	6.4	7.1	Ad: F=4.539
	Std. Deviation	3.64	3.48	3.13	2.28	2.34	3.20	3.06	2.46	Finish: F=.018
	Minimum	2	2	2	4	5	3	2	2	Interact.: F=.569
	Maximum	13	14	13	11	13	14	13	12	
$Cred_{Ad3}$	Mean	10.4	11.1	11.6	12.9	12.9	11.9	9.0	10.1	Ad: F=7.871*
	Std. Deviation	4.57	3.68	4.09	2.38	2.94	4.53	4.19	3.24	Finish: F=1.092
	Minimum	3	6	3	10	7	6	3	4	Interact.: F=.582
	Maximum	18	17	19	15	18	21	18	16	
A_{Ad1}	Mean	2.5	2.8	3.6	3.6	3.6	3.2	3.3	3.6	Ad: F=6.711*
	Std. Deviation	1.07	0.94	0.93	0.84	1.16	0.90	0.84	0.70	Finish: F=.064
	Minimum	1	1	2	2	2	2	1	2	Interact.: F=1.058
	Maximum	5	4	5	5	5	5	5	5	
A_{Ad3}	Mean	8.8	10.7	14.9	13.7	14.7	12.3	12.8	14.4	Ad: F=3.719
	Std. Deviation	4.15	3.92	3.70	3.77	4.09	4.38	3.64	3.70	Finish: F=.001
	Minimum	3	5	8	5	8	6	7	8	Interact.:
	Maximum	16	18	21	17	21	18	21	19	F=2.715**
$Beliefs_{Brand}$	Mean	4.1	4.7	5.2	4.7	5.1	4.4	2.8	3.2	Ad: F=7.668*
	Std. Deviation	1.58	1.54	0.93	1.21	1.32	1.51	1.43	0.89	Finish: F=.035
	Minimum	1.0	2.0	3.5	2.0	1.0	1.5	1.0	1.7	Interact.: F=2.113
	Maximum	7.0	7.0	7.0	6.0	7.0	6.5	6.0	4.7	
A_{Brand1}	Mean	1.7	1.9	2.5	2.4	2.5	2.3	1.9	1.9	Ad: F=14.003**
	Std. Deviation	0.76	0.64	0.51	1.13	0.51	0.58	0.58	0.58	Finish: F=.217
	Minimum	1	1	2	1	2	1	1	1	Interact.: F=.718
	Maximum	3	3	3	4	3	3	3	3	
A_{Brand3}	Mean	8.9	10.4	15.6	14.9	15.8	12.1	10.5	12.6	Ad: F=3.315
	Std. Deviation	4.42	4.29	3.09	3.48	3.41	3.91	3.75	2.53	Finish: F=.022
	Minimum	3	4	10	10	8	6	3	7	Interact.:
	Maximum	18	18	21	20	21	19	17	18	F=4.609**
PI_{Brand} (LI)	Mean	2.3	2.3	2.8	2.4					Ad: F=2.163
	Std. Deviation	1.11	0.98	0.87	1.01					Finish: F=.419
	Minimum	1	1	1	1					Interact.: F=.872
	Maximum	4	4	4	4					
PI_{Brand} (HI)	Mean					5.8	4.4	3.0	3.8	Ad: F=2.485
	Std. Deviation					2.26	2.55	1.91	2.26	Finish: F=.086
	Minimum					2	1	1	1	Interact.:
	Maximum					10	11	7	8	F=4.861**

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 39. Study 2: Descriptive statistics for measures in the pretest. (The n *given for each ad is the lowest over all measures in the table. In some cases the* n *was 1 or 2 higher.)*

Table 39 presents descriptive statistics for three measures of ad credibility: One single-item and two multiple-item measures. The five-point single-item measure

was constructed for Study 2 and, hence, not included in Study 1. There was a statistically significant main effect for both *ad* and *degree of finish* in the two-way ANOVA of $Cred_{Ad1}$, i.e., the mean score was not equal over either ads or finished and rough versions of the ads. Only the finished version of the painkiller ad had answers on all five steps of the scale, while the other ads had answers on all but the highest and lowest step.

The two-item measure of ad credibility was the same as in Study 1, while the three-item measure was made up of the same two statements as the two-item measure and a new translation of the third item in Study 1 ("Biased - Unbiased"), that was made for Study 2. The behavior of the two measures was almost identical. The rank order of the mean scores for the four ads was identical for the finished ads, while the rank order for the rough sketch ads was different for the two highest scoring ads (coffee and pension plan). The two measures were similar also with regard to the range of values. Scores ranged from 2 to 14 on the two-item measure and from 3 to 21 on the three-item measure.

The range of answers to the *ad liking* measure (A_{Ad1}) was between 1 and 5 on the five-point scale for the finished versions of the painkiller and jeans ads; otherwise the scores were between 1 and 4 or 2 and 5. A statistically significant main effect was found for *ad* in the ANOVA, i.e., the mean scores were not equal for the four ads. The main difference was between the painkiller ad and the other three ads. The mean scores for the painkiller ad were 2.5 and 2.8, whereas the mean scores for the other ads ranged between 3.2 and 3.6.

No statistically significant main effect was observed for *ad* in the ANOVA of the multiple measure of *ad attitude* (A_{Ad3}), but a statistically significant interaction effect between *ad* and *degree of finish* was found. The tendency for

the mean scores was the same as for the ad liking measure, i.e., the mean score was lower for the painkiller ad than for the other ads. The range of answers was from 3 to 21 over all ads, but no ad had answers at both ends of the scale.

The mean score on the *brand benefit beliefs* measure was lower for the jeans ad than for the other ads, and the ANOVA showed a statistically significant main effect for *ad*. The answers were dispersed over the scale for all ads, but only the finished versions of the painkiller and pension plan ads had answers at both ends of the scale.

A statistically significant main effect for *ad* was shown by the ANOVA of the single-item *brand attitude* (A_{Brand1}) measure. The mean scores for the painkiller and jeans ads were lower than the mean scores for the coffee and pension plan ads. The rough sketch of the coffee ad was the only ad execution that had a response at the top end of the scale. The maximum for all other executions was 3, but all ads had answers on the lowest scale step.

There were no statistically significant main effects in the ANOVA for the multiple-item *brand attitude* (A_{Brand3}) measure, but the interaction effect was significant. None of the ad executions had answers at both ends of the scale, but over the ads, answers ranged from 3 to 21.

No statistically significant effects were found in the ANOVA run on the low-involvement (five-point) *brand purchase intention* measure. The mean scores were rather similar for both finished and rough versions of the painkiller and coffee ads. It is noteworthy that none of the subjects chose the scale step 5; the maximum was 4 for all four executions. For the high-involvement (eleven-point) brand purchase intention measure, there was a statistically significant interaction effect, but not any significant main effects. Responses ranged from 1 to 11 for

the rough version of the pension plan ad, but no responses were greater than 8 for the jeans ad.

One of the research questions for Study 2, research question 7, concerned whether results from advertising pretests with rough sketch versions of ads are similar to the results from pretests with finished versions of the ads. Ideally, the number of subjects in Study 2 should have been larger, to permit thorough statistical testing of whether degree of finish causes differences in pretest scores and, if so, when and how this happens. Given the actual number of subjects in the study, this was not possible, but the descriptive statistics in Table 39 provide an indication of whether differences existed in the pretest scores. The ANOVA in Table 39 showed only a statistically significant main effect for the finish of the ad for one variable, namely $Cred_{Ad2}$, but there were some statistically significant interaction effects between *ad* and *degree of finish*. The interaction effects were disordinal, i.e., the direction of the difference in mean scores varied between the ads (cf. Hair et al., 1998). For the painkiller and jeans ads, the mean scores were higher for the rough version of the ad than for the finished version, while the reverse was true for the coffee and pension plan ads. This indicates that pretests of rough and finished versions of ads may result in different pretest scores, but that the direction of the difference varies between ads. It is also of some interest that the differences between rough and finished versions of the ads were very small for two of the measures, namely A_{Ad1} and A_{Brand1}, which could indicate that degree of finish does not influence all types of measures in the same way. In all, the results from Study 2 indicate that the degree of finish influences the test scores in advertising pretests, but on the basis of the present results little can be said about when and how.

The subjects in Study 2 wrote down fewer thoughts or feelings in the CRM than the subjects in Study 1; the average was about four per ad compared to about six

per ad in Study 1. This difference can probably be explained by the fact that the subjects in Study 2 had to do both the CRM *and* fill out an ACL. An average of more than four responses is still quite impressive as it means that the average subject wrote down a total of more than 16 thoughts or feelings. The CRM responses and an internal non-response that never exceeded two subjects for any of the combinations of ads and executions indicate that the experimental realism (Swieringa and Weick, 1982) in Study 2 was on a satisfactory level.

Few differences in the mean scores on the measures in Study 2 were found when they were compared to the mean scores in Study 1. The scores for the painkiller ad were consistently lower in Study 1, but the difference was, in most cases marginal. All in all, it appears that the results in Study 2 were to a large extent comparable to the results in Study 1.

4.2.3 Study 1: Posttest

The posttest contained two groups of questions: First, a number of ad and brand awareness measures and, second, the same communication effects measures that were included in the pretest.

The frequencies and percentages for the ad and brand awareness measures are shown in Table 40. The values for the ad recognition measure show that almost all subjects recognized the ads when they saw them in the posttest questionnaire. The share of subjects ranged between 92 and 97 percent over the four ads. The share of subjects who recalled the ad when the product category was given as cue was lower. The recall was between 30 and 40 percent. The recall of the brand name in connection with ad recall was even lower, namely between 2 and 15 percent. The difference between the four ads was rather large for the brand-prompted ad recall measure (*BPAR*). The lowest share, 31 percent, was for the

painkiller ad and the highest, 61 percent, was for the coffee ad. The other two ads were in between: 40 percent for the pension plan ad and 48 percent for the jeans ad.

Few subjects recalled the brand when the product category was given as cue. The levels ranged between 2 and 11 percent. The levels of brand recognition were higher, but with some differences between the four ads. The lowest brand recognition was 32 percent (painkiller brand) and the highest was 72 percent (coffee brand). The pension plan brand was recognized by 39 percent and the jeans brand by 52 percent.

		Painkillers	Coffee	Pension plan	Jeans
$Recognition_{Ad}$	Frequency	98	68	72	111
	Percent	95%	92%	96%	97%
$Recall_{Ad}$, correct ad	Frequency	41	22	23	43
	Percent	40%	30%	31%	38%
$Recall_{Ad}$, correct brand	Frequency	2	11	2	9
	Percent	2%	15%	3%	8%
BPAR	Frequency	32	45	30	55
	Percent	31%	61%	40%	48%
$Recognition_{Brand}$	Frequency	33	53	29	59
	Percent	32%	72%	39%	52%
$Recall_{Brand}$	Frequency	2	8	4	4
	Percent	2%	11%	5%	4%

Table 40. Study1: Frequencies for posttest measures of ad and brand awareness.

Three communication effects were measured in the posttest. Descriptive statistics for *brand benefit beliefs*, *brand attitude* (single-item and multiple-item), and *brand purchase intention* are shown in Table 41. Overall, statistically significant differences were found in the means over the four ads on all measures, and the responses were rather well dispersed over the scales.

		Painkillers	Coffee	Pension plan	Jeans	ANOVA
$Beliefs_{Brand}$	Mean	3.7	4.6	4.4	3.0	F=39.479***
	Std. Deviation	1.15	1.07	.75	1.12	
	Minimum	1.0	1.8	2.3	1.0	
	Maximum	6.5	7.0	6.3	5.3	
	n	91	70	67	104	
A_{Brand1}	Mean	1.6	2.3	2.0	1.6	F=27.911***
	Std. Deviation	.63	.63	.57	.56	
	Minimum	1	1	1	1	
	Maximum	3	3	3	3	
	n	94	70	68	106	
A_{Brand3}	Mean	9.1	13.5	12.5	10.5	F=33.505***
	Std. Deviation	3.24	3.40	2.63	2.97	
	Minimum	3	4	7	3	
	Maximum	18	21	20	21	
	n	93	70	68	105	
PI_{Brand} (LI)	Mean	1.7	2.4			t=4.520*** (indep. samples)
	Std. Deviation	.92	.88			
	Minimum	1	1			
	Maximum	5	5			
	n	103	74			
PI_{Brand} (HI)	Mean			3.4	2.3	t=4.108*** (indep. samples)
	Std. Deviation			1.88	1.70	
	Minimum			1	1	
	Maximum			9	8	
	n			75	114	

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 41. Study 1: Descriptive statistics for posttest measures of communication effects.

The scores on the brand benefit beliefs measure ranged from 1.0 to 7.0, i.e., there were responses on both the low and high ends of the scale. The lowest mean score was 3.0 (jeans) and the highest mean score was 4.6 (coffee). The mean score for the pension plan brand was almost as high as for the coffee brand, while the mean score for the painkiller brand was in between.

None of the four brands had responses on the highest single-item brand attitude scale step ("the single best brand"), but there were responses on all other scale steps. The lowest mean scores were for the painkiller and jeans brands, and the highest was for the coffee brand. The pension plan brand was in between.

There were responses over the entire scale on the multiple-item brand attitude measure. Both the coffee brand and the jeans brands had responses up to the possible maximum of 21, and the painkiller and the jeans brands had the lowest

possible responses, i.e., 3. The lowest mean score was for the painkiller brand, and the highest for the coffee brand. The mean score for the pension plan brand was the second highest, and the mean score for the jeans brand the second lowest.

There were no responses on the highest scale steps on the high-involvement purchase intention measure, but there were scores at the lowest end of the scale. The mean score was higher for the pension plan brand, but both brands had mean scores at the low end of the scale. Interestingly, some subjects did choose the highest scale step ("definitely will try") for both the painkiller and coffee brands on the low-involvement purchase intention measure. The mean scores were on the lower half of the scale, with the score for the painkiller brand lower than the score for the coffee brand.

An interesting issue is whether any difference exist in mean scores between the pretest and the posttest in Study 1. The mean scores from the pretest and the posttest, taken from Table 35 and Table 41, are shown together in Table 42. Note that the figures are based on all the subjects who participated in the pretest and posttest, respectively. The tendency seen in Table 42 is that the mean scores in the posttest were lower than the mean scores in the pretest. The only exceptions were the two measures of brand attitude for the painkiller brand.

	Painkillers		**Coffee**		**Pension plan**		**Jeans**	
	Pretest (n=80)	Posttest (n=91-103)	Pretest (n=55)	Posttest (n=70-74)	Pretest (n=59)	Posttest (n=67-75)	Pretest (n=86-87)	Posttest (n=104-114)
$Beliefs_{Brand}$	4.0	3.7	4.9	4.6	4.9	4.4	3.3	3.0
A_{Brand1}	1.5	1.6	2.5	2.3	2.3	2.0	1.9	1.6
A_{Brand3}	7.9	9.1	14.4	13.5	14.1	12.5	12.0	10.5
PI_{Brand} (LI)	2.0	1.7	2.6	2.4	-	-	-	-
PI_{Brand} (HI)	-	-	-	-	4.5	3.4	3.0	2.3

Table 42. Study 1: Mean scores on communication effect measures in pretest and posttest (All subjects).

As a further comparison between the pretest and posttest, the mean scores on the communication effects measures for the subjects who participated in both tests are presented in Table 43. The tendency that can be observed in Table 43 is the same as in Table 42: the mean scores in the posttest were lower than the mean scores in the pretest, except, again, for the two brand attitude measures for the painkiller brand. The statistical significance of the differences in scores was tested with paired samples *t*-tests. The difference between the two tests was statistically significant for all measures for the two high-involvement brands, but not for the two low-involvement brands. The only statistically significant difference for the low-involvement brands was that the multiple-item measure of brand attitude was higher in the posttest than in the pretest.

	Painkiller (*n*=71)		**Coffee** (*n*=47)		**Pension plan** (*n*=51)		**Jeans** (*n*=77)	
	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
$Beliefs_{Brand}$	3.9	3.6	4.8	4.6	5.0***	4.5***	3.3***	2.9***
A_{Brand1}	1.5	1.6	2.4	2.4	2.3***	2.1***	1.9***	1.7***
A_{Brand3}	7.5***	8.7***	14.2	13.7	14.0***	12.7***	12.1***	10.5***
PI_{Brand}	2.0	1.7	2.5	2.5	4.5***	3.5***	3.1***	2.4***

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 43. Study 1: Mean scores on communication effect measures in both studies (Pretest subjects only).

The high positive response on the $Recognition_{Ad}$ measure shows that practically all subjects in the study have, at least minimally, processed the four ads. Many also recalled the ads, as shown by the results on the $Recall_{Ad}$ and *BPAR* measures. The number of subjects who responded positively on the brand questions was lower; brand recall was just above or below 10 percent for all four brands.

The low brand recall is not surprising as it would be quite an accomplishment if four previously unknown brands were recalled to a large extent, following only two or three ad exposures in the study, since several well-established brands

exist in each of the product categories. The rather low scores on $Recognition_{Brand}$ are somewhat harder to explain. One possible explanation could be that several subjects only indicated brands which they recognized as brands available on the Swedish market and, knowingly or unknowingly, chose not to indicate the brands in the study. Another factor which could have contributed to the low score was that brand recognition was measured with plain text, i.e., the brand name was not written with the typeface used in the ad and, no logotype was included, which would make the brand easier to recognize.

The fact that the number of people who indicated that they recognized the brand is lower than the number of people who responded to the questions about brand benefit beliefs, brand attitude and brand purchase intention is somewhat disturbing. It is logically impossible to hold any beliefs, attitudes or intentions about a brand which is unknown (cf. Rossiter and Percy, 1997). An unappealing possibility is that some of the subjects responded to the communication effect questions without actually knowing what brand they were evaluating. Some ground exists for not accepting that possibility as true. First, the ad recognition measure indicates that the ads were processed by most of the subjects. When filling out the posttest questionnaire, it would be natural for the subjects to assume that the questions about a painkiller brand, a coffee brand, a pension plan brand and a jeans brand were about the products they had seen in the ads earlier in the study. This would mean that the subjects who did not recognize the brand name in the $Recognition_{Brand}$ measure, either recognized the brand name when they saw it in connection with the questions or that they answered the questions with the "brand in the ad in the study" in mind. Second, some subjects did not respond to some or all of the communication effects measures for one or two of the brands, but they did respond to the same measures for the other brands. This indicates that they preferred not to answer than to try to have an opinion about brands that were unknown to them. It is not unreasonable to

assume that most subjects in the study behaved in the same way, which would mean that most of those who answered the communication effects questions had some sense of which brand they were evaluating.

The main difference between the ads in the pretest, i.e., that the mean scores for the painkiller ad were consistently lower, remained in the posttest, but only compared to the coffee and pension plan brand. The jeans brand had mean scores more or less on the same level as the painkiller brand. It is also of some interest that generally the mean scores were lower in the posttest than in the pretest. This indicates that evaluations made at some distance in time from exposure to an ad get to be lower than evaluations made immediately after exposure to an ad, but an alternative explanation could be wearout of the advertising effects (cf. Blair, 1987; Rossiter and Percy, 1997).

4.3 Comparison of single-item and multiple-item measures

The first research question for Study 1 was whether *multiple-item* measures of *attitude toward the ad* and *brand attitude* are interchangeable. To investigate this, both single-item and multiple-item operationalizations of A_{Ad} and A_{Brand} were included in the pretest and posttest questionnaires in Study 1. The same research question was posed in Study 2 and another question was added concerning the same issue, i.e., single-item and multiple-item measurement, for ad credibility. Consequently, the pretest questionnaire in Study 2 included both single-item and multiple-item operationalizations of ad credibility, attitude toward the ad and brand attitude. The aim of the statistical analyses of the two research questions is to investigate whether any substantial differences exist between single-item and multiple-item measures or whether these are actually interchangeable.

The first step in the analysis was to correlate the single-item and multiple items operationalizations for each of the constructs and to calculate the single-item reliabilities following a procedure from Wanous and Reichers (1996) and Wanous et al (1997). An additional analysis, performed only with the data from Study 1, was to compare the R^2 in regression models that differed only with regard to whether a single-item or multiple-item measure was used as one of the independent variables. The results from these initial analyses are shown in Table 44 (Study 1) and Table 45 (Study 2).

The correlations between the single-item and multiple-item measures were high and statistically significant ($p < .01$) in both studies, which provides a first indication that the two types of measures are interchangeable. The lowest correlation, .70, was for $A_{Brand(Post)}$ and the highest, .89, was for A_{Ad}.

The next test was to calculate a measure of the single-item reliability following a procedure used in Wanous and Reichers (1996) and Wanous et al (1997). The starting point is the formula for correction of attenuation (e.g., Nunnally and Bernstein, 1994: 257):

$$r'_{12} = \frac{r_{12}}{\sqrt{r_{11} r_{22}}} \qquad (4\text{-}1)$$

where r'_{12} = expected correlation between two perfectly reliable variables

r_{12} = correlation between variables 1 and 2

r_{11} = reliability of variable 1

r_{22} = reliability of variable 2

If r'_{12} is assumed to be 1.0, i.e., if the two variables in the equation were perfectly reliable, it follows that the correlation between the two measures is perfect, and Equation 4-1 can be rearranged into the following equation:

$$r_{12} = \sqrt{r_{11}} \cdot \sqrt{r_{22}} \qquad (4\text{-}2)$$

If the correlation between the two variables and the reliability of one of the variables are known, the remaining reliability can be estimated by solving Equation 4-2. This relationship was used to estimate the reliabilities of the single-item measures in Study 1 and Study 2.

It is reasonable to assume that the single-item and multiple-item measures in Study 1 and Study 2 are measures of the same underlying construct. Assuming this, the correlation between them, if the measures were perfectly reliable, would be 1.0. With the available data it is possible to calculate both the correlations between the two types of measures, as well as the reliability, i.e., Cronbach's alpha, for the multiple-item measures. With the correlations and the reliabilities for the multiple-item measures, Equation 4-2 can be applied and the single-item reliability can be solved. The correlations, multiple-item reliabilities and single-item reliabilities are shown Table 44 and Table 45.

	A_{Ad}	A_{Brand}	
		Pretest	Posttest
Correlation (Pearson)	.89***	.76***	.70***
Multiple-item reliability (alpha)	.92	.93	.87
Single-item reliability	.86	.62	.56
No. of pairs of models tested	50	20	5
Mean difference in R^2	.02	.02	.04
Max. difference in R^2	.09	.05	.10
Number of significant differences	0	0	0

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 44. Study 1: Comparison of single-item and multiple-item measures.

The single-item reliabilities ranged from .56, for $A_{Brand(Post)}$, to .86, for A_{Ad} in Study 1. For $Cred_{Ad}$ and A_{Ad}, the single-item reliabilities were only marginally lower than the multiple-item reliabilities. The differences ranged between .01 and .06. For A_{Brand} the differences were larger, between .31 and .32. The single-item reliabilities for A_{Ad} (> .80) and $Cred_{Ad}$ (> .70) are more than acceptable according to commonly used standards for group research (cf. DeVellis, 1991; Nunnally and Bernstein, 1994). On the other hand, the single-item reliabilities for A_{Brand} are somewhat low, but close to .65, the level that, according to DeVellis (1991), is minimally acceptable. It should also be kept in mind that the estimated single-item reliabilities are minimum reliabilities and that the real reliabilities might be higher, but not lower (Wanous and Reichers, 1996; Wanous et al., 1997). In all, the estimated reliabilities indicate that lack of reliability should not be a great concern for the single-item measures of the three constructs.

	$Cred_{Ad}$	A_{Ad}	A_{Brand}
Correlation (Pearson)	.75***	.82***	.74***
Multiple-item reliability (alpha)	.75	.85	.92
Single-item reliability	.74	.79	.60

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 45. Study 2: Comparison of single-item and multiple-item measures.

The next comparison of the two measure types was meant to reveal if any differences existed in their explanatory power in regression models. A total of 150 regression models were run, with either a single-item measure or a multiple-item measure of one of the three constructs as one of the independent variables. The dependent variables in the regression models were PI_{Brand}, $Beliefs_{Brand}$, or A_{Brand}, while the independent variables varied depending on which variable was dependent in the model. When A_{Brand} was dependent variable, for example, the independent variables in one of the pairs of models that were tested were CRM_{Net}, $Cred_{Ad}$, $Beliefs_{Brand}$ and either A_{Ad1} or A_{Ad3}. For regression models with

more than one construct with alternative measures, the type of measure was only varied for one construct in each pair of models. The R^2 of the models varied between .12 and .89, but the variation in R^2 was, as will be shown, mainly between the pairs of regressions, not within them.

The lower half of Table 44 summarizes the results from the comparisons of 75 pairs of regression models with data from Study 1. The difference in R^2 between the models went both ways, i.e., in some cases the model with the single-item measure had a higher R^2 than the model with the multiple-item measure, and vice versa. The mean difference in R^2 that is reported in Table 44 is based on the absolute value of the differences and it is reported separately for A_{Ad}, $A_{Brand(Pre)}$ and $A_{Brand(Post)}$. The table also reports the maximum difference in R^2. The mean differences in R^2 were low, between .02 and .04, and the maximum differences were not very large, particularly not for $A_{Brand(Pre)}$. Following earlier research (e.g., Gunnarsson, 1999) the multiple R's from the regressions were used to test the statistical significance of the difference in R^2s (cf. Howell, 1997). For each pair of models the multiple R's were Fisher's z' transformed and the difference between them tested for statistical significance (cf. Cohen and Cohen, 1975). The analyses showed that none of the differences was statistically significant. Taken together, the analyses of the R^2s do not indicate that single-item measures of A_{Ad} or A_{Brand} are better or worse than multiple-item measures of the same constructs in explaining the variation in related variables.

As a final step in the analysis of the measure types, it was decided to perform an analysis of variance to test whether the mean scores on the two types of measures were equal. This was tested with *repeated measures* MANOVA on the data from both studies. Repeated measures MANOVA controls for individual differences, which means that multiple observations of the same variable from the same subject can be analyzed (cf. Bray and Maxwell, 1985; Hair et al.,

1998). In the analyses, the observations for the different ads were assumed to be independent and type of ad was treated as a between-subjects factor. It was also assumed that the single-item and multiple-item measures were two instances of the same measure, i.e., the repeated measures in the study. To allow a comparison of the mean scores on the two types of measures, the variables included in the measures were standardized and an average index was calculated for the multiple-item measures.

For both studies a 2 (*type of measure*) by 4 (*ad*) repeated measures MANOVA was carried out, in which *type of measure* was treated as a within-subjects factor and *ad* was treated as between-subjects factor. The results are shown in Table 46 (Study 1) and Table 47 (Study 2).

Construct	Source	Sum of Squares[a]	df	Mean Square	*F*	Sig.
A_{Ad} (n=279)	Within-Subjects Effects					
	Type of measure	.007	1	.007	.069	.793
	Type of measure * Ad	.472	3	.157	1.557	.200
	Between-Subjects Effects					
	AD	183.095	3	61.032	54.321	.000
$A_{Brand(Pre)}$ (n=281)	Within-Subjects Effects					
	Type of measure	.012	1	.012	.054	.817
	Type of measure * Ad	3.258	3	1.086	5.061	.002
	Between-Subjects Effects					
	AD	149.177	3	49.726	43.527	.000
$A_{Brand(Post)}$ (n=335)	Within-Subjects Effects					
	Type of measure	.013	1	.013	.049	.826
	Type of measure * ad	2.444	3	.815	3.036	.029
	Between-Subjects Effects					
	AD	126.426	3	42.142	36.281	.000

[a]Type III Sum of Squares

Table 46 Study 1: Repeated measures MANOVA with single-item and multiple-item measures.

In both studies, a statistically significant main effect was observed for *ad*, i.e., the mean scores were not equal for the four ads in either study. In light of the descriptive statistics presented earlier, this result was expected. The analyses did

not find a statistically significant main effect for *type of measure*, which means that the standardized mean scores were equal over the single-item and multiple-item measures. In three of the six models that were tested, a statistically significant interaction effect was found between *type of measure* and *ad*. Since no main effect was found for *type of measure* and the results for the interaction between *type of measure* and *ad* were mixed (i.e., only some of the interaction effects were significant), the results from the repeated measures MANOVA did not support that a difference exists between single-item and multiple-item measures of ad credibility, attitude toward the ad and brand attitude.

Construct	Source	Sum of Squares[a]	df	Mean Square	*F*	Sig.
$Cred_{Ad}$ (*n*=158)	Within-Subjects Effects					
	Type of measure	.001	1	.001	.004	.949
	Type of measure * Ad	2.287	3	.762	3.515	.017
	Between-Subjects Effects					
	AD	22.407	3	7.469	5.587	.001
A_{Ad} (*n*=160)	Within-Subjects Effects					
	Type of measure	.000	1	.000	.000	.993
	Type of measure * Ad	.033	3	.011	.063	.979
	Between-Subjects Effects					
	AD	46.570	3	15.523	11.651	.000
$A_{Brand(Pre)}$ (*n*=155)	Within-Subjects Effects					
	Type of measure	.001	1	.001	.004	.950
	Type of measure * Ad	.958	3	.319	1.305	.275
	Between-Subjects Effects					
	AD	67.737	3	22.579	18.285	.000

[a]Type III Sum of Squares

Table 47. Study 2: Repeated measures MANOVA with single-item and multiple-item measures.

In the repeated measures MANOVAs it was assumed that the responses for the different ads were independent. The realism of this assumption might be challenged, for example, on the grounds that individuals might have a tendency to respond toward the high or low end of the scale. An alternative would have been to not treat the observations as independent in the analyses and test a 4 by 2 entirely within-subjects design and thereby control for individual effects over

the ads. A drawback with of design is that subjects with internal non-response on one or more of the questions for each construct are excluded from the analysis. As a total of 16 questions were asked for the two measures of each construct for the four ads, the risk of some non-response was rather high. In addition, the number of subjects would have been further reduced since only responses from subjects who met the relevance criteria, i.e., had bought, used or intended to buy the product category, *for all four product categories*, could be included in the analyses. This meant that the number of subjects would have been just above 30 for Study 1 and about 20 for Study 2, which was judged to be too few observations. In light of the low number of subjects and a lack of indications that the assumption of independence of the answers over ads is unrealistic, it was decided to rely on the models with *ad* as a between-subjects factor.

The results in this section indicate that the single-item measures of ad credibility, attitude toward the ad and brand attitude are interchangeable with the multiple-item measures of the same constructs. The two types of measures were strongly correlated, the estimated single-item reliabilities were acceptable or good, no statistically significant differences were found in R^2 in regression models with either single-item or multiple-item measures and the repeated measures MANOVA found no main effects for type of measure. Together, these results suggest that the single-item measures can be substituted for multiple-item measures for the three constructs tested in this study.

On the basis of the results in this section, it was decided to use the single-item measures of *attitude toward the ad* and *brand attitude* in the remaining analyses in the thesis. For *ad credibility*, the two-item measure was chosen in order to keep comparability in the results between the two studies.

4.4 Comparison of adjective checklists and cognitive response measurement

The second research question concerned the two suggested measures of acceptance, adjective checklists and cognitive response measurement. Are these two measures interchangeable or are they complementary? Both measures were included in both studies, but in Study 1, ACL and CRM were included in different questionnaires, while they were included in the same questionnaire in Study 2. This means that the data from Study 1 can only be used for group comparisons, while the data from Study 2 allows for individual level comparisons.

The analyses of the ACLs and CRM are divided into three parts. In the first part the focus is on whether ACLs overdiagnose ads, i.e., elicit responses that did not actually occur (cf. Rossiter and Percy, 1997). In the second part, the results from principal component analyses of the ACLs will be presented and discussed. These results are not directly related to the research question about ACLs and CRM, but the factors from the analyses were used as variables in the analyses in the third part.

In the third part of the analyses, the focus is on the explanatory power of variables based on ACLs and CRM. In these analyses, variables based on the ACLs and the CRM in the two studies were used as independent variables in regression analyses. This meant that two separate models for each dependent variable had to be run with the data from Study 1, since the ACL and the CRM in Study 1 were included in separate questionnaires. In Study 2, all subjects both filled out an ACL and responded to CRM. Consequently, only one model for each dependent variable had to be run, and it was possible to test whether the *ACL factors* or the net cognitive response made a statistically significant contribution to explaining the variance in the dependent variable in a model that

already had the other acceptance measure as an independent variable. Because of these differences, the analyses for Study 1 and Study 2 are presented in separate sections.

All the analyses done to compare ACLs and CRM will be summarized and commented on in a separate section following the analyses.

4.4.1 Do ACLs overdiagnose ads?

One of the reasons that Rossiter and Percy (1997) assert that cognitive response measurement is more valid than adjective checklists is that ACLs may overdiagnose ads, i.e., the ACL prompts responses to the ad that did not actually occur during the processing. The nature of processing makes it difficult to test this assertion, as measurement of processing responses generally takes place after the processing, and is therefore based on reconstructions of the processing and not on the actual processing. To measure processing responses during the processing is probably not a viable alternative, as measurement is likely to interfere with the processing.

In the present context one quantitative evaluation can be made of whether ACLs overdiagnose ads. This involves comparing the number of responses to the ACLs with the number of responses to the CRM in the same study. It is less likely that CRM would prompt responses to the ad which did not actually occur during the processing than would ACLs, since CRM does not provide response alternatives. This means that if many more responses are made to the ACL than the CRM, this could indicate that ACLs overdiagnose ads. For Study 1, the comparison is made between the two pretest groups and for Study 2, it is made on an individual level within the only pretest group. The mean number of CRM and ACL responses in both studies is shown in Table 48.

		All	Painkillers	Coffee	Pension plan	Jeans
Study 1	CRM_{Count}	5.9	5.8	6.5	5.8	5.8
	ACL_{Count}	14.2	12.3	14.9	13.2	16.1
Study 2	CRM_{Count}	4.3	4.3	4.4	4.4	4.3
	ACL_{Count}	7.1	6.0	7.4	6.7	8.2

Table 48. Mean number of cognitive responses (CRM_{Count}) and mean number of items checked in the adjective checklists (ACL_{Count}).

The results from both studies indicate that ACLs elicit a greater number of responses than CRM, indicating that ACLs possibly overdiagnose ads. Table 48 shows a clear difference in the number of responses to the CRM and the ACLs in the two studies. In Study 1, the mean number of CRM responses was less than half of the mean number of ACL responses for all four ads. The difference between the two types of measures was less marked in Study 2, but still fairly large. The mean number of CRM responses was between half and two thirds of the mean number of ACL responses. When the differences in the mean number of responses in Study 2 were tested using paired samples *t*-tests, the differences were statistically significant ($p < .01$) for all four ads.

4.4.2 Principal components analysis of the two ACLs

To obtain a more manageable number of variables for the remaining analyses a principal components analysis was conducted on the ACL items. Following earlier research on ACLs (e.g., Wells, 1964; Zinkhan and Fornell, 1985), this analysis was made with the data for all four ads in the same model. This meant that the principal components analysis could be carried out on 144 observations. Both the *KMO* measure (.827) and *Bartlett's test* ($p < .000$) indicated that the data was suitable for factor analysis (cf. Sharma, 1996; Hair et al., 1998).

A number of heuristic rules are available for determining the number of factors to extract in a factor or principal component analysis. In the analysis of the ACL

items from Study 1 two of these rules of thumb gave conflicting results. The *scree test criterion* suggested five factors, whereas the *latent root criterion* suggested eight factors. The choice was made to extract eight factors, because the factors in the rotated eight-factor solution were easier to interpret than the rotated factors in the five-factor solution. Ease of interpretation has been pointed out as an important criterion for the number of factors to extract (Stewart, 1981; Sharma, 1996). An additional reason was the higher total variance explained in the eight-factor solution; the total variance explained in the eight factor solution was 64 percent compared with 54 percent in the five-factor solution. The items in the ACL and their factor loadings from the final, EQUAMAX rotated, solution are presented in Table 49.

Factor	Loadings							
Appealing								
Beautiful	.74							
Pleasant	.66							
Gentle	.69							
Appealing	.66							
Attractive	.55			.56				
In good taste	.73							
Comforting	.75							
Fresh	.58							
Lively								
Colorful		.88						
Lively		.52						
Colorful		.88						
Sharp, bright, clear		.67						
Silly								
Irritating			.56					
Silly			.75					
Pointless			.68					
Phony			.64					
Forgettable			.59					–.46
Exciting								
Exciting				.76				
Interesting				.58				
Worth looking at				.54				
Fascinating				.41				
Dull				–.55				
Familiar								
Seen-a-lot					.67			
Familiar					.80			
Worn-out					.74			
Meaningful								
Meaningful						.68		
Convincing						.70		
Important to me						.66		
Strong						.44		
Honest						.64		
Simple								
Easy to understand							.83	
Simple							.73	
Memorable								
Easy to remember								.64
Worth remembering								.71
New, different								.41

n = 144
Total variance explained: 64%
KMO: .83
Bartlett: $p < .000$
Rotation: *Equamax*

Table 49. Study 1: Eight-factor solution obtained in an analysis of the Adjective Checklist (loadings below .40 are not shown in the table).

The main purpose of the principal component analysis was data reduction, but a comment on the interpretability and labeling of the factors is warranted. The first two factors are made up of items from the ACL in Wells (1964): In the first factor ($ACL_{Appealing}$) all of the items but one comes from the *attractiveness* dimension and in the second factor (ACL_{Lively}) all items come from the *vitality* dimension. The items in the third factor (ACL_{Silly}) come from the *irritation* or *dislike* dimensions found in most studies with ACLs (e.g., Wells et al., 1971; Aaker and Bruzzone, 1981; Aaker and Stayman, 1990; Biel and Bridgwater, 1990; Bruzzone and Tallyn, 1997). All items, except one, in the fourth factor ($ACL_{Exciting}$) come from the *attractiveness* dimension in Wells (1964). The fifth factor ($ACL_{Familiar}$) is similar to *familiar* in Aaker and Stayman (1990) and *familiarity* in Bruzzone and Tallyn (1997), and the sixth factor ($ACL_{Meaningful}$) is similar to *meaningfulness* in Wells (1964) and *personal relevance* in Wells et al. (1971). The two items in the seventh factor (ACL_{Simple}) come from *meaningfulness* in Wells (1964), while the eighth factor ($ACL_{Memorable}$) resembles *importance* in Bruzzone and Tallyn (1997). The similarity to factors obtained in earlier studies indicates that the factors obtained in the principal components analysis of the ACL in Study 1 are meaningful. The only large difference was that the *attractiveness* dimension in Wells (1964) was divided into two factors. This division does not seem unreasonable as it is possible to interpret both factors in a meaningful way. The items in $ACL_{Exciting}$ seem to indicate ability to arouse interest, while the items in $ACL_{Appealing}$ indicate the aesthetic qualities of ads.

Factor loadings from a principal components analysis of the ACL in Study 2 (reduced to almost half of the number of items compared to Study 1) are shown in Table 50, along with some basic statistical information. The number of factors was decided on in the same manner as for the ACL in Study 1, and the analysis

resulted in five factors. These were labeled using five of the labels from the principal components analysis of the ACL items in Study 1.

Factor	Loadings			
Meaningful				
Pointless	–.80			
Meaningful	.70			
Silly	–.66			
Convincing	.56			
Worn-out	–.52			
Memorable				
Worth remembering		.81		
Easy to remember		.78		
Interesting		.54		
Appealing				
Comforting			.80	
Beautiful			.79	
Simple				
Simple				.75
Easy to understand		.43		.65
Familiar				.64
Lively				
Colorful				.86
Sharp, bright, clear				.64
Exciting		.42		.47

n =156
Total variance explained: 63%
KMO: .76
Bartlett: $p < .000$
Rotation: *Equamax*

Table 50. Study 2: Five-factor solution obtained in an analysis of the Adjective Checklist (loadings below .40 are not shown in the table).

The principal component analysis of the Study 2 ACL did not result in factors directly comparable to those obtained in the analysis of the ACL in Study 1. Table 50 shows that the two items from $ACL_{Familiar}$ in Study 1, i.e., *Worn-out* and *Familiar*, wound up in two separate factors. The same happened with the two items from $ACL_{Exciting}$, i.e., *Interesting* and *Exciting*. The two items from ACL_{Silly}, i.e., *Pointless* and *Silly*, merged with the two items from $ACL_{Meaningful}$, i.e., *Meaningful* and *Convincing*, and with *Worn-out* from $ACL_{Familiar}$, into a new factor. The two items from $ACL_{Memorable}$ stayed together, but were joined by an

additional item, and the same happened to the two items from ACL_{Simple}. The two items from $ACL_{Appealing}$ stayed together in one factor, without being joined by any additional items. Despite the changes in dimensions between the two studies, the five factors were considered meaningful and were retained for further analysis.

The change in dimensions between the two studies means that the results of the principal components analysis of the Study 1 ACL are not validated. To check whether the number of items in themselves could have contributed to the change in dimensions between studies, a new analysis of the Study 1 ACL was carried out. In the new analysis, only the 16 items that were included in the Study 2 ACL were included in the analysis. The results from this principal components analysis indicated that the number of items could have contributed to the results obtained in the previous analyses. The results in the new analysis of the Study 1 ACL were similar to those obtained in the analysis of the Study 2 ACL. Six factors had an eigenvalue greater than one, but only three items did not belong to the same factors as they did in the Study 2 analysis. The items *worn-out* and *familiar* moved to a new factor and *exciting* moved from *lively* to *memorable.* These results indicate that the lower number of items in the second study could have contributed to the different factor structure.

4.4.3 Explanatory power of the ACL and CRM in Study 1

The analysis of the explanatory power of ACLs and CRM was done by means of regression analyses with the other intermediate pretest measures as dependent variables for both studies. This meant that two separate models for each dependent variable had to be run with the data from Study 1, since the ACL and the CRM in Study 1 were included in separate questionnaires. These pairs of

models will be discussed in this section, while the analyses of the data from Study 2 will be discussed in the next section.

In the analyses, the factors obtained in the principal components analysis of the adjective checklist and the quantitative measure based on the cognitive response measurement, net cognitive response, were included among the independent variables along with other intermediate pretest measures, in a number of regression analyses. The dependent variables in the models were ad credibility, attitude toward the ad, brand benefit beliefs and brand attitude. The discussion of the analyses will focus mainly on the explanatory ability of the *ACL factors* and CRM_{Net} and on whether one of the measures is better than the other at explaining the variance in the dependent variables. Both types of regression models will be shown and discussed for each dependent variable.

Ad credibility. There were no other independent variables than the *ACL factors* or CRM_{Net} in the regression models with $Cred_{Ad2}$ as dependent variable. The results in Table 51 (ACL models) and Table 52 (CRM models) show that the independent variables in both models managed rather well, with one exception, to explain the variance in the dependent variable. The R^2 ranged between .48 and .76 for the ACL models and between .14 and .66 for the CRM models. All models were statistically significant. A comparison of the adjusted R^2s show that two of the ACL models had a higher adjusted R^2 than the CRM models, and vice versa. Together, the results indicate that the two types of acceptance measures can contribute to an understanding of why an ad is perceived as credible, but the results do not indicate if one of the two types of measures is better than the other at explaining the dependent variable.

	All *Beta*	Painkillers *Beta*	Coffee *Beta*	Pension Plan *Beta*	Jeans *Beta*
$ACL_{Appealing}$	.30***	.29	.43**	.02	.37**
ACL_{Lively}	−.05	.11	.02	.08	−.01
ACL_{Silly}	−.45***	−.44**	−.52***	−.32**	−.41**
$ACL_{Exciting}$	.21***	.36**	.23	.15	.21
$ACL_{Familiar}$	−.01	.02	.12	−.10	−.02
$ACL_{Meaningful}$	.42***	.26*	.30	.50***	.26*
ACL_{Simple}	.21***	.47***	.30*	.26*	.04
$ACL_{Memorable}$	.20***	.14	.30*	.39**	.13
n	143	40	29	31	43
F	25.39***	5.51***	3.82**	8.54***	3.85***
R^2	.60	.59	.60	.76	.48
Adj. R^2	.58	.48	.45	.67	.35

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 51. Study 1: Pretest regression models with $Cred_{Ad2}$ *as dependent variable and* ACL factors *as independent variables.*

When the beta coefficients in Table 51 are scrutinized, it is clear that not all *ACL factors* contributed to explaining the dependent variables in all models. Two of the factors, ACL_{Lively} and $ACL_{Familiar}$, were not statistically significant in any of the models, while ACL_{Silly} was significant in all the models. The results for the other factors were mixed: $ACL_{Meaningful}$ and ACL_{Simple} were statistically significant in models for three of the ads, $ACL_{Memorable}$ and $ACL_{Appealing}$ were statistically significant for two of the ads, and $ACL_{Exciting}$ for one of the ads. It should be noted that an influence of $ACL_{Meaningful}$ on $Cred_{Ad2}$ was not entirely unexpected, since both variables included an item based on the word *convincing*.

In the CRM models (Table 52), the beta coefficient for CRM_{Net} was statistically significant for all four ads, but it was about half the size in the model for the painkiller ad, compared with the models for the other three ads.

	All *Beta*	Painkillers *Beta*	Coffee *Beta*	Pension plan *Beta*	Jeans *Beta*
CRM_{Net}	.68***	.38**	.81***	.69***	.62***
n	136	40	26	28	42
F	113.65***	6.23**	45.60***	23.58***	24.70***
R^2	.46	.14	.66	.48	.38
Adj. R^2	.45	.12	.64	.46	.37

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 52. Study 1: Pretest regression models with $Cred_{Ad2}$ *as dependent variable and* CRM_{Net} *as independent variable.*

Attitude toward the ad. In the models with A_{Ad1} as dependent variable, $Cred_{Ad2}$ was included as an independent variable in addition to *ACL factors* or CRM_{Net}. The ACL models are shown in Table 53 and the CRM models in Table 54. Both types of models, with one exception, explained more than half of the variation in the dependent variable and all the models were statistically significant. The R^2s in the models do not indicate whether the *ACL factors* or CRM_{Net} is better at explaining the variance in A_{Ad1}. A comparison of the adjusted R^2s reveals that the ACL model explained more variance for one of the ads, the CRM model more variance for another ad, and that for two of the ads the adjusted R^2 was about the same.

	All *Beta*	Painkillers *Beta*	Coffee *Beta*	Pension plan *Beta*	Jeans *Beta*
$ACL_{Appealing}$	.37***	.42*	.33*	.32*	.30*
ACL_{Lively}	–.10**	–.12	.06	–.02	–.00
ACL_{Silly}	–.26***	–.53**	–.15	–.19	–.25*
$ACL_{Exciting}$	.35***	.32*	.55***	.57***	.28**
$ACL_{Familiar}$	.12**	.14	.01	.07	.15
$ACL_{Meaningful}$	.16**	.21	–.00	.26	.23*
ACL_{Simple}	.21***	.20	.27*	.28	.27**
$ACL_{Memorable}$	.07	–.08	.26**	.19	.16
$Cred_{Ad2}$	.25***	.29	.30*	.03	.27*
n	143	40	29	31	43
F	34.72***	4.57***	8.43***	4.47***	6.57***
R^2	.70	.58	.80	.66	.64
Adj. R^2	.68	.45	.70	.51	.54

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 53. Study 1: Pretest regression models with A_{Ad1} *as dependent variable and* ACL factors *and* $Cred_{Ad2}$ *as independent variables.*

Two *ACL factors* in Table 53 were statistically significant in the regression models for all four ads: $ACL_{Appealing}$ and $ACL_{Exciting}$. Two of the factors, ACL_{Lively} and $ACL_{Familiar}$, were not statistically significant in any of the models for the four ads. The remaining factors were statistically significant in one or two models. The non-ACL variable $Cred_{Ad2}$ was significant in two models, for the coffee ad and the jeans ad. In the CRM models, shown in Table 54, CRM_{Net} was statistically significant in two of the four models for the different ads and $Cred_{Ad2}$ was statistically significant in three.

A comparison of the beta coefficients in the regression models with *ACL factors* and CRM_{Net} as independent variables indicates that ACL and CRM variables are not entirely interchangeable. At least two statistically significant *ACL factors* were found in each of the ACL models, while CRM_{Net} was not statistically significant in two CRM models. One reason for this difference seems to be that the *ACL factors* to some extent overlap with ad credibility. The beta coefficient for $Cred_{Ad2}$ is rather large and statistically significant for three of the ads in the CRM models, while the beta coefficient is smaller and statistically significant for only two of the ads in the ACL models. Somewhat surprisingly, the overlap between *ACL factors* and $Cred_{Ad2}$ does not appear to be caused mainly by $ACL_{Meaningful}$, which includes one of the items in $Cred_{Ad2}$ (*Convincing*), since $ACL_{Meaningful}$ was only significant in one of the four models.

	All *Beta*	Painkillers *Beta*	Coffee *Beta*	Pension plan *Beta*	Jeans *Beta*
CRM_{Net}	.53***	.48***	.27	.25	.49***
$Cred_{Ad2}$	.36***	.22	.61***	.54**	.43***
n	135	40	26	28	41
F	134.83***	10.35***	27.79***	14.97***	43.52***
R^2	.67	.36	.71	.54	.70
Adj. R^2	.67	.32	.68	.51	.68

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 54. Study 1: Pretest regression models with A_{Ad1} *as dependent variable and* CRM_{Net} *and* $Cred_{Ad2}$ *as independent variables.*

Brand benefit beliefs. One of the ACL models, the model for the pension plan ad, with $Beliefs_{Brand}$ as dependent variable, was not statistically significant and will not be discussed. All other models, both ACL models (Table 55) and CRM models (Table 56), were statistically significant. The R^2s in the models ranged from .13 to .66; two ACL models had a higher adjusted R^2 than the corresponding CRM models, and one CRM model had a higher adjusted R^2 than the corresponding ACL model.

	All Beta	Painkillers Beta	Coffee Beta	Pension plan Beta	Jeans Beta
$ACL_{Appealing}$	–.13	.22	–.13	.06	–.03
ACL_{Lively}	–09	.29*	–.04	–.12	.29
ACL_{Silly}	–.04	–.25	.13	.06	.03
$ACL_{Exciting}$	–08	–.11	.16	.19	.32*
$ACL_{Familiar}$	–.16**	.29*	–.09	.08	–.16
$ACL_{Meaningful}$	–.18*	–.14	.47**	.06	–.20
ACL_{Simple}	–.03	.26	–.30	.05	–.01
$ACL_{Memorable}$	–14*	–.05	.06	.06	.42**
$Cred_{Ad2}$	.18	–.04	.21	.31	.27
A_{Ad1}	.25*	.51***	.17	.02	–.23
n	143	40	29	31	43
F	5.81***	5.55***	2.21*	.61	2.223*
R^2	.31	.66	.55	.23	.41
Adj. R^2	.25	.54	.30	.00	.23

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 55. Study 1: Pretest regression models with $Beliefs_{Brand}$ *as dependent variable and* ACL factors, $Cred_{Ad2}$ *and* A_{Ad1} *as independent variables.*

Few statistically significant beta coefficients were found for the *ACL factors*. Table 55 shows that $ACL_{Appealing}$, ACL_{Silly} and ACL_{Simple} were not statistically significant in any of the models and that the other *ACL factors* were statistically significant only in one model each. Of the other two variables, A_{Ad1} and $Cred_{Ad2}$, the only statistically significant coefficient was for A_{Ad1} in the painkiller model.

None of the CRM_{Net} or A_{AD1} coefficients in the CRM models was statistically significant. The only statistically significant coefficients were for $Cred_{Ad2}$, which were statistically significant in the models for three of the ads.

In the same manner as when A_{Ad1} was dependent variable, a difference was found between the ACL and CRM models with regard to $Cred_{Ad2}$. Also in this case, $Cred_{Ad2}$ was only statistically significant in the CRM models.

	All *Beta*	**Painkillers** *Beta*	**Coffee** *Beta*	**Pension plan** *Beta*	**Jeans** *Beta*
CRM_{Net}	–.02	.17	–.42	.37	–.01
$Cred_{Ad2}$	.59***	.47***	.64*	.13	.43*
A_{Ad1}	–.06	.06	.42	.18	.02
n	135	40	26	28	41
F	17.31***	6.52***	7.33***	4.81**	3.00**
R^2	.28	.35	.50	.38	.20
Adj. R^2	.27	.30	.43	.30	.13

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 56. Study 1: Pretest regression models with $Beliefs_{Brand}$ *as dependent variable and* CRM_{Net}, $Cred_{Ad2}$ *and* A_{Ad1} *as independent variables.*

Brand attitude. In the models with A_{Brand1} as dependent variable the ACL model for the pension plan ad was not statistically significant and will therefore be disregarded in the discussion of the models. The other seven models were statistically significant. The explanatory power of the independent variables was fairly good or good. The R^2 of the ACL models ranged between .64 and .78. (Table 57) and the R^2 of the CRM models ranged between .34 and .75 (Table 58). Interestingly, the adjusted R^2s of the ACL models were consistently higher

than the adjusted R^2s of the CRM models, even if the difference was substantial only for the painkiller ad. This indicates that the explanatory power of *ACL factors* could be higher than the explanatory power of CRM_{Net} when brand attitude is the dependent variable.

	All *Beta*	**Painkillers** *Beta*	**Coffee** *Beta*	**Pension plan** *Beta*	**Jeans** *Beta*
$ACL_{Appealing}$	.20**	.52***	–.01	.18	.09
ACL_{Lively}	–.11*	–.19	–.09	–.15	.03
ACL_{Silly}	.03	–.28*	.04	.14	.10
$ACL_{Exciting}$	.11	.22*	–.07	.28	.09
$ACL_{Familiar}$	.12*	.56***	–.42**	.20	.12
$ACL_{Meaningful}$	.05	–.02	.18	.05	–.02
ACL_{Simple}	.03	.33**	–.23	–.39	.09
$ACL_{Memorable}$	.00	–.10	.04	–.21	.24*
$Cred_{Ad2}$	.11	.29*	–.12	.66*	–.08
A_{Ad1}	.14	.41**	–.12	.03	.19
$Beliefs_{Brand}$	.48***	–.08	.70***	.15	.58***
n	143	40	29	31	43
F	15.39***	9.26***	5.23***	1.42	4.94***
R^2	.56	.78	.77	.45	.64
Adj. R^2	.53	.70	.62	.13	.51

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 57. Study 1: Pretest regression models with A_{Brand1} *as dependent variable and* ACL factors, $Cred_{Ad2}$, A_{Ad1} *and* $Beliefs_{Brand}$ *as independent variables.*

There were few beta coefficients for the *ACL factors* which were statistically significant. Table 57 shows that $ACL_{Familiar}$ was statistically significant in two models, but with different signs before the coefficient. Two factors, ACL_{Lively} and $ACL_{Meaningful}$, were not statistically significant in any of the models. The remaining factors were statistically significant in one model each. In two of the models, $Cred_{Ad2}$ was significant and in the other two models $Beliefs_{Brand}$ was significant.

Table 58 shows that CRM_{Net} was statistically significant in only one model. This was the pension plan model, i.e., the ad for which the ACL model was not significant. The pension plan model was also the only one in which A_{Ad1} was

statistically significant. In the other models, $Cred_{Ad2}$ and $Beliefs_{Brand}$ contributed to explaining the variance, except for the coffee ad in which only $Beliefs_{Brand}$ was statistically significant. The models with brand attitude as dependent variable, also appears to indicate a difference between the ACL and CRM model regarding ad credibility. In the ACL models, $Cred_{Ad2}$ was only statistically significant if $Beliefs_{Brand}$ were not, while both variables were significant in two of the CRM models.

	All *Beta*	**Painkillers** *Beta*	**Coffee** *Beta*	**Pension plan** *Beta*	**Jeans** *Beta*
CRM_{Net}	.16	.05	.46	.39*	.03
$Cred_{Ad2}$	.25**	.33*	–.18	.24	.37*
A_{Ad1}	.17*	.00	.01	.35*	.02
$Beliefs_{Brand}$	.33***	.30*	.59**	–.14	.39**
n	135	40	26	28	41
F	39.56***	4.45**	6.56***	9.38***	7.45***
R^2	.55	.34	.75	.62	.45
Adj. R^2	.54	.26	.56	.55	.39

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 58. Study 1: Pretest regression models with A_{Brand1} *as dependent variable and* CRM_{Net}, $Cred_{Ad2}$, A_{Ad1} *and* $Beliefs_{Brand}$ *as independent variables.*

To summarize the results from the Study 1 regressions with *ACL factors* or CRM_{Net} as independent variables, it can be said that both types of variables contributed to explaining the variance in almost all the dependent variables. One exception to this was found: When $Beliefs_{Brand}$ was the dependent variable, CRM_{Net} was not statistically significant in any of the models. The R^2s of the models were in many cases high, up to .80, and more often above .50 than below.

The explanatory power was not consistently better for one of the two types of variables. Three of the ACL models had a higher adjusted R^2 than the corresponding CRM models when brand attitude was the dependent variable, but no clear pattern emerged for the other dependent variables. Each of the two

types of independent variables was the better one in almost an equal number of cases.

Some variation was found between the ACL and CRM models in terms of which other independent variables were statistically significant. Ad credibility tended not to be statistically significant in the ACL models, while it tended to be significant in the CRM models. Several differences emerged between the models for the different ads, both in terms of which independent variables were statistically significant and how much of the variance in the dependent variable was explained.

4.4.4 Explanatory power of the ACL and CRM in Study 2

The next round of regression analyses was done with the data from Study 2. In Study 2 all subjects responded both to an ACL and CRM, which means that the two alternative measures can be compared within-subjects, and not between-subjects, as was done with the Study 1 data. The focus in the analysis was on finding out whether the *ACL factors* contributed to explaining the variation in the dependent variables given that CRM_{Net} was already included in the model, and, conversely, whether CRM_{Net} contributed to the R^2 in models which already included *ACL factors* as independent variables. This was tested by running two types of pairs of regression models for the same dependent variables as in the analyses of the data from Study 1. In the first of the pairs of models, a model with only CRM_{Net} as one of the independent variables was compared with a model in which both CRM_{Net} and *ACL factors* were among the independent variables. In the second type, the first model had only *ACL factors* among the independent variables, whereas the second model had both *ACL factors* and CRM_{Net} among the independent variables. The aim in the pairwise comparisons of regression models was to ascertain whether R^2 increased when either *ACL*

factors or CRM_{Net} was added to the model. The statistical significance of the change in R^2 between the models was tested by means of a partial F test.

For each of the dependent variables, one set of regression models will be discussed. Whether the models include *ACL factors*, CRM_{Net} or both, will depend on the outcome of the two types of pairwise model comparisons. If either of the two acceptance measures does not contribute to explaining the variance in the dependent variable in any of the models for the four ads, they will be excluded from the final model.

Ad credibility. Both the five *ACL factors* and CRM_{Net} were included in the models with $Cred_{Ad2}$ as dependent variable. All four models were statistically significant and the R^2 ranged between .35 and .69. As can be seen in Table 59, CRM_{Net} was only statistically significant in one of the models, while at least one *ACL factor* was significant in each of the four models. It is interesting to note that three statistically significant *ACL factors* were found in the model for the painkiller ad, the same model in which CRM_{Net} was significant.

	All *Beta*	**Painkillers** *Beta*	**Coffee** *Beta*	**Pension plan** *Beta*	**Jeans** *Beta*
CRM_{Net}	.13*	.32**	–.13	.06	.00
$ACL_{Meaningful}$	.42***	.21*	.27	.49***	.40***
$ACL_{Memorable}$	.41***	.32**	.58***	.51***	.20
$ACL_{Appealing}$	.16**	.08	.17	.00	.40**
ACL_{Simple}	.12**	.26**	.20	.11	.08
ACL_{Lively}	.12**	.20*	.06	.13	.18
n	152	41	32	35	44
F	26.84***	12.76***	2.21*	10.29***	5.79***
R^2	.53	.69	.35	.69	.48
Adj. R^2	.51	.64	.19	.62	.40

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 59. Study 2: Pretest regression models with $Cred_{Ad2}$ *as dependent variable and* CRM_{Net} *and* ACL factors *as independent variables.*

Attitude toward the ad. All the models with A_{Ad1} as dependent variable, shown in Table 60, were statistically significant and the R^2s were high, between .57 and .81. CRM_{Net} was statistically significant in two of the models and at least one significant *ACL factor* was found in each of the four models. Three variables were not statistically significant in any of the models for the four ads: ACL_{Simple}, ACL_{Lively} and $Cred_{Ad2}$.

	All *Beta*	**Painkiller** *Beta*	**Coffee** *Beta*	**Pension plan** *Beta*	**Jeans** *Beta*
CRM_{Net}	.30***	.33**	.22	.10	.48***
$ACL_{Meaningful}$	.27***	.46***	.23	.19	.24*
$ACL_{Memorable}$	.29***	.30***	.56***	.40***	.15
$ACL_{Appealing}$	.23***	.10	.07	.27**	–.05
ACL_{Simple}	.09*	–.03	.08	–.02	.03
ACL_{Lively}	.08	.00	.19	.05	.13
$Cred_{Ad2}$	.15**	.14	.12	.27	.20
n	152	41	32	35	44
F	42.47***	20.25***	6.28***	15.20***	6.86***
R^2	.67	.81	.65	.80	.57
Adj. R^2	.66	.77	.54	.75	.49

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 60. Study 2: Pretest regression models with A_{Ad1} *as dependent variable and* CRM_{Net}, ACL factors *and* $Cred_{Ad2}$ *as independent variables.*

Brand benefit beliefs. The *ACL factors* did not add to the explanation of the variance $Beliefs_{Brand}$ and were therefore not included in the final models. All the models in Table 61 were statistically significant, but no significant variables were found in the pension plan model. The R^2s of the models ranged between .30 and .53 and the explained variance was mainly due to Ad_{Cred2}, which was the only statistically significant independent variable in all three models. CRM_{Net} was only statistically significant in one of the models, namely for the coffee ad, and A_{Ad1} was not significant in any model.

	All *Beta*	Painkillers *Beta*	Coffee *Beta*	Pension plan *Beta*	Jeans *Beta*
CRM_{Net}	.15*	.02	.30**	.19	.18
$Cred_{Ad2}$	.50***	.40*	.59***	.23	.37**
A_{Ad1}	−.02	.16	−.01	.23	.13
n	154	41	34	34	45
F	24.18***	5.18***	11.29***	5.29***	6.05***
R^2	.33	.30	.53	.35	.31
Adj. R^2	.31	.24	.48	.28	.26

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 61. Study 2: Pretest regression models with $Beliefs_{Brand}$ *as dependent variable and* CRM_{Net}, $Cred_{Ad2}$ *and* A_{Ad1} *as independent variables.*

Brand attitude. The pairwise evaluations of models with A_{Brand1} as dependent variable showed that neither the *ACL factors* nor CRM_{Net} added any explanatory power to the models if the other variable was already included. When the regression models were inspected, CRM_{Net} was found not to be statistically significant in any of the models that were evaluated. As a consequence, the final models presented in Table 62 include only the *ACL factors* among the independent variables. The model for the coffee ad was not statistically significant, while the other three were. The R^2s ranged between .38 and .69. $ACL_{Appealing}$ was the only *ACL factor* that was statistically significant in any model. Of the other variables, $Cred_{Ad2}$ and A_{Ad1} were not statistically significant in any of the models, while $Beliefs_{Brand}$ were significant in all models.

	All *Beta*	Painkiller *Beta*	Coffee *Beta*	Pension plan *Beta*	Jeans *Beta*
$ACL_{Meaningful}$	.02	.20	−.04	.06	−.13
$ACL_{Memorable}$	−.01	.09	−.01	.04	.05
$ACL_{Appealing}$	.18**	.03	−.09	.27*	.31*
ACL_{Simple}	.01	−.06	−.06	.11	.15
ACL_{Lively}	−.01	.05	−.11	.14	.24
$Cred_{Ad2}$	.11	.22	−.21	.22	−.03
A_{Ad1}	.20**	.24	.07	.01	−.03
$Beliefs_{Brand}$	.44***	.27*	.73**	.44**	.39**
n	151	43	30	35	43
F	16.86***	6.33***	1.55	7.15***	2.62**
R^2	.49	.60	.37	.69	.38
Adj. R^2	.46	.50	.13	.59	.24

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 62. Study 2: Pretest regression models with A_{Brand1} *as dependent variable and* CRM_{Net}, ACL factors, $Cred_{Ad2}$, A_{Ad1} *and* $Beliefs_{Brand}$ *as independent variables.*

To summarize, the results of the analyses with data from the pretest in Study 2 were similar to those of the analyses of the Study 1 pretest data. Whether *ACL factors* or CRM_{Net} were better at explaining the variance in the dependent variables could not be concluded from the results. Both types of variables contributed to explaining the variance in $Cred_{Ad2}$ and A_{Ad1}, the *ACL factors* did not contribute to explaining the variance in $Beliefs_{Brand}$, and CRM_{Net} did not contribute to explaining the variance in A_{Brand1}.

Rather than regarding the two acceptance measures as alternatives to each other, they could be regarded as complementary. If they are, both should be included in advertising pretests. This would be associated with increased costs for handling more extensive questionnaires, e.g., more time required for the coding answers, and the burden on respondents would be increased. To shed light on the possible gains from routinely including both types of measures in pretests, the average gain in R^2 by adding one of the measures was calculated for the sets of models with Ad_{Cred2}, A_{Ad1}, $Beliefs_{Brand}$ and A_{Brand1} as dependent variables. The results of these calculations are shown in Table 63.

Dependent variable	ACL factors added	CRM_{Net} added
Ad_{Cred2}	.28	.02
A_{Ad1}	.15	.05
$Brand_{Beliefs}$	.02	.07
A_{Brand1}	.01	.05

Table 63. Average gain in R^2 when ACL factors *or* CRM_{Net} *are added to regression models which already include the other type of variable.*

The average gains in R2, shown in Table 63, were fairly large when *ACL factors* were added to regression models with Ad_{Cred2} or A_{Ad1} as dependent variable and small when CRM_{Net} was added to models with the same dependent variables. The reverse was true for the models with $Brand_{Beliefs}$ and A_{Brand1} as dependent variables, but the difference between the two types of variables was very small and it is doubtful, therefore, whether a difference in explanatory power actually exists. Together, these results indicate, albeit weakly, that if the gain in R^2 is the sole criterion for deciding on whether to use both types of acceptance measures, it might be worthwhile adding an ACL to a questionnaire which already includes CRM, but not to add CRM to a questionnaire with an ACL.

To take the issue of whether both acceptance measures should be used one step further, the adjusted R^2s in the Study 1 models for $Cred_{Ad2}$ and A_{Ad1} were compared with the adjusted R^2s in the equivalent Study 2 models. It turned out that exactly half of the Study 2 models had higher adjusted R^2s than the equivalent Study 1 models which had either *ACL factors* or CRM_{Net} among the independent variables. This result, along with those discussed previously, indicate that there is no obvious gain in explanatory power to be made from using both types of acceptance measures.

4.4.5 Concluding comment on ACLs and CRM

The quantitative comparison of the ACL and CRM responses showed clearly that the adjective checklists in both studies elicited a larger number of responses than the cognitive response measurement. The mean number of CRM responses was between one third and half of the mean number of ACL responses. This could indicate that ACLs overdiagnose ads, i.e., that they elicit responses which did not occur (cf. Rossiter and Percy, 1997), but the quantitative analysis alone does not provide a solid foundation for any definite conclusions.

The principal components analysis of the ACL items in Study 1 resulted in eight interpretable factors that accounted for more than 60 percent of the variance in the items. These factors were similar to factors found in earlier research. The principal component analysis of the ACL items in Study 2 resulted in five interpretable factors that also explained more than 60 percent of the variance in the items. The factor structure found in Study 2 was in some respects different from the one found in Study 1, but a re-analysis of the items in Study 1 indicated that one of the reasons for the difference could be the difference in number of items between the two studies.

The regression analyses aimed at investigating the explanatory power of the *ACL factors* and net cognitive response yielded mixed results with regard to whether one variable type was better than the other at explaining the dependent variables. The explanatory power, in terms of adjusted R^2, was not consistently better for one or the other of two variable types.

The regression results indicate that the *ACL factors* to some extent overlap with other constructs measured in the pretest questionnaire. The most notable of these is ad credibility, which was statistically significant in many of the CRM models in the Study 1 analyses, but not in the corresponding ACL models. This

indicates that if a choice is made between using either ACL or CRM, care should be taken in the choice of other variables to include in the questionnaire and the analyses.

In the regression models from both studies, some of the *ACL factors* were only statistically significant in one or two models for some of the dependent variables. Given the significance ($p < .10$) level used as cut-off point in the analyses, some of these variables probably could be excluded from future ACLs without any statistically significant loss in explanatory power. Further, one *ACL factor* in Study 2, namely ACL_{Simple}, was not statistically significant in any model. This issue will not be further explored here, but if explanatory power was the main criterion, it would probably be possible to reduce the number of items in ACLs with little loss of explanatory power.

The results in the regression analyses were not entirely in line with the results in earlier research. In the studies by Zinkhan and Fornell (1985) and Zinkhan and Burton (1989) a number of *ACL factors* were statistically significant in regression models with brand attitude as dependent variable. In the present research the relationship between *ACL factors* and brand attitude was tenuous. While certain factors were statistically significant in some models, not one factor was consistently significant, i.e., in all or most models. The likely reason for this difference is that the regression models in the present study, unlike the models in the other two studies, included, apart from the *ACL factors*, several other variables, in particular brand beliefs. The results in the present study is in line with the results in Mittal (1990) and Fishbein and Middlestadt (1995) who found that the relationship between certain independent variables and attitudes are inflated unless cognitive beliefs are accounted for.

Both similarities and differences were found when the results for models including CRM_{Net} were compared to results in earlier research. One earlier study found a positive correlation between cognitive response variables and attitude toward the ad (liking the ad) and cognitive response variables and brand attitude (Lebenson and Blackston, 1997). Also, cognitive response variables were statistically significant in a regression model with brand attitude as dependent variable in another study (Wright, 1973). The results in the present research support a positive relationship between cognitive response variables and attitude toward the ad. The regression results for CRM_{Net} and attitude toward the ad were the same in both studies. CRM_{Net} was statistically significant in two of the four models and the coefficient in the other two models was positive and fairly large, but not statistically significant. The results in the models with brand attitude as dependent variable do not support a relationship between cognitive response variables and brand attitude. In the first study, CRM_{Net} was statistically significant in one of the four models, but it was not included in the models from the second study as it did not contribute to explaining the variance in A_{Brand} when the other variables were included.

The discussion of adjective checklists and cognitive response measurement has so far focused on explanatory power and the statistical significance of regression coefficients. An issue which has not yet been discussed is the diagnostic power of the two measures. Do these measures provide qualitative information that helps understanding why the creative execution of the ads works or does not work? This issue is not central to this thesis, but an illustration of the differences in the diagnostic information from the two types of measures is provided in Table 64. The table shows the 15 most frequently checked ACL items and some of the CRM responses to the coffee ad in Study 1.

Coffee ad		
15 most checked ACL items		Thoughts/feelings from CRM
Pleasant	90%	Package "...looks like a shampoo bottle" "...looks like a bottle with after shave" "...looks like a package of shower soap or a bottle of perfume" "...looks like a bottle of perfume" "Strange hat on the package - looks cheap" "Strange package"
Simple	86%	
Easy to understand	79%	
Beautiful	76%	
Appealing	76%	
In good taste	76%	
Seen-a-lot	72%	
Comforting	69%	Wrong category/brand "Sensuality - perfume advertising?" "Makes me think of ice cream, Carte D'ôr (?)" "I think of some other brand" "Feels like perfume advertising" "Reminds me of the ads for Emporio Armanis new perfume..." "Nice advertising which reminds me of many ads for coffee" "For some reason gives me sense of chocolate" "Dark chocolate"
Gentle	69%	
Attractive	69%	
Fresh	62%	
Familiar	55%	
Worn-out	45%	
Exciting	41%	
Forgettable	38%	

Table 64. Example of diagnostic information from adjective checklists and cognitive response measurement (for the coffee ad in Study 1).

The most popular ACL items were mainly positive: The ad was felt to be *pleasant*, *simple*, *easy to understand,* and so on. On the negative, side it was felt to be *seen-a-lot*, *worn out* and *forgettable*. The CRM responses presented in Table 64 are all comments about either the package shown in the picture in the ad or references to product categories other than coffee or to other brands than the brand in the ad. The package comments indicate that the package in the ad, either due to its design or how it is portrayed in the ad, can be mistaken for a bottle of shampoo or perfume. The references to other product categories or brands indicate that the ad might be mixed up with other ads for other product categories or brands of coffee.

The strength of the diagnostic information provided by the ACL is that it is quantified. It can be seen, for example, that 69 percent of the subjects in the advertising pretest felt that the ad was *attractive*. A limitation is that the answers are a result of what is asked. The subjects can only indicate items which appear on the list; they cannot convey other responses they might have had during the

processing of the ad. The CRM, on the other hand, allows the subjects to convey everything they felt during the processing of the ad, but the information is unstructured and in many cases not meaningful to quantify. An interesting observation is that the CRM responses seem to be able to reveal information which is most often associated with pure qualitative research methods, such as focus groups or in-depth interviews. The subjects, for example, not only said that the package did not look like a coffee package, they also said what kind of package they thought it looked like.

4.5 Relationships between the measures in the pretest

The third research question concerns how measures commonly included in advertising pretests are related to each other. In this section, results from a number of regression analyses aimed at answering that question will be presented and discussed.

The analyses were done using with data from both empirical studies, but in contrast to the earlier regression analyses, the analyses in this section will not include variables from adjective checklists or net cognitive response. The focus in the previous section was mainly on the ability of *ACL factors* and CRM_{Net} to explain the variation in the dependent variables. In this section, however, the focus will be on the relationship between the other constructs measured in the study, e.g., the relationship between attitude toward the ad and brand attitude. By excluding *ACL factors* and CRM_{Net}, the regression models became more parsimonious and the possible overlap between them and the other variables, e.g., ad credibility, was eliminated. The exclusion also made it possible to include the subjects from both pretest groups in Study 1 in the same analyses, thereby doubling the number of observations in each regression model. The regression models from both studies will be discussed together as the two

questionnaires, apart from the ACL and CRM, were virtually identical. There will be no analyses with ad credibility or attitude toward the ad as dependent variables, since there are no theoretically sound variables that could be used as independent variables when *ACL factors* and CRM_{Net} are excluded. (The correlation, which is identical to the beta coefficient in a bivariate regression, between $Cred_{Ad2}$ and A_{ADI} varied between .55 and .77 in Study 1 and .44 and .76 in Study 2.)

Brand benefit beliefs. The independent variables in the model with $Beliefs_{Brand}$ as dependent variable were $Cred_{Ad2}$ and A_{Ad1}. The results for Study 1 are shown in Table 65 and the results for Study 2 in Table 66. All models in both studies were statistically significant and the R^2s ranged between .19 and .35 for the Study 1 models, and between .27 and .46 for the Study 2 models.

	All *Beta*	**Painkillers** *Beta*	**Coffee** *Beta*	**Pension plan** *Beta*	**Jeans** *Beta*
$Cred_{Ad2}$	.46***	.35***	.28	.32**	.44***
A_{Ad1}	.05	.32***	.34*	.21	.00
n	279	80	55	59	85
F	43.73***	20.83***	13.56***	8.47***	9.75***
R^2	.24	.35	.34	.23	.19
Adj. R^2	.24	.33	.32	.20	.17

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 65. Study 1: Pretest regression models with $Beliefs_{Brand}$ *as dependent variable and* $Cred_{Ad2}$ *and* A_{Ad1} *as independent variables.*

Of the independent variables, $Cred_{Ad2}$ was statistically significant in three of the models for both the studies, and A_{Ad1} was significant in two of the Study 1 models, but none of the Study 2 models. In Study 1, both $Cred_{Ad2}$ and A_{Ad1} were significant in the painkiller ad model, but in the coffee ad model only A_{Ad1} was significant. It is of some interest that the beta coefficients for $Cred_{Ad2}$ and A_{Ad1} were positive in all models but one, even if they were not statistically significant in all cases. The results indicate that ad credibility had most influence on brand

benefit beliefs, but also that a relationship might exist between attitude toward the ad and brand benefit beliefs.

	All *Beta*	Painkillers *Beta*	Coffee *Beta*	Pension plan *Beta*	Jeans *Beta*
$Cred_{Ad2}$	.49***	.32*	.63***	.26	.39**
A_{Ad1}	.09	.24	.09	.35	.22
n	157	43	34	35	45
F	33.31***	7.28***	13.09***	7.83***	8.42***
R^2	.30	.27	.46	.33	.29
Adj. R^2	.29	.23	.42	.29	.25

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 66. Study 2: Pretest regression models with $Beliefs_{Brand}$ *as dependent variable and* $Cred_{Ad2}$ *and* A_{Ad1} *as independent variables.*

Brand attitude. An issue of interest in the regression models with brand attitude as dependent variable is whether brand benefit belief influences brand attitude (cf. Mittal, 1990; Fishbein and Middlestadt, 1995 ; Fishbein and Middlestadt, 1997; Rossiter and Percy, 1997). The first test of this was to use partial *F* test to see if $Beliefs_{Brand}$ made a statistically significant contribution to R^2 when added to a model that already had $Cred_{Ad2}$ and A_{Ad1} as independent variables. The increase in R^2 was statistically significant in all models in both studies, except in the model for the pension plan ad in Study 1. The average increase in R^2 was .14 (Study 1) and .19 (Study 2). Regression models with $Cred_{Ad2}$, A_{Ad1} and $Beliefs_{Brand}$ as independent variables are shown in Table 67 (Study 1) and Table 68 (Study 2). All models were statistically significant and the R^2s ranged between .36 and .51 for the Study 1 models and .29 and .61 for the Study 2 models.

	All *Beta*	Painkillers *Beta*	Coffee *Beta*	Pension plan *Beta*	Jeans *Beta*
$Cred_{Ad2}$	.18***	.40***	.01	.35**	.09
A_{Ad1}	.31***	.22**	.08	.26*	.20*
$Beliefs_{Brand}$	.38***	.20*	.66***	.09	.53***
n	279	80	55	59	85
F	96.85***	22.22***	17.76***	10.40***	22.58***
R^2	.51	.47	.51	.36	.46
Adj. R^2	.51	.45	.48	.33	.44

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 67. Study 1: Pretest regression models with A_{Brand1} as dependent variable and $Cred_{Ad2}$, A_{Ad1} and $Beliefs_{Brand}$ as independent variables.

These results support the notion that brand benefit beliefs influence brand attitude and indicate that attitude toward the ad in some cases influences brand attitude. The results regarding ad credibility were less clear. The variable $Beliefs_{Brand}$ was statistically significant in the models for all ads except for the pension plan ad in Study 1. Of the other two variables, $Cred_{Ad2}$ was statistically significant in two of the Study 1 models and none of the Study 2 models, and A_{Ad1} was significant in three of the Study 1 models and one of the Study 2 models.

	All *Beta*	Painkillers *Beta*	Coffee *Beta*	Pension plan *Beta*	Jeans *Beta*
$Cred_{Ad2}$	.12	.24	–.15	.19	.10
A_{Ad1}	.27***	.40**	–.03	.20	.00
$Beliefs_{Brand}$	.45***	.26**	.71***	.52***	.48***
n	155	43	32	35	45
F	45.06***	18.16***	5.41***	16.50***	5.59***
R^2	.47	.58	.37	.61	.29
Adj. R^2	.46	.55	.30	.58	.24

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 68. Study 2: Pretest regression models with A_{Brand1} as dependent variable and $Cred_{Ad2}$, A_{Ad1} and $Beliefs_{Brand}$ as independent variables.

Purchase intention. Three regression models with PI_{Brand} as dependent variable were evaluated. The first model included only A_{Brand1} as independent variable, the second A_{Brand1} and $Beliefs_{Brand}$, and the third A_{Brand1}, $Beliefs_{Brand}$, $Cred_{Ad2}$ and

A_{Ad1}. A partial F test showed that both additions of variables contributed statistically significantly to increasing R^2 in three of the Study 1 models. The picture was less clear for the Study 2 models; the two additions of variables only contributed statistically significantly to R^2 in one case. Despite this, a decision was taken to run the models with all variables for both studies. The average gain in R^2 from adding $Beliefs_{Brand}$ was .06 in Study 1 and .09 in Study 2 and the average gain in R^2 from adding $Cred_{Ad2}$ and A_{Ad1} to the model was .14 in Study 1 and .04 in Study 2; the models are shown in Table 69 (Study 1) and Table 70 (Study 2). All models in both studies were statistically significant and the R^2s ranged between .33 and .57 (Study 1) and .27 to .72 (Study 2).

	All *Beta*	**Painkillers** *Beta*	**Coffee** *Beta*	**Pension plan** *Beta*	**Jeans** *Beta*
$Cred_{Ad2}$	.23***	.10	.11	.42***	.25**
A_{Ad1}	.14**	.35**	.56***	.03	.05
$Beliefs_{Brand}$	.24***	.07	.11	−.07	.38***
A_{Brand1}	.17**	.17	−.07	.45***	.15
n	279	80	55	59	85
F	48.05***	9.30***	10.53***	17.82***	17.11***
R^2	.41	.33	.46	.57	.46
Adj. R^2	.40	.30	.41	.54	.43

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 69. Study 1: Pretest regression models with PI_{Brand} *as dependent variable and* $Cred_{Ad2}$, A_{Ad1}, $Beliefs_{Brand}$ *and* A_{Brand1} *as independent variables.*

In Study 1, the models for the ads for the low- and high-involvement products appeared to differ. In the models for the painkiller and coffee ads, A_{Ad1} was the only variable which was statistically significant. In the models for the pension plan and jeans ads, on the other hand, A_{Ad1} was not statistically significant. Instead, Ad_{Cred2} was statistically significant in both models and $Beliefs_{brand}$ and A_{Brand1} were significant in one model each.

The picture was different for the Study 2 models. A_{Ad1} was statistically significant in the coffee and pension plan ad models, $Cred_{Ad2}$ in the jeans ad,

Beliefs$_{Brand}$ in the coffee ad model, and A_{Brand1} in the painkiller and jeans ad models.

	All *Beta*	Painkillers *Beta*	Coffee *Beta*	Pension plan *Beta*	Jeans *Beta*
$Cred_{Ad2}$	.24**	.16	.16	−.18	.29*
A_{Ad1}	.17*	.17	.33*	.46**	−.08
$Beliefs_{Brand}$	.12	.11	.43*	.05	.02
A_{Brand1}	.21**	.53***	−.21	.38	.36**
n	154	43	32	34	45
F	20.44***	24.06***	5.77***	6.13***	3.76**
R^2	.35	.72	.46	.46	.27
Adj. R^2	.34	.69	.38	.38	.20

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 70. Study 2: Pretest regression models with PI_{Brand} *as dependent variable and* $Cred_{Ad2}$, A_{Ad1}, $Beliefs_{Brand}$ *and* A_{Brand1} *as independent variables.*

The results discussed in this section indicate that the perceived credibility of the ads influences attitude toward the ad and brand benefit beliefs. The correlation between Ad_{Cred2} and A_{Ad1} was rather strong and Ad_{Cred2} was statistically significant in six of the eight models with *Beliefs*$_{Brand}$ as dependent variable. The results were less clear for the relationships between ad credibility and brand attitude, and ad credibility and brand purchase intention. Ad_{Cred2} was statistically significant in two out of eight models with A_{Brand1} as dependent variable and three out of eight models with PI_{Brand} as dependent variable. Both of these results are in line with earlier findings. A study which did not include brand beliefs or brand purchase intention found that ad credibility was related to attitude toward the ad and brand attitude in a structural equation model (MacKenzie and Lutz, 1989).

Some evidence was found indicating that attitude toward the ad influenced brand benefit beliefs: A_{Ad1} was statistically significant in two out of eight models with *Beliefs*$_{Brand}$ as dependent variable and the coefficient was positive in all models but one. Further, results indicated that there might be a relationships

between attitude toward the ad and brand attitude and attitude toward the ad and brand purchase intention: A_{Adl} was statistically significant in four out of eight models. The influence of attitude toward the ad on brand attitude and brand purchase intention is in line with earlier results, but in the present study the relationship is less clear than in earlier studies (e.g., MacKenzie et al., 1986; MacKenzie and Lutz, 1989; Brown and Stayman, 1992; Lebenson and Blackston, 1997) This is probably due to the inclusion of brand beliefs in the models (cf. Mittal, 1990; Fishbein and Middlestadt, 1995; 1997).

The results in the regression analyses indicated that there was a relationship between brand benefit beliefs and brand attitude. The variable $Beliefs_{Brand}$ was statistically significant in seven of the eight models with A_{Brandl} as dependent variable. This result is clearly in line with the findings of Mittal (1990) and Fishbein and Middlestadt (1995) and with theoretical expectations (cf. Eagly and Chaiken, 1993).

Weak support was found for a relationship between brand benefit beliefs and brand purchase intention: $Beliefs_{Brand}$ was statistically significant in two out of eight models with PI_{Brand} as dependent variable. Somewhat surprisingly, as theory leads us to expect otherwise (cf. Eagly and Chaiken, 1993), little support was found for a relationship between brand attitude and brand purchase intention: A_{Brandl} was statistically significant in three out of eight models. Several earlier studies have found brand attitude to be positively related to brand purchase intention (e.g., Zinkhan and Fornell, 1985; Laroche et al., 1996), while others have suggested that the relationship might differ in strength between product types (Lebenson and Blackston, 1997).

The results in the regression models with $Beliefs_{Brand}$ and A_{Brandl} as dependent variables lend some weak support to the *dual mediation hypothesis*, which posits

that attitude toward the ad influences both brand cognitions (i.e., brand beliefs) and brand attitude (cf. MacKenzie et al., 1986; MacKenzie and Lutz, 1989; Brown and Stayman, 1992). A_{Ad1} was statistically significant in two out of eight models with $Beliefs_{Brand}$ as dependent variable and in four out of eight models with A_{Brand1} as dependent variable. $Beliefs_{Brand}$ was statistically significant in seven out of eight models with A_{Brand1} as dependent variable.

Two sets of results were not in line with the dual mediation hypothesis, or any of three other suggested models of how attitude toward the ad influences brand attitude, i.e., *the affect transfer hypothesis*, *the reciprocal mediation hypothesis* and *the independent influences hypothesis* (cf. MacKenzie et al., 1986; Brown and Stayman, 1992). $Cred_{Ad2}$ was statistically significant in six out of eight regression models with $Beliefs_{Brand}$ as dependent variable and in two out of eight models with A_{Brand1} as dependent variable. These relationships, if ad credibility is regarded as a type of *ad cognitions*, are not suggested in any of the above-mentioned hypotheses of how attitude toward the ad influences brand attitude.

A number of differences were found between the models for the two studies. In some instances, variables were statistically significant in the Study 1 model but not significant in the equivalent Study 2 model. Some of these cases were probably a result of the smaller number of observations in the Study 2 models, since the sign and the size of the coefficients were similar. In other cases, variables were statistically significant in the Study 2 model but not in the equivalent Study 1 model. Moreover, two of the models for the pension plan ad differed rather much between the studies with regard to the relative importance of variables. The coefficients for $Cred_{Ad2}$ and A_{Ad1} were fairly large and statistically significant in the Study 1 model with A_{Brand1} as dependent variable, while $Beliefs_{Brand}$ was the only significant variable in the equivalent Study 2 model. Similar differences were found between the Study 1 and Study 2 models

for the painkiller and pension plan ads with PI_{Brand} as dependent variable. These differences were probably caused by the fact that almost half of the subjects in Study 2 evaluated rough sketch versions of the ads.

4.6 Relationships between the measures in the posttest

The fourth research question, which will be addressed in this section, concerns the relationships between measures commonly used in advertising posttests. The analyses were done in the same way as when the relationships between the pretest variables were investigated, but since a posttest was only included in the first study, no results for Study 2 are available.

A number of regression analyses with the measures in the posttest in Study 1 are presented and discussed in this section. In all regression models, the data from all the three groups, ACL, CRM and control, in Study 1 were analyzed together.

Three of the measures in the posttest questionnaire were excluded from the analysis since there was virtually no variation in them. More than 90 percent of the subjects recognized the ads and, hence, $Recognition_{Ad}$ could not be included in the analyses. The problem was the reverse with correct brand in the measure of ad recall and brand recall. There were few positive responses on the variables $Recall_{Ad}$ (correct brand) and $Recall_{Brand}$: Between 2 and 11 subjects connected the correct brand with the recalled ads and between two and eight subjects could recall the brands. Consequently, both $Recall_{Ad}$ (correct brand) and $Recall_{Brand}$ were excluded from the analyses. Analyses were run for three dependent variables: $Beliefs_{Brand}$, A_{Brandl} and PI_{Brand}. In all models, $Cut\text{-}through_{Ad}$, $BPAR$ and $Recognition_{Brand}$ were included as independent variables; $Beliefs_{Brand}$ was included as independent variable in two models and A_{Brandl} was included one model.

Brand benefit beliefs. The independent variables in the models with $Beliefs_{Brand}$ as dependent variable were $Cut\text{-}through_{Ad}$, *BPAR* and $Recognition_{Brand}$. The results of the analyses, shown in Table 71, did not indicate any clear relationship between any of these three independent variables and the dependent variable. Only two of the models for the four ads were statistically significant and in one of them, none of the independent variables was significant. The R^2 in the two statistically significant models was .19 and .20. The only statistically significant variable in all models was $Cut\text{-}through_{Ad}$ in the coffee ad model.

	All *Beta*	**Painkillers** *Beta*	**Coffee** *Beta*	**Pension plan** *Beta*	**Jeans** *Beta*
$Cut\text{-}through_{Ad}$	–.06	–.09	.25*	.08	–.13
BPAR	.07	.07	.07	.27	–.04
$Recognition_{Brand}$	.13	–.05	.22	.16	.17
n	330	91	70	67	102
F	3.60**	.22	5.30***	5.16***	.97
R^2	.03	.01	.19	.20	.03
Adj. R^2	.02	.00	.16	.16	.00

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 71. Study 1: Posttest regression models with $Beliefs_{Brand}$ *as dependent variable and* $Cut\text{-}through_{Ad}$, BPAR *and* $Recognition_{Brand}$ *as independent variables.*

Brand attitude. In the models with A_{Brand1} as dependent variable, $Beliefs_{Brand}$ was added to the independent variables. All the models for the four ads were statistically significant and the R^2s varied between .24 and .39. The only statistically significant variable in the models was $Beliefs_{Brand}$, which was significant in the models for all four ads.

	All *Beta*	Painkillers *Beta*	Coffee *Beta*	Pension plan *Beta*	Jeans *Beta*
Cut-through$_{Ad}$	–.08	–.12	.08	.02	–.11
BPAR	.04	–.04	–.11	.22	.09
Recognition$_{Brand}$	.19**	.10	.19	–.03	.19
Beliefs$_{Brand}$	.54***	.52***	.54***	.40***	.39***
n	330	91	70	67	102
F	49.44***	9.16***	10.24***	5.84***	7.68***
R^2	.38	.30	.39	.27	.24
Adj. R^2	.37	.27	.35	.23	.21

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 72. Study 1: Posttest regression models with A$_{Brand1}$ *as dependent variable and* Cut-through$_{Ad}$, BPAR, Recognition$_{Brand}$ *and* Beliefs$_{Brand}$ *as independent variables.*

Brand purchase intention. In the models with PI_{Brand} as dependent variable, A_{Brand1} was added to the independent variables. All models were statistically significant and the R^2s varied between .17 and .46. As was the case in the previous models in this section, the two ad recall variables and the brand recognition variable contributed little to explaining the variance in the dependent variable. $Cut\text{-}through_{Ad}$ was statistically significant in one model, but the beta coefficient was about half the size of the other two significant beta coefficients. $Recognition_{Brand}$ was also statistically significant in one model, but the coefficient was, contrary to expectations, negative. $Beliefs_{Brand}$ and A_{Brand1} were statistically significant in three models each; both variables were statistically significant in the models for the two high-involvement product ads, while $Beliefs_{Brand}$ was the only significant variable in the coffee ad model, and A_{Brand1} was significant in the painkiller ad model (together with $Recognition_{Brand}$).

	All *Beta*	Painkillers *Beta*	Coffee *Beta*	Pension plan *Beta*	Jeans *Beta*
Cut-through_{Ad}	.01	–.07	.11	.18*	.02
BPAR	.04	–.08	.14	.04	.05
$\text{Recognition}_{Brand}$	–.07	.16	–.35**	–.15	–.19
Beliefs_{Brand}	.27***	.06	.44***	.31**	.34***
A_{Brand1}	.34***	.33**	–.01	.44***	.39***
n	330	91	70	67	102
F	26.06***	3.37**	3.71**	10.20***	10.66***
R^2	.29	.17	.22	.46	.36
Adj. R^2	.28	.12	.16	.41	.32

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 73. Study 1: Posttest regression models with PI_{Brand} *as dependent variable and* Cut-through_{Ad}, BPAR, $\text{Recognition}_{Brand}$, Beliefs_{Brand} *and* A_{Brand1} *as independent variables.*

The analyses of the posttest data from Study 1 did not indicate that ad cut-through, brand-prompted ad recognition or brand recognition were related to any of the communication effects. The variables *Cut-through*$_{Ad}$, and *Recognition*$_{Brand}$ were statistically significant in one model each, whereas *BPAR* was not statistically significant in any of the models. This should not be taken as a clear indication that these variables are unimportant in an advertising posttest context since method factors cannot be ruled out. It is not clear what the implications are of including brands in the study that were not available on the Swedish market, and of inserting the ads in rather unusual media.

A clear indication of a relationship between brand benefit beliefs and brand attitude was found. *Beliefs*$_{Brand}$ was statistically significant in all of the four models with A_{Brand1} as dependent variable. Indications of a direct relationship were also found between brand benefit beliefs and brand purchase intention: *Beliefs*$_{Brand}$ was statistically significant in three of the four models with PI_{Brand} as dependent variable. In the model with PI_{Brand} as dependent variable, an indication of a direct relationship could be seen between brand attitude and brand purchase intention since A_{Brand1} was significant in three of the four models. These results are in line with theory and earlier research, except for the direct

influence of brand benefit beliefs on brand purchase intention. Brand beliefs are expected to influence brand attitude and brand attitude in turn is expected to influence brand purchase intention (cf. Mittal, 1990; Eagly and Chaiken, 1993; Fishbein and Middlestadt, 1995; Fishbein and Middlestadt, 1997; Rossiter and Percy, 1997).

4.7 Predictive ability of pretest measures

In the final section of this chapter, the results from the analyses of the relationship between the variables in the pretest and the posttest will be presented. These analyses relate to the research question regarding the ability of commonly used pretest measures to predict the communication effects measured in posttests. Can results from advertising pretests be used to predict what results an ad or campaign will get in a posttest after the campaign has finished? The data in the analyses were from the pretest and posttest in Study 1 and, hence, the subjects in the control group, who did not do the pretest, were not included in the analysis.

In the first round of analyses, models with *ACL factors* and CRM_{Net} among the independent variables were run. For both types of variables, there were very few instances in which the coefficients for the variables were statistically significant *and* had the expected sign. The overall impression was that a very weak or non-existent relationship existed between the *ACL factors* and CRM_{Net} and the variables in the posttest. The few statistically significant coefficients were, in light of the cut-off point for statistical significance ($p < .10$), interpreted as a result of random variation. The results of the initial analyses led to the decision to exclude both *ACL factors* and CRM_{Net} from further analyses. This decision also meant that the number of observations in the regression models doubled, since CRM and ACL were measured in two different groups of pretest subjects.

It should be noted that the multicollinearity diagnostics did not indicate any problems in any of the models.

The results of the analyses with three dependent variables will not be discussed in detail. A large number of logistic regression models with *Cut-through*$_{Ad}$, *BPAR* and *Recognition*$_{Brand}$ as dependent variables showed that the variables in the pretest, with a few minor exceptions, did not contribute to explaining the variance in the dependent variables. Consequently, the discussion will move onto the models with *Beliefs*$_{Brand(Post)}$, $A_{Brand1(Post)}$ and $PI_{Brand(Post)}$ as dependent variables.

Brand benefit beliefs. The independent variables in the model with *Beliefs*$_{Brand(Post)}$ as dependent variable were A_{Ad1}, $Cred_{Ad2}$ and *Beliefs*$_{Brand(Pre)}$. The models for the four ads, shown in Table 74, were statistically significant with R^2s between .18 and .60. The independent variable *Beliefs*$_{Brand(Pre)}$ was statistically significant in all the models, which was not the case for any of the other independent variables.

	All *Beta*	**Painkillers** *Beta*	**Coffee** *Beta*	**Pension plan** *Beta*	**Jeans** *Beta*
$Cred_{Ad2}$	.09	.16	.21	.03	−.03
A_{Ad1}	−.04	−.05	−.19	.15	.14
$Beliefs_{Brand(Pre)}$	.63***	.35**	.75***	.52***	.57***
n	245	71	47	51	76
F	63.43***	4.93***	21.20***	9.63***	14.50***
R^2	.44	.18	.60	.38	.38
Adj. R^2	.43	.14	.57	.34	.35

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 74. Study 1: Between-test regression models with $Beliefs_{Brand(Post)}$ *as dependent variable and* $Cred_{Ad2}$, A_{Ad1} *and* $Beliefs_{Brand(Pre)}$ *as independent variables.*

To test for a statistically significant relationship between *Beliefs*$_{Brand(Post)}$ and the constructs measured in the pretest that were not included in the first model,

another regression model was run. This model had the same independent variables as the model in Table 74 and, in addition, also had $A_{Brand1(Pre)}$ and $PI_{Brand(Pre)}$ as independent variables. In one instance, one of the two additional variables was statistically significant: $A_{Brand1(Pre)}$ was significant in the model for the pension plan ad. It is interesting to note that $Beliefs_{Brand(Pre)}$ remained statistically significant in the model. The other three models remained virtually unchanged.

Brand attitude. The models with brand attitude as dependent variable are shown in Table 75. All models were statistically significant and the R^2s varied between .31 and .39. In all models, $A_{Brand1(Pre)}$ was statistically significant. For $Cred_{Ad2}$ and $Beliefs_{Brand(Pre)}$, the results were less clear. A statistically significant *positive* relationship was found between $Cred_{Ad2}$ and the dependent variable in the painkiller ad model while a significant *negative* relationship between $Cred_{Ad2}$ and $A_{Brand(Post)}$ was found in the jeans ad model. $Beliefs_{Brand(Pre)}$ was statistically significant and negative in the painkiller ad model.

	All *Beta*	**Painkillers** *Beta*	**Coffee** *Beta*	**Pension plan** *Beta*	**Jeans** *Beta*
$Cred_{Ad2}$	.02	.25*	.30	.03	−.36**
A_{Ad1}	−.07	−.05	−.25	−.21	.13
$Beliefs_{Brand(Pre)}$	.05	−.27**	.27	.00	.19
$A_{Brand1(Pre)}$	.61***	.49***	.35**	.66***	.50***
n	245	71	47	51	76
F	35.76***	7.32***	6.73***	6.09***	8.52***
R^2	.37	.31	.39	.35	.32
Adj. R^2	.36	.27	.33	.29	.29

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 75. Study 1: Between-test regression models with $A_{Brand1(Post)}$ *as dependent variable and* $Cred_{Ad2}$, A_{Ad1} $Beliefs_{Brand(Pre)}$ *and* $A_{Brand1(Pre)}$ *as independent variables.*

Brand purchase intention. In the analyses with $PI_{Brand(Post)}$ as dependent variable, all constructs measured in the pretest, except ACL and CRM variables, were

included in the model, which is shown in Table 76. All models were statistically significant and the R^2s ranged from .17 to .51. The variable $PI_{Brand(Pre)}$ was statistically significant for all models, except the painkiller ad model. The pretest measure of brand benefit beliefs, $Beliefs_{Brand(Pre)}$, was not statistically significant in any model; A_{Ad1} was significant in one model, but the sign of the coefficient was contrary to expectations. $Cred_{Ad2}$ was statistically significant in the two ads for low-involvement products.

	All *Beta*	**Painkillers** *Beta*	**Coffee** *Beta*	**Pension plan** *Beta*	**Jeans** *Beta*
$Cred_{Ad2}$	.13	.31*	.46**	−.27	−.21
A_{Ad1}	−.09	−.29*	−.15	.06	.04
$Beliefs_{Brand(Pre)}$	−.02	−.23	−.01	.12	.19
$A_{Brand1(Pre)}$	.18**	.24	−.06	.15	−.03
$PI_{Brand(Pre)}$	.44***	.15	.44**	.66***	.69***
n	245	71	47	51	76
F	23.34***	2.75**	6.68***	8.16***	14.73***
R^2	.33	.17	.45	.48	.51
Adj. R^2	.31	.11	.38	.42	.48

*** $p < .01$, ** $p < .05$, * $p < .10$

Table 76. Study 1: Between-test regression models with $PI_{Brand(Post)}$ *as dependent variable and* $Cred_{Ad2}$, A_{Ad1}, $Beliefs_{Brand(Pre)}$, $A_{Brand1(Pre)}$ *and* $PI_{BRand(Pre)}$ *as independent variables.*

In sum, two clear indications were found in the regression models with measures from the pretest as independent variables and measures from the posttest as dependent variables. First, the processing measures (i.e., the ACL and CRM related variables, ad credibility and attitude toward the ad) do not predict the communication effects outcome. The processing measure variables were, with a few exceptions, not at all related to the posttest communication effects measures. Second, the best predictors of the communication effects are measurements of the same communication effects in the pretest. In the models with $Beliefs_{Brand(Post)}$ as dependent variable, $Beliefs_{Brand(Pre)}$ was the only statistically significant variable in all the models. In the models with $A_{Brand(Post)}$ as dependent variable, $A_{Brand(Pre)}$ was statistically significant in all models, while

the results for the other independent variables were mixed and contradictory to theoretical expectations. The results in the regressions with $PI_{Brand(Post)}$ as dependent variable were in line with those for the other two dependent variables. $PI_{Brand(Pre)}$ was statistically significant in the models for three of the ads, while $Cred_{Ad2}$, which was significant in two models, was the only other variable that was significant and had a coefficient with an expected sign.

At the end of Chapter 2, it was suggested that the correspondence between measures (cf. Eagly and Chaiken, 1993; Fishbein and Middlestadt, 1995) could lead us to expect that a pretest measure of a construct could be a good predictor of the same construct in a posttest. An alternative suggestion was that the formation of communication effects takes longer time than the time period between exposure and measurement in the pretest and that more than one exposure situation might be needed for communication effects to be formed. The alternative suggestion implies that processing measures might be better than communication effect measures at predicting communication effects at a later stage. The results presented in this section support that the correspondence between measures is more important than the time span between exposure and measurement and the number of exposures.

5 Conclusions

This chapter starts with a summary of the results presented in Chapter 4, followed by a discussion of the conclusions that can be drawn from the results and their possible implications. Thereafter the limitations of the two empirical studies will be discussed. The chapter concludes with a comment on managerial implications and suggestions for further research.

5.1 Summary of the results

A total of seven research questions (presented in Section 2.15) formed the basis of the two empirical studies in this thesis. The analyses presented in the previous chapter were aimed at answering these questions. In this section, the results of the analyses for each of the research questions will be summarized and the results with regard to the relationships between constructs will then be summarized in a figure.

Research questions 1 and 6 asked whether single-item and multiple-item operationalizations of ad credibility, attitude toward the ad and brand attitude were interchangeable. This was investigated in a series of analyses that showed that the two types of operationalizations were correlated and that the estimated single-item reliabilities were acceptable or good. The analyses also showed that the measures did not differ in explanatory power in regression analyses and repeated measures MANOVAs did not show any statistically significant main effects for type of measure. The conclusion was therefore was therefore drawn that the measures were interchangeable and, as a consequence, the single-item operationalizations of attitude toward the ad and brand attitude were used in the remaining analyses in the thesis.

The second research question concerned whether adjective checklists and cognitive response measurement are interchangeable or complementary. The first issue for investigation was whether ACLs overdiagnose ads, which has been suggested in the literature (Rossiter and Percy, 1997). The analysis showed that the ACLs in both studies elicited many more responses than the CRM. In Study 1, the mean number of ACL responses was twice the mean number of CRM responses for all four ads. The difference was smaller in Study 2, but still substantial. The greater number of responses elicited by ACLs could indicate that they overdiagnose ads, but caution is warranted as the two types of responses are different in terms of their content and the effort required of respondents.

The main comparison between ACLs and CRM was carried out by using variables from the two measures as independent variables in regression models. The results did not provide any solid ground for deciding whether ACL variables are better at predicting other pretest variables than CRM variables, or vice versa. The explained variance (R^2)in the regression models did not differ systematically in favor of either of the two measures.

The data from Study 2 allowed an investigation of whether the two measures were complementary since both types of measures were included in the pretest questionnaire. An analysis was done to see if adding ACL variables to the independent variables in a model that already has CRM variables among the independent variables increased the explained variance of the model. The analysis was also done the other way around, i.e., CRM variables were added to a model with ACL variables. This analysis indicated that adding an ACL to a pretest that already has CRM could improve the ability to explain other pretest variables, but that it is doubtful whether the same is true if CRM is added to a pretest that already has an ACL.

In addition to the comparison of the explanatory power of the two types of variables, the regression analyses including the ACL and CRM variables also provided information on the relationship between the two measures and other measures. Some *ACL factors* were almost consistently statistically significant in the regression models with $Cred_{Ad}$ and A_{Ad} as dependent variables. Also CRM_{Net} was statistically significant in almost all models with $Cred_{Ad}$ and A_{Ad} as dependent variables. The relationships between *ACL factors* and CRM_{Net}, and brand benefit beliefs and brand attitude were more tenuous, but the results do not rule out that the variables were related.

Research question 3 concerned the relationships between intermediate advertising effectiveness measures included in advertising pretests. This question was examined using regression analyses with data from both Study 1 and Study 2. Regression models were run with attitude toward the ad (A_{Ad1}), brand benefit beliefs ($Beliefs_{Brand}$), brand attitude (A_{Brand1}) and brand purchase intention (PI_{Brand}) as dependent variables. The independent variables in these models were the intermediate measures assumed to precede each of the constructs, except the ACL and CRM variables. The results of the analyses were mainly, but not entirely, in line with results in earlier research that had tested these relationships.

The analyses indicated that ad credibility ($Cred_{Ad2}$) influenced A_{Ad1} and $Beliefs_{Brand}$, but that it was only weakly related to A_{Brand1} and PI_{Brand}. A weak indication of a relationship was found between A_{Ad1} and $Beliefs_{Brand}$ and A_{Brand1}, but not between A_{Ad1} and any other variables. A clear indication of a relationship was found between $Beliefs_{Brand}$ and A_{Brand1} and a somewhat weaker indication of a relationship between $Beliefs_{Brand}$ and PI_{Brand}. The analyses also showed that A_{Brand1} in some, but not all, cases influenced PI_{Brand}.

Research question 4, which concerned the relationships between intermediate measures in advertising posttests, was investigated with a number of regression analyses of the data from Study 1. Unfortunately, the measures of ad recognition, ad recall (with correct branding) and brand recall had to be excluded from the analyses because of their skewed response patterns. It was possible to include ad cut-through ($Cut\text{-}through_{Ad}$), brand-prompted ad recall ($BPAR$) and brand recognition ($Recognition_{Brand}$) in the analyses, but, with a few exceptions, these variables were not statistically significantly related to any of the other posttest variables. The other posttest measures, i.e., $Beliefs_{Brand}$, A_{Brand1} and PI_{Brand}, were related to each other as expected from theory and earlier research. The analyses clearly indicated that $Beliefs_{Brand}$ is directly related to both A_{Brand1} and PI_{Brand} and that A_{Brand1} influences PI_{Brand}.

Research question 5 was investigated in much the same fashion as research questions 3 and 4, but both logistical and multiple regression analyses were used. The analyses showed that pretest measures of *processing* could not predict communication effects in the posttest, while pretest measures of *communication effects* could. The regression models with data from both the pretest and the posttest showed that the processing measures, i.e., *ACL factors* and CRM_{Net}, $Cred_{Ad2}$ and A_{Ad1}, did not explain any of the variance in the posttest variables. The only relationships that were found in the regression models were between the pretest and posttest measures of the same constructs. The analysis clearly indicated that a relationship between $Beliefs_{Brand(Pre)}$ and $Beliefs_{Brand(Post)}$, $A_{Brand1(Pre)}$ and $A_{Brand1(Post)}$, and $PI_{Brand(Pre)}$ and $PI_{Brand(Post)}$. In addition, the descriptive results showed that the mean scores on the communication effects measures in the posttest were lower than the mean scores on the communication effects measures in the pretest, i.e., pretest scores should be adjusted downwards when level of posttest communication effects are predicted.

The results of the analyses pertaining to research questions 3, 4 and 5 are summarized in Figure 4. A relationship between two constructs has been drawn with a continuous arrow if the analyses consistently indicated a relationship, i.e., if six out of eight or three out of four regression coefficients were statistically significant and positive and no negative coefficients were to be found in any of the models. In cases where the indications of a relationship were weaker, i.e., there were at least two out of eight statistically significant coefficients, but no statistically significant negative coefficients, the relationship between constructs is depicted with a dashed arrow. Figure 4 also shows the relationship between adjective checklists and cognitive response measurement and the intermediate constructs, i.e., the results from the analyses pertaining to research question 2. In the figure, the two measures have been drawn together and relationships with the intermediate constructs are indicated even if only one of the measures was related to it.

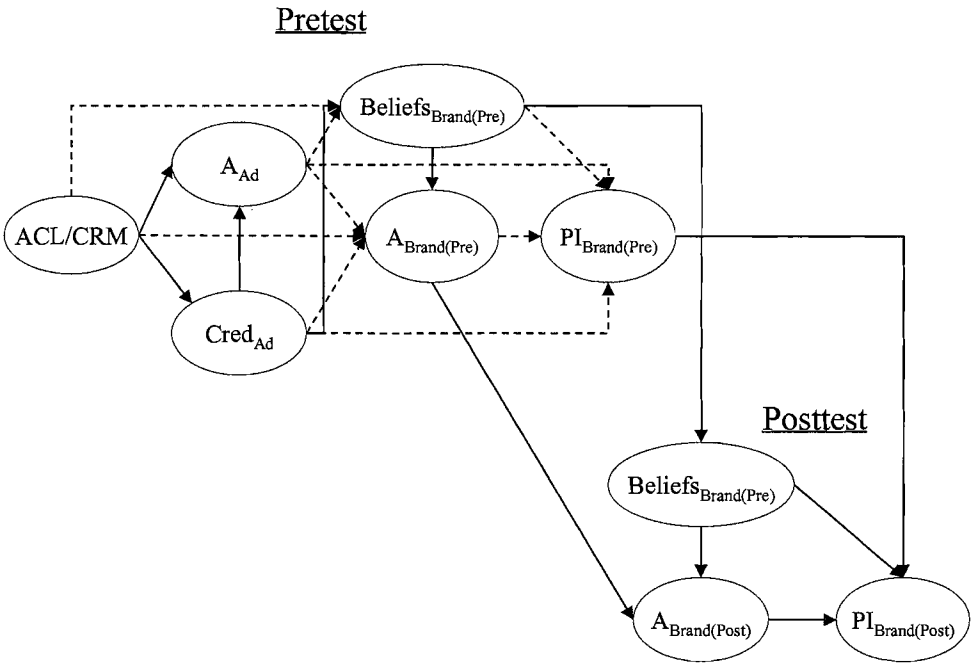

Figure 4. Relationships identified between constructs in the analyses of the two empirical studies.

Research question 7 was only investigated in Study 2 and it concerned whether rough sketch and finished versions of ads give similar results in advertising pretests. Unfortunately, the limited number of subjects in Study 2 did not allow for a thorough investigation of this research question, but the descriptive statistics indicated that pretest results differ depending on the finish of ads. The mean scores on the processing and communication effects measures were different for finished and rough versions of the four ads, but the direction of the difference varied. There were also two measures, single-item attitude toward the ad (A_{Ad1}) and single-item brand attitude (A_{Brand1}), for which the mean scores were similar for rough and finished versions of the ads.

5.2 Discussion

The three research problems for this thesis originated in the abstract concepts of constructs, operationalizations and nomological networks (Cronbach and Meehl, 1955). The aim in this thesis has been to use these abstract concepts to elicit and investigate issues that have concrete practical implications for both academics and practitioners. In this section, some conclusions and implications of relevance to both academics and practitioners will be discussed, while a number of implications of particular relevance to practitioners will be discussed in Section 5.3.

The literature review and the results in the analysis of the relationships between the pretest variables demonstrated the importance of including as many relevant parts of the nomological network (Cronbach and Meehl, 1955) as possible. The results showed, for example, that brand benefit beliefs influence brand attitude and that the relationship between attitude toward the ad and brand attitude was more tenuous, which was in line with previous research (Gardner, 1985; Mittal, 1990; Fishbein and Middlestadt, 1995). In research which has excluded brand beliefs, the relationship between attitude toward the ad and brand attitude has generally been found to be strong (e.g., MacKenzie et al., 1986; MacKenzie and Lutz, 1989; Brown and Stayman, 1992).

Another illustration of the importance of including all relevant variables is provided by the results of the analyses including the variables from adjective checklists and cognitive response measurement. The relationship between ACL or CRM variables and brand attitude was tenuous, while earlier research which has excluded other variables has found it to be strong (e.g., Wright, 1973;

Zinkhan and Fornell, 1985; Zinkhan and Burton, 1989).[3] The present and earlier results show that academics and practitioners alike should try to measure and include the relevant constructs in the nomological network that their target construct(s) are part of. The reason for this is that omission of important constructs is likely to lead to erroneous conclusions, e.g., that a direct relationship exists between two constructs when in fact it does not exist, or that the strength of the influence of one construct on another is exaggerated. The results in this thesis provide some guidance as to which constructs should be included to avoid excluding important variables in advertising research; Figure 4 in the previous section shows which constructs are likely to be related in advertising research.

In the introduction to Chapter 2, it was argued that intermediate advertising effectiveness measures in many cases could be better as measures of advertising effectiveness than sales or other outcome variables since many other factors besides advertising influence the behavioral outcome. This means that prediction of the outcome in terms of intermediate effects should be of interest to advertisers. The results of the analysis of the ability of pretest measures to predict post-campaign communication effects showed that processing measures, e.g., ad credibility and attitude toward the ad, could not predict the outcome in terms of communication effects at a later point in time. These results indicate, for example, that the attention that liking the ad has received as a predictor of advertising effectiveness (e.g., Haley and Baldinger, 1991) has not been entirely warranted. The strong relationship between communication effects in the pretest and corresponding communication effects in the posttest indicates that if prediction of campaign effectiveness is the sole purpose of pretesting, then it

[3] Note that the results in the study by Wright (1973) are not directly comparable to the results in the present study since he coded the cognitive responses into the categories *counterargument*, *support argument* or *source derogation*. It cannot be ruled out that this could have contributed to the difference in results.

should suffice to measure communication effects in the pretest. The main purpose of the processing measures should be diagnostic, i.e., they should mainly be used to understand why or why not certain communication effects were achieved (cf. Rossiter and Percy, 1997). This use of the processing measures is supported by the results in the analyses of the pretest data, which indicated that certain processing measures were related to certain communication effects measured in the pretest.

The empirical tests of whether attitude toward the ad (liking the ad), brand attitude and ad credibility could be operationalized with single-item measures indicated that this was, in fact, possible. The results showed that, for all practical purposes, single-item and multiple-item measures of the three constructs were interchangeable. These results indicate that practitioners can continue to use single-item operationalizations of attitude toward the ad and brand attitude without any negative effects on the quality of their research. They also suggest that academics could shorten questionnaires, and thereby reduce the burden placed on respondents as well as the costs associated with collecting and entering data, without impairing the quality of their research.

The literature on attitude toward the ad and liking the ad indicated that the two are, in fact, the same construct. This was confirmed by the analysis that investigated whether measures of the two were interchangeable. That attitude toward the ad and liking the ad are the same construct might not come as a great surprise to many researchers, academic or practitioners within the field of advertising. Nevertheless, it is of value to the research community to be able to conclude that this is the case. On the basis of this conclusion, it is possible to consolidate the literature and earlier research on the two constructs and reduce the risk of misunderstanding what is actually assessed by the two measures (cf. Peter, 1981).

The analyses of the adjective checklists and the cognitive response measurement in the two studies did not reveal whether one or the other was better at explaining the variance in various dependent variables. In the analyses of the data from Study 1, the adjusted R^2 was in some cases higher with *ACL factors* in the model and in some cases higher with CRM_{Net} in the model; the results were similar in the analyses of the Study 2 data. The comparisons between the two measures were limited in two respects. First, the ACLs were not tailor-made for each ad, as suggested by Rossiter and Percy (1997), which might have improved the explanatory power of the ACL variables. Second, the CRM responses were not divided into different types of responses, e.g., ad-related or brand-related, before the net cognitive response was calculated, which could have improved the explanatory power of the CRM variables. Before more sophisticated empirical work has been carried out, it is not possible to conclude whether ACLs or CRM is better at explaining the variance in dependent variables or whether the two measures are complementary.

The literature review suggested that the theoretical foundation for adjective checklists and, to a lesser extent, cognitive response measurement is weak and it was suggested that the two measures might also measure underlying constructs other than *acceptance*. Moreover, the results in the regression analyses indicated that the two measures and the measure of ad credibility overlap and that ACLs and CRM, in their present form, are of limited value in explaining brand beliefs and brand attitude. Taken together, the literature review and the empirical results point to a need for more work on theory regarding adjective checklists and cognitive response measurement.

The main theoretical need is to identify the constructs that are, or could be, measured by ACLs, CRM or both. If the constructs are clearly identified, then

measures could be developed for each construct and the relationship between each construct and other constructs analyzed separately. This would make the purpose of measurement clear, and analytical results could be interpreted within a more clearly defined theoretical framework. Constructs that might underlie ACLs or CRM could be found, for example, in the processing step in the six-step effects sequence (Rossiter and Percy, 1997), which includes not only acceptance but also *attention*, *learning* and *emotional responses*.

The literature review indicated that measures of brand benefits have been neglected by many advertising researchers for the last twenty years in spite of sound theoretical and empirical support for their inclusion in advertising effectiveness research. The results in this thesis further strengthen the argument that brand beliefs should be included in advertising effectiveness measurement since a clear indication of a relationship between brand beliefs and brand attitude was found, and between brand beliefs and brand purchase intention, although it was slightly weaker. The measurement of brand beliefs requires more work by researchers than, for example, measures of ad credibility or attitude toward the ad since they cannot be standardized for different ads. Brand beliefs are of central importance for the formation of brand attitudes and should therefore not be left out of advertising effectiveness measurement.

5.2.1 Limitations of the empirical studies

In the introduction to Chapter 3, mundane realism (Swieringa and Weick, 1982) in laboratory research and its possible negative effect on external validity was discussed. It was argued that mundane realism should be of limited importance in the present context as the type of generalizability sought was *theory application* and that it could even be an advantage to reduce variance extraneous to theory (Calder et al., 1981; 1982). According to this line of thinking, the

results in this thesis should be applicable to the real world, but the exact outcome in a given situation should vary both according to theoretical expectations and depending on situational influences extraneous to theory. Lack of mundane realism, in short, is not regarded as a serious limitation of the two studies in this thesis.

A concrete limitation of the empirical studies is that the pre-measure survey did not include a measure of whether the subjects were familiar with the brands used in the empirical studies. Consequently, it is not certain that the brands were unknown to all subjects as they might have come across them during visits to other countries. If this is true for some of the subjects, it would mean that their responses to the measures in the study might have been influenced not only by the ads in the study but also by their earlier experience. Although the effect of this on these studies is probably minor, it could be of value in future research to control for possible effects by measuring brand awareness before choosing the brands to be included in test ads. An alternative procedure would be to use fake brands.

The use of brands that are supposedly new to a market might limit the results of the studies because it cannot be taken for granted that advertising works in the same way for established and new brands. Certain findings indicate, for example, that new brands are more responsive to advertising than established brands (e.g., Blair and Rosenberg, 1994; Lodish et al., 1995). For this reason, it cannot be ruled out that the relationships between the constructs in the thesis would have been different if the ads had been for established brands. Caution is warranted, therefore, before results are generalized to include established brands.

The ads in the empirical studies were magazine ads and might limit, therefore, the extent to which the results can be generalized to ads in other media. Research comparing how advertising works in different media is rare, but most models of how advertising works assume, implicitly or explicitly, that it works in the same way irrespective of media (cf. Aaker et al., 1992; Rossiter and Percy, 1997; Tellis, 1998; Vakratsas and Ambler, 1999). Although it seems unlikely, for instance, that the six-step effects sequence (Rossiter and Percy, 1997) should be different for different media, it cannot be ruled out that the processing of ads is different for different media. Attitude toward the ad, for example, might influence brand attitude more in TV advertising than in magazine advertising. This could mean that the relationships which involve the processing constructs found in the empirical studies might not be exactly the same for other media types. However, the limitation with regard to media type need not mean that the relationships between the communication effects constructs are different for advertising in other media types. Moreover, the other findings in the study, e.g., that single-item measures of three constructs are interchangeable with multiple-item measures of the same constructs, should not be influenced by the choice of advertising media.

The final limitation to be discussed concerns the measurement of brand benefit beliefs. The benefits that were included in the measures of brand benefit beliefs were judgmentally selected and pretested in a small group. Measurement was made by asking, for each benefit, to what extent the product in the ad delivered that benefit. According to Fishbein and Middlestadt (1995), salient beliefs should be elicited from a sample of the population and both belief strength and the evaluation of each belief should be included in the measurement. It is possible, therefore, that the brand benefit beliefs in the empirical studies were not measured satisfactorily, both with regard to which benefits were measured and the evaluative component of the beliefs. However, the relationship between

brand benefit beliefs and the other measures made theoretical sense, which indicates that the benefit beliefs measurement in the studies was not entirely off track. In both the pretests and the posttest, the brand benefit beliefs measure was strongly related to brand attitude and to brand purchase intention. These relationships provide an indication of construct validity (cf. Cronbach and Meehl, 1955; Peter, 1981).

5.3 Managerial implications

In addition to the implications relevant to academics and practitioners alike (which were discussed in Section 5.2) the research in this thesis has specific implications of particular relevance to advertising practitioners. Perhaps the most important implication is that no single, universal measure of advertising effectiveness exists. This has been pointed out by other authors (e.g., Hansen, 1998) and the present research has found no reason to challenge this view. The research in this thesis illustrates further that advertising is a complex process which cannot be evaluated and understood with a single measure, e.g., attitude toward the ad or brand recall.

Another management implication that has been discussed by other authors is the danger of relying on standardized advertising effectiveness measurement (cf. Percy and Rossiter, 1997; Rossiter and Percy, 1997). In the present work, the measurement of brand benefit beliefs, brand attitude and brand purchase intention illustrates the importance of using tailor-made research. The benefits that are included in the brand benefit belief measurement are different for different product categories and they could be different for different brands within the same product category. In the case of brand attitude and brand purchase intention, it is important that the measures correspond as closely as possible – with regard to target, action, context and time – to the object of

measurement (cf. Eagly and Chaiken, 1993; Fishbein and Middlestadt, 1995). For example, measures of brand attitude should be specific with regard to usage situation and measures of brand purchase intention should be specific with regard to time period and behavior in terms of trial, purchase or usage (cf. Rossiter and Percy, 1997).

The differences between the four ads in the relationships in many of the regression analyses points to the importance of learning from advertising effectiveness measurement how the advertising for a particular brand within a particular product category works. Even if the majority of the constructs underlying the measures included in this thesis are relevant for most advertising, the relative importance of different constructs will vary between different product categories, different brands and different ads for the same brand (cf. Mehta, 1994). Each advertiser should keep track of effectiveness measurement results to build up an understanding of how their advertising works and to learn before future campaigns.

5.4 Further research

One area requiring further research is was briefly mentioned in the discussion of the results (Section 5.2), namely the theoretical background of adjective checklists and cognitive response measurement. Which constructs are actually measured by these two measures needs to be clarified and an examination of whether the measures, the coding of the responses or the measurement procedure are in need of change is required. This must be investigated both by a review of theory, from the field of advertising and other relevant fields, and with empirical studies.

A solid theoretical background for adjective checklists could, for example, help to shed light on the issue of whether the items in ACLs should be standardized and applicable to all ads (e.g., Wells, 1964; Leavitt, 1970; Wells et al., 1971; Leavitt, 1975; Aaker and Bruzzone, 1981) or whether they should be tailor-made for each ad (Rossiter and Percy, 1997). The two main benefits of standardized ACLs is that they do not require any work when the questionnaire is designed and that results for different ads can be easily compared if the same items are used for all ads. A drawback is that standardized ACLs might not be able to elicit the responses relevant for a particular ad, as they do not cover ad-specific features. These two approaches to designing ACLs could be compared in the same fashion as ACLs and CRM were compared in the first empirical study in this thesis, i.e., by including them in different pretest questionnaires that evaluate the same ads.

Another question for further research is whether certain groups of ads behave similarly in effectiveness measurement. The results in this thesis indicate that the relationships between the constructs underlying the measures are different for different ads, and the question therefore, is whether all ads are unique or if certain groups of ads elicit similar responses from the target audience? Rossiter and Percy (1997), for example, suggest that different constructs vary in relevance in the four cells in the *Rossiter-Percy grid*, e.g., that attitude toward the ad is mainly relevant for low-involvement transformational advertising. It would be of great value if the relationship between different advertising constructs were investigated for groups of ads with common characteristics, e.g., ads belonging to the same cell in the Rossiter-Percy grid or ads for products of a certain type. This research could be carried out in the same fashion as the empirical research in this study, but with a larger number of ads divided into, for example, two groups with similar characteristics. To reduce the burden on

respondents, adjective checklists or cognitive response measurement could be omitted from the pretest questionnaire.

A final suggestion for further research concerns the measurement of brand benefit beliefs. The benefits included in the brand benefit beliefs measures in the empirical studies were judgmentally selected. An alternative advocated by, among others, Fishbein and Middlestadt (1995) is to elicit salient beliefs from a sample of the target population. It would be useful to compare these two methods, and perhaps others, of selecting benefits or attributes to include in advertising research. One issue in this context is whether benefits which could be, but are not necessarily, associated with all brands within a product category are more or equally relevant than benefits associated with the ad(s) for a particular brand.[4] An example of this is provided by the coffee ad included in the two empirical studies. The ad had a sensual, almost erotic touch and the responses in CRM indicated that it conveyed benefits unassociated with the actual product, e.g., tenderness, sensuality, or sexual gratification. The product category of coffee, in contrast, would probably elicit benefits or attributes along such as, good taste, value for money and exclusive, if a group of consumers were asked to list benefits or attributes associated with the coffee in general. The question is whether benefits related to product category or ad, would be the most important for the formation of brand attitude following exposure to an ad? It would be possible to use alternative procedures to elicit benefits and to include the alternative benefits in different pretests. The alternative brand benefit belief measures could then be compared by including them in regression analyses with brand attitude as the dependent variable.

[4] The author is indebted to Larry Percy for bringing this issue to his attention.

6 References

Aaker, David A. (1992), "The Value of Brand Equity," *Journal of Business Strategy,* (4), 27-32.

Aaker, David A. (1996), "Measuring Brand Equity Across Products and Markets," *California Management Review,* (3), 102-120.

Aaker, David A., Rajeev Batra and John G. Myers (1992), *Advertising Management.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

Aaker, David A. and Phillip K. Brown (1972), "Evaluating Vehicle Source Effects," *Journal of Advertising Research,* (4), 11-16.

Aaker, David A. and Donald E. Bruzzone (1981), "Viewer Perceptions of Prime-Time Television Advertising," *Journal of Advertising Research,* (5), 15-23.

Aaker, David A. and James M. Carman (1982), "Are You Overadvertising?," *Journal of Advertising Research,* (4), 57-70.

Aaker, David A. and John G. Myers (1982), *Advertising Management.* Englewood Cliffs, N.J.: Prentice-Hall, Inc.

Aaker, David A. and Donald Norris (1982), "Characteristics of TV Commercials Perceived as Informative," *Journal of Advertising Research,* (2), 61-70.

Aaker, David A. and Douglas M. Stayman (1990), "Measuring Audience Perceptions of Commercials and Relating Them to Ad Impact," *Journal of Advertising Research,* (4), 7-17.

Adams, Anthony J. and Margaret Henderson Blair (1992), "Persuasive Advertising and Sales Accountability: Past Experience and Forward Validation," *Journal of Advertising Research,* (2), 20-25.

Anderson, James C. and David W. Gerbing (1988), "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach," *Psychological Bulletin,* (3), 411-423.

Ang, Ien (1991), *Desperately Seeking the Audience.* London: Routledge.

Appel, Valentine (1993), "Anatomy of a Magazine Audience Estimate: The ARF Comparability Study Revisited," *Journal of Advertising Research,* (1), 11-17.

Appel, Valentine and Babette Jackson (1975), "Copy Testing in a Competitive Environment," *Journal of Marketing,* (1), 84-86.

Axelrod, Joel N. (1963), "Induced Moods and Attitudes toward Products," *Journal of Advertising Research,* 19-24.

Axelrod, Joel N. (1968), "Attitude Measures That Predict Purchase," *Journal of Advertising Research,* (1), 3-17.

Bagozzi, Richard P. and Alvin J. Silk (1983), "Recall, Recognition, and the Measurement of Memory for Print Advertisements," *Marketing Science,* (2), 95-134.

Barclay, William D., Richard M. Doub and Lyron T. McMurtrey (1965), "Recall of TV Commercials by Time and Program Slot," *Journal of Advertising Research,* 41-47.

Baron, Reuben M. and David A. Kenny (1986), "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations," *Journal of Personality and Social Psychology,* (6), 1173-1182.

Batra, Rajeev, Donald R. Lehman, Joanne Burke and Jae Pae (1995), "When Does Advertising Have an Impact? A Study of Tracking Data," *Journal of Advertising Research,* (5), 19-32.

Batra, Rajeev and Michael L. Ray (1986), "Affective Responses Mediating Acceptance of Advertising," *Journal of Consumer Research,* (2), 234-249.

Baumgartner, Hans and Christian Homburg (1996), "Applications of Structural Equation Modeling in Marketing and Consumer Research: A Review," *International Journal of Research in Marketing,* (2), 139-161.

Bearden, William O., Richard G. Netemeyer and Mary F. Mobley (1993), *Handbook of Marketing Scales*. Newbury Park, California: SAGE Publications, Inc.

Belk, Russel W. (1985), "Issues in the Intention - Behavior Discrepancy," in *Research in Consumer Behavior: A Research Annual,* J. N. Sheth, ed. Stanford, CT: JAI Press, Inc.,.1-34.

Bergkvist, Lars (1999a), "Competing in Advertising Effectiveness: An Analysis of the 1996 British Advertising Effectiveness Award Case Histories," Paper presented at the AEJMC annual convention, New Orleans, August 4-7.

Bergkvist, Lars (1999b), "Swedish Advertising Effectiveness: An analysis of the Swedish 1998 Advertising Effectiveness Awards," Paper presented at the 15th Nordic Conference on Business Studies, Helsinki, Finland, August 19-21.

Biel, Alexander L. (1996), "Do You Really Want to Know?," *Journal of Advertising Research,* (2), RC/2-RC/7.

Biel, Alexander L. and Carol A. Bridgwater (1990), "Attributes of Likable Television Commercials," *Journal of Advertising Research,* (3), 38-44.

Blair, Margaret Henderson (1987), "An Empirical Investigation of Advertising Wearin and Wearout," *Journal of Advertising Research,* 45-50.

Blair, Margaret Henderson and Karl E. Rosenberg (1994), "Convergent Findings Increase Our Understanding of How Advertising Works," *Journal of Advertising Research,* (3), 35-45.

Bloom, Derek, Andrea Jay and Tony Twyman (1977), "The Validity of Advertising Pretests," *Journal of Advertising Research,* (2), 7-16.

Bogart, Leo (1976), "Mass advertising: the message, not the measure," *Harvard Business Review,* 107-116.

Bogart, Leo, B. Stuart Tolley and Frank Orenstein (1970), "What One Little Ad Can Do," *Journal of Advertising Research,* (4), 3-13.

Bray, James H. and Scott E. Maxwell (1985), *Multivariate Analysis of Variance.* Beverly Hills: SAGE Publications.

Briggs, Rex and Nigel S. Hollis (1997), "Advertising on the Web: Is There Response before Click-Through?," *Journal of Advertising Research,* (2), 33-45.

Broach, V. Carter, Thomas J. Page and R. Dale Wilson (1997), "The Effects of Program Context on Advertising Effectiveness," in *Measuring Advertising Effectiveness,* W. D. Wells, ed. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.,.203-214.

Broadbent, Simon (1995), *Best Practice in Campaign Evaluation.* London: IPA.

Broadbent, Tim (1999), "What's the Use of Tracking Studies?," *Admap,* (400), 18-20.

Brown, Gordon (1985), "Tracking Studies and Sales Effects: A U.K. Perspective," *Journal of Advertising Research,* (1), 52-64.

Brown, Nigel A. and Ronald Gatty (1967), "Rough vs. Finished TV Commercials in Telpex Tests," *Journal of Advertising Research,* (4), 21-24.

Brown, Steven P. and Douglas M. Stayman (1992), "Antecedents and Consequences of Attitude toward the Ad: A Meta-analysis," *Journal of Consumer Research,* (1), 34-51.

Bruzzone, Donald E. and Deborah J. Tallyn (1997), "Linking Tracking to Pretesting with an "ARM"," *Journal of Advertising Research,* (3), 74-79.

Buzzell, Robert D., Marshall Kolin and Malcolm P. Murphy (1965), "Television Commercial Test Scores and Short-Term Changes in Market Shares," *Journal of Marketing Research,* 307-313.

Cacioppo, John T. and Richard E. Petty (1981), "Social Psychological Procedures for Cognitive Response Assessment: The Thought-Listing Technique," in *Cognitive Assessment,* T. V. Merluzzi, C. R. Glass and M. Genest, eds. New York: The Guilford Press,.309-342.

Calder, Bobby J., Lynn W. Phillips and Alice M. Tybout (1981), "Designing Research for Application," *Journal of Consumer Research,* (2), 197-207.

Calder, Bobby J., Lynn W. Phillips and Alice M. Tybout (1982), "The Concept of External Validity," *Journal of Consumer Research,* (3), 240-244.

Chattopadhyay, Amitava and Joseph W. Alba (1988), "The Situational Importance of Recall and Inference in Consumer Decision Making," *Journal of Consumer Research,* (1), 1-12.

Churchill, Gilbert A. (1979), "A Paradigm for Developing Better Measures of Marketing Constructs," *Journal of Marketing Research,* (1), 64-73.

Clancy, Kevin J. and Robert Garsen (1970), "Why Some Scales Predict Better," *Journal of Advertising Research,* (5), 33-38.

Clancy, Kevin J. and David M. Kweskin (1971), "TV Commercial Recall Correlates," *Journal of Advertising Research,* (2), 18-20.

Clancy, Kevin J. and Lyman E. Ostlund (1976), "Commercial Effectiveness Measures," *Journal of Advertising Research,* (1), 29-34.

Cohen, Jacob and Patricia Cohen (1975), *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Crane, Lauren E. (1964), "How Product, Appeal, and Program Affect Attitudes toward Commercials," *Journal of Advertising Research,* (1), 15-18.

Crites, Stephen L., Leandre R. Fabrigar and Richard E. Petty (1994), "Measuring the Affective and Cognitive Properties of Attitudes: Conceptual and Methodological Issues," *Personality and Social Psychology Bulletin,* (6), 619-634.

Cronbach, Lee J. and Paul E. Meehl (1955), "Construct Validity in Psychological Tests," *Psychological Bulletin,* (4), 281-302.

Danaher, Peter J. and Terence W. Beed (1993), "A Coincidental Survey of People Meter Panelists: Comparing What People Say With What They Do," *Journal of Advertising Research,* (1), 86-92.

DeVellis, Robert F. (1991), *Scale Development*. Newbury Park, London: SAGE Publications.

Dickson, Peter R. (1991), "GRP: A Case of a Mistaken Identity," *Journal of Advertising Research,* 55-59.

du Plessis, Erik (1994), "Recognition versus Recall," *Journal of Advertising Research,* (3), 75-91.

DuBois, Cornelius (1974), "Copy Testing," in *Handbook of Marketing Research,* R. Ferber, ed. McGraw-Hill, Inc.,.4/132-134/150.

Duckworth, Gary, ed. (1997), *Advertising Works 9. IPA Advertising Effectiveness Awards 1996.* Henley-on-Thames, UK: NTC Publications Limited.

Eagly, Alice H. and Shelly Chaiken (1993), *The Psychology of Attitudes*. Fort Worth, TX: Harcourt Brace Jovanovich, Inc.

Eastlack, Joseph O., Jr. and Ambar G. Rao (1989), "Advertising Experiments at the Campbell Soup Company," *Marketing Science,* (1), 57-71.

Engel, James F., Roger D. Blackwell and Paul W. Miniard (1995), *Consumer Behavior.* Forth Worth, TX: The Dryden Press.

Ephron, Erwin (1995), "More Weeks, Less Weight: The Shelf-space Model of Advertising," *Journal of Advertising Research,* (3), 18-23.

Fenwick, Ian and Marshall D. Rice (1991), "Reliability of Continuous Measurement Copy-Testing Methods," *Journal of Advertising Research,* (1), 23-29.

Finn, Adam and Ujwal Kayandé (1997), "Reliability Assessment and Optimization of Marketing Measurement," *Journal of Marketing Research,* (2), 262-275.

Fishbein, Martin and Susan Middlestadt (1995), "Noncognitive Effects on Attitude Formation and Change: Fact or Artifact?," *Journal of Consumer Psychology,* (2), 181-202.

Fishbein, Martin and Susan E. Middlestadt (1997), "A Striking Lack of Evidence for Nonbelief-Based Attitude Formation and Change: A Response to Five Commentaries," *Journal of Consumer Psychology,* (1), 107-115.

Flandin, M.P., E. Martin and L.P. Simkin (1992), "Advertising Effectiveness Research: A Survey of Agencies, Clients and Conflicts," *International Journal of Advertising,* (3), 203-214.

Gardner, Meryl Paula (1985), "Does Attitude Toward the Ad Affect Brand Attitude Under a Brand Evaluation set?," *Journal of Marketing Research,* (2), 192-198.

Greyser, Stephen A. (1973), "Irritation in Advertising," *Journal of Advertising Research,* (1), 3-10.

Grover, Ronald (1999), *What Game? Look at the Ad.* Business Week. European Edition, 47.

Grønhaug, Kjell, Olav Kvitastein and Sigmund Grønmo (1991), "Factors Moderating Advertising Effectiveness as Reflected in 333 Tested Advertisements," *Journal of Advertising Research,* (5), 42-50.

Gunnarsson, Jonas (1999), "Cherchez la Femme? Differential Validity in Models of Risky Household Financial Behavior," *Journal of Consumer Policy,* (3), 281-312.

Hair, Joseph F. , Rolph E. Anderson, Ronald L. Tatham and William C. Black (1998), *Multivariate Data Analysis*. Upper Saddle River, New Jersey: Prentice Hall.

Haley, Russel I. and Allan L. Baldinger (1991), "The ARF Copy Research Validity Project," *Journal of Advertising Research,* (2), 11-32.

Hansen, Flemming (1998), "Advertising Research: Testing Communication Effects," in *ESOMAR Handbook of Market and Opinion Research,* C. McDonald and P. Vangelder, eds. Amsterdam: European Society for Opinion and Marketing Research,.653-724.

Haugtvedt, Curtis P. and Joseph R. Priester (1997), "Conceptual and Methodological Issues in Advertising Effectiveness: An Attitude Strength Perspective," in *Measuring Advertising Effectiveness,* W. D. Wells, ed. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.,.79-93.

Hollis, Nigel S. (1995), "Like It or Not, Liking Is Not Enough," *Journal of Advertising Research,* (5), 7-16.

Howard, J.A. (1989), *Consumer Behavior in Marketing Strategy*. Englewood Cliffs, NJ: Prentice Hall.

Howell, David C. (1997), *Statistical Methods for Psychology*. Belmont, CA: Duxbury Press.

Info (1999), *Byrånumret 1999*. Info.

Jamieson, Linda F. and Frank M. Bass (1989), "Adjusting Stated Intention Measures to Predict Trial Purchase of New Products: Comparison of Models and Methods," *Journal of Marketing Research,* (3), 336-345.

Johnson, Roger M. (1974), "Measuring Advertising Effectiveness," in *Handbook of Marketing Research,* R. Ferber, ed. McGraw-Hill, Inc.,.4/151-154/164.

Jones, John Philip (1995), *When Ads Work. New Proof that Advertising Triggers Sales*. New York: Lexington Books.

Jones, John Philip (1997), "What Does Effective Frequency Mean in 1997," *Journal of Advertising Research,* (4), 14-20.

Juster, F. Thomas (1966), "Consumer Buying Intentions and Purchase Probability: An Experiment in Survey Design," *Journal of the American Statistical Association,* (315), 658-696.

Kalwani, Manohar U. and Alvin J. Silk (1982), "On the Reliability and Predictive Validity of Purchase Intention Measures," *Marketing Science,* (3), 243-286.

Keller, Kevin Lane (1993), "Conceptualizing, Measuring, and Managing Customer-Based Brand Equity," *Journal of Marketing,* (1), 1-22.

Kokkinaki, Flora and Peter Lunt (1999), "The Effect of Advertising Message Involvement on Brand Attitude Accessibility," *Journal of Economic Psychology,* (1), 41-51.

Kotler, Philip (1997), *Marketing Management. Analysis, Planning, Implementation, and Control.* London: Prentice-Hall International Limited.

Krugman, Herbert E. (1972), "Why Three Exposures May Be Enough," *Journal of Advertising Research,* (6), 11-14.

Kumar, V., David A. Aaker and George S. Day (1999), *Essentials of Marketing Research*. New York: John Wiley & Sons, Inc.

LaBarbera, Priscilla A. and Joel D. Tucciarone (1995), "GSR Reconsidered: A Behavior-Based Approach to Evaluating and Improving the Sales Potency of Advertising," *Journal of Advertising Research,* (5), 33-53.

Laroche, Michel, Chankon Kim and Lianxi Zhou (1996), "Brand Familiarity and Confidence as Determinants of Purchase Intention: An Empirical Test in a Multiple Brand Context," *Journal of Business Research,* (2), 115-120.

Laskey, Henry A., Richard J. Fox and Melvin R. Crask (1995), "The Relationship Between Advertising Message Strategy and Television Commercial Effectiveness," *Journal of Advertising Research,* (2), 31-39.

Lavidge, Robert J. and Gary A. Steiner (1961), "A Model for Predictive Measurements of Advertising Effectiveness," *Journal of Marketing,* (6), 59-62.

Leavitt, Clark (1970), "A Multidimensional Set of Rating Scales for Television Commercials," *Journal of Applied Psychology,* (5), 427-429.

Leavitt, Clark (1975), "Theory as a Bridge Between Description and Evaluation of Persuasion," in *Advances in Consumer Research,* M. J. Schlinger, ed. Ann Arbor: Association for Consumer Research,.607-613.

Lebenson, Peggy and Max Blackston (1997), "Pretesting Advertising Using Cognitive Response Analysis: Some Conclusions and Hypotheses Based on Practical Experience," in *Measuring Advertising Effectiveness,* W. D. Wells, ed. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.,.107-127.

Lilien, Gary L., Philip Kotler and K. Sridhar Moorthy (1992), *Marketing Models*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Lipstein, Benjamin (1984), "An Historical Retrospective of Copy Research," *Journal of Advertising Research,* (6), 11-14.

Lodish, Leonard M., Magid Abraham, Stuart Kalmenson, Jeanne Livelsberger, Beth Lubetkin, Bruce Richardson and Mary Ellen Stevens (1995), "How T.V. Advertising Works: A Meta-Analysis of 389 Real World Split Cable T.V. Advertising Experiments," *Journal of Marketing Research,* 125-139.

Lord, Kenneth R., Myung-Soo Lee and Paul L. Sauer (1995), "The Combined Influence Hypothesis: Central and Peripheral Antecedents of Attitude toward the Ad," *Journal of Advertising,* (1), 73-85.

Lucas, Darrell Blaine and Steuart Henderson Britt (1963), *Measuring Advertising Effectiveness*. New York: McGraw-Hill Book Company, Inc.

LVMH (1998), *Annual Report.*

MacKenzie, Scott B. and Richard J. Lutz (1989), "An Empirical Examination of the Structural Antecedents of Attitude Toward the Ad in an Advertising Pretesting Context," *Journal of Marketing,* (2), 48-65.

MacKenzie, Scott B., Richard J. Lutz and George E. Belch (1986), "The Role of Attitude Toward the Ad as a Mediator of Advertising Effectiveness: A Test of Competing Explanations," *Journal of Marketing Research,* (2), 130-143.

Madden, Thomas J., Chris T. Allen and Jacquelyn L. Twible (1988), "Attitude Toward the Ad: An Assessment of Diverse Measurement Indices Under Different Processing "Sets"," *Journal of Marketing Research,* (3), 242-252.

Malhotra, Naresh K. (1996), *Marketing Research. An Applied Orientation.* Upper Saddle River, New Jersey: Prentice Hall Inc.

Mallet, Daniel T. (1993), "The Relationship of Screen-in Rates and Readership Levels in MRI and SMRB," *Journal of Advertising Research,* (1), 18-22.

McKenna, William J. (1988), "The Future of Electronic Measurement Technology in U.S. Media Research," *Journal of Advertising Research,* RC/3-RC/7.

McQuarrie, Edward F. and J. Michael Munson (1991), "A Revised Product Involvement Inventory: Improved Usability and Validity," in *Advances in Consumer Research,* J. F. Sherry and B. Sternthal, eds. Provo, UT: The Association for Consumer Research,.108-115.

Mehrotra, Sunil, Stuart Van Auken and Subhash C. Lonial (1981), "Adjective Profiles in Television Copy Testing," *Journal of Advertising Research,* (4), 21-25.

Mehta, Abhilasha (1994), "How Advertising Response Modeling (ARM) Can Increase Ad Effectiveness," *Journal of Advertising Research,* (3), 62-74.

Mittal, Banwari (1990), "The Relative Roles of Brand Beliefs and Attitude Toward the Ad as Mediators of Brand Attitude: A Second Look," *Journal of Marketing Research,* (2), 209-219.

Morwitz, Vicki G. (1997), "Why Consumers Don't Always Accurately Predict Their Own Future Behavior," *Marketing Letters,* (1), 57-70.

Morwitz, Vicki G., Eric Johnson and David Schmittlein (1993), "Does Measuring Intent Change Behavior?," *Journal of Consumer Research,* (1), 46-61.

Morwitz, Vicki G. and David Schmittlein (1992), "Using Segmentation to Improve Sales Forecasts Based on Purchase Intent: Which "Intenders" Actually Buy?," *Journal of Marketing Research,* (4), 391-405.

Murry, John P. and Peter A. Dacin (1996), "Cognitive Moderators of Negative-Emotion Effects: Implications for Understanding Media Context," *Journal of Consumer Research,* (4), 439-447.

Naccarato, John L. and Kimberley A. Neuendorf (1998), "Content Analysis as a Predictive Methodology: Recall, Readership, and Evaluations of Business-to-Business Print Advertising," *Journal of Advertising Research,* (3), 19-33.

Nelson, Michelle R., Sharon Shavitt, Angela Schennum and Jacqueline Barkmeier (1997), "Prediction of Long-Term Advertising Effectiveness: New Cognitive Response Approaches," in *Measuring Advertising Effectiveness,* W. D. Wells, ed. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.,.133-155.

Nunnally, Jum C. and Ira H. Bernstein (1994), *Psychometric Theory*. New York: McGraw-Hill, Inc.

O'Mahoney, Sheila and Tony Meenaghan (1997), "The Impact of Celebrity Endorsements on Consumers," *Irish Marketing Review,* (2), 15-24.

Patzer, Gordon L. (1991), "Multiple Dimensions of Performance for 30-Second and 15-Second Commercials," *Journal of Advertising Research,* (31), 18-25.

Pedrick, James H. and Fred S. Zufryden (1991), "Evaluating the Impact of Advertising Media Plans: A Model of Consumer Purchase Dynamics Using Single-Source Data," *Marketing Science,* (2), 111-130.

Percy, Larry and John R. Rossiter (1997), "A Theory-Based Approach to Pretesting Advertising," in *Measuring Advertising Effectiveness,* W. D. Wells, ed. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.,.267-281.

Peter, Paul J. (1979), "Reliability: A Review of Psychometric Basics and Recent Marketing Practices," *Journal of Marketing Research,* (1), 6-17.

Peter, Paul J. (1981), "Construct Validity: A Review of Basic Issues and Marketing Practices," *Journal of Marketing Research,* (2), 133-145.

Petty, Richard E. and John T. Cacioppo (1984), "The Effects of Involvement on Responses to Argument Quantity and Quality: Central and Peripheral Routes to Persuasion," *Journal of Personality and Social Psychology,* (1), 69-81.

Petty, Richard E., John T. Cacioppo and David Schumann (1983), "Central and Peripheral Routes to Advertising Effectiveness: The Moderating Role of Involvement," *Journal of Consumer Research,* (2), 135-146.

Pieters, Rik G.M. and Tammo H.A. Bijmolt (1997), "Consumer Memory for Television Advertising: A Field Study of Duration, Serial Position, and Competition Effects," *Journal of Consumer Research,* (4), 362-372.

Pieters, Rik G.M. and Marianne de Klerk-Warmerdam (1996), "Ad-Evoked Feelings: Structure and Impact on A_{ad} and Recall," *Journal of Business Research,* (2), 105-114.

Plummer, Joseph T. (1972), "Evaluating TV Commercial Test," *Journal of Advertising Research,* (5), 21-27.

Popper, Karl R. (1989), *Conjectures and Refutations. The Growth of Scientific Knowledge.* London: Routledge.

Procter & Gamble (1998), *Annual Report 1998.*

Reynolds, Thomas J., Jerry C. Olson and John P. Rochon (1997), "A Strategic Approach to Measuring Advertising Effectiveness," in *Measuring Advertising Effectiveness,* W. D. Wells, ed. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.,.337-355.

Rosenberg, Karl E. and Margaret H. Blair (1994), "*Observations:* The Long and Short of Persuasive Advertising," *Journal of Advertising Research,* (4), 63-69.

Rossiter, John R. (1981), "Predicting Starch Scores," *Journal of Advertising Research,* (5), 63-68.

Rossiter, John R. (1999), "Construct Measurement in Marketing," Working Paper, Department of Marketing, University of Wollongong.

Rossiter, John R. and Peter J. Danaher (1998), *Advanced Media Planning.* Boston: Kluwer Academic Publishers.

Rossiter, John R. and Geoff Eagleson (1994), "Conclusion from the ARF's Copy Research Validity Project," *Journal of Advertising Research,* (3), 19-32.

Rossiter, John R. and Larry Percy (1987), *Advertising and Promotion Management.* Singapore: McGraw-Hill, Inc.

Rossiter, John R. and Larry Percy (1997), *Advertising Communications & Promotion Management.* International Edition: McGraw-Hill.

Rossiter, J.R., L. Percy and R.J. Donovan (1991), "A Better Advertising Planning Grid," *Journal of Advertising Research,* (5), 11-21.

Rubens, William (1989), "We Don't Care About Research Quality Anymore," *Journal of Advertising Research,* (1), RC/3-RC/6.

Rust, Roland T. and Bruce Cooil (1994), "Reliability Measures for Qualitative Data: Theory and Implications," *Journal of Marketing Research,* (1), 1-14.

Schlinger, Mary Jane (1979), "A Profile of Responses to Commercials," *Journal of Advertising Research,* (2), 37-46.

Schlinger, Mary Jane and Leila Green (1980), "Art-Work Storyboards versus Finished Commercials," *Journal of Advertising Research,* (6), 19-23.

Scholten, Marc (1996), "Lost and Found: The Information-Processing Model of Advertising Effectiveness," *Journal of Business Research,* (2), 97-104.

Schroeder, Gary, Bruce C. Richardson and Avu Sankaralingam (1997), "Validating STAS Using BehaviorScan(R)," *Journal of Advertising Research,* (4), 33-43.

Sharma, Subhash (1996), *Applied Multivariate Techniques.* New York: John Wiley & Sons, Inc.

Shavitt, Sharon and Timothy C. Brock (1990), "Delayed Recall of Copytest Responses: The Temporal Stability of Listed Thoughts," *Journal of Advertising,* (4), 6-17.

Singh, Surendra N. and Catherine A. Cole (1993), "The Effects of Length, Content, and Repetition on Television Commercial Effectiveness," *Journal of Marketing Research,* 91-104.

Smith, Robert E. and William R. Swinyard (1982), "Information Response Models: An Integrated Approach," *Journal of Marketing,* (1), 81-93.

Stapel, Jan (1998), "Recall and Recognition: A Very Close Relationship," *Journal of Advertising Research,* (4), 41-45.

Stephens, Debra L. and J. Edward Russo (1997), "Extensions of the Cognitive Response Approach to Predicting Postadvertisement Attitudes," in *Measuring Advertising Effectiveness,* W. D. Wells, ed. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.,.157-178.

Stewart, David W. (1981), "The Application and Misapplication of Factor Analysis in Marketing Research," *Journal of Marketing Research,* (1), 51-62.

Stewart, David W. (1986), "The Moderating Role of Recall, Comprehension, and Brand Differentiation on the Persuasiveness of Television Advertising," *Journal of Advertising Research,* (2), 43-47.

Stewart, David W. and David H. Furse (1984), "Analysis of the Impact of Executional Factors on Advertising Performance," *Journal of Advertising Research,* (6), 23-26.

Stewart, David W. and David H. Furse (1986), *Effective Television Advertising. A Study of 1000 Commercials*. Lexington, MA: Lexington Books.

Stewart, David W. and Scott Koslow (1989), "Executional Factors and Advertising Effectiveness: A Replication," *Journal of Advertising,* (3), 21-32.

Strong, Edward K. , Jr. (1912), "The Effect of Length of Series Upon Recognition Memory," *Psychological Review,* (1), 447-462.

Sutherland, Max (1993), *Advertising and the Mind of the Consumer. What Works, What Doesn't and Why?* Australia: Allen & Unwin.

Sveriges Reklamförbund, ed. (1998), *Advertising effectiveness awards 98.* Stockholm: Arvinius Förlag AB.

Swieringa, Robert J. and Karl E. Weick (1982), "An Assesment of Laboratory Experiments in Accounting," *Journal of Accounting Research,* 56-101.

Tavassoli, Nader T., Clifford J. Shultz and Gavan J. Fitzsimons (1995), "Program Involvement: Are Moderate Levels Best for Ad Memory and Attitude Toward the Ad?," *Journal of Advertising Research,* (5), 61-72.

Tellis, Gerard J. (1998), *Advertising and Sales Promotion Strategy.* Reading, MA: Addison Wesley Longman, Inc.

Tripp, Carolyn, Thomas D. Jensen and Les Carlson (1994), "The Effects of Multiple Product Endorsement by Celebrities on Consumers' Attitudes and Intentions," *Journal of Consumer Research,* (4), 535-547.

Unilever (1998), *Annual Accounts 1998 (English Version in Pounds Sterling).*

Unnava, Rao H., Sanjeev Agarwal and Curtis P. Haugtvedt (1996), "Interactive Effects of Presentation Modality and Message-Generated Imagery on Recall of Advertising Information," *Journal of Consumer Research,* (1), 81-88.

Vakratsas, Demetrios and Tim Ambler (1999), "How Advertising Works: What Do We Really Know?," *Journal of Marketing,* (1), 26-43.

Walker, David and Tony M. Dubitsky (1994), "Why Liking Matters," *Journal of Advertising Research,* (3), 9-18.

Wanous, John P. and Arnon E. Reichers (1996), "Estimating the Reliability of a Single-Item Measure," *Psychological Reports,* (2), 631-634.

Wanous, John P., Arnon E. Reichers and Michael J. Hudy (1997), "Overall Job Satisfaction: How Good Are Single-Item Measures?," *Journal of Applied Psychology,* (2), 247-252.

Wansink, Brian and Michael L. Ray (1992), "Estimating an Advertisement's Impact on One's Consumption of a Brand," *Journal of Advertising Research,* (3), 9-16.

Wansink, Brian, Michael L. Ray and Rajeev Batra (1994), "Increasing Cognitive Response Sensitivity," *Journal of Advertising,* (2), 65-75.

Wells, William D. (1964), "EQ, Son of EQ, and the Reaction Profile," *Journal of Marketing,* (4), 45-52.

Wells, William D. (1993), "Discovery-oriented Consumer Research," *Journal of Consumer Research,* (4), 489-504.

Wells, William D., Clark Leavitt and Maureen McConville (1971), "A Reaction Profile For TV Commercials," *Journal of Advertising Research,* (6), 11-17.

Wright, Peter (1980), "Message-Evoked Thoughts: Persuasion Research Using Thought Verbalizations," *Journal of Consumer Research,* (2), 151-175.

Wright, Peter L. (1973), "The Cognitive Processes Mediating Acceptance of Advertising," *Journal of Marketing Research,* (1), 53-62.

Zaltman, Gerald and Christine Moorman (1989), "The Management and Use of Advertising Research," *Journal of Advertising Research,* (December/January), 11-18.

Zielske, Hubert A. (1959), "The Remembering and Forgetting of Advertising," *Journal of Marketing,* (3), 239-243.

Zielske, Hubert A. (1982), " Does Day-After Recall Penalize "Feeling" Ads?," *Journal of Advertising Research,* (1), 19-22.

Zielske, Hubert A. and Walter A. Henry (1980), "Remembering and Forgetting Television Ads," *Journal of Advertising Research,* (2), 7-13.

Zinkhan, George M. and Scot Burton (1989), "An Examination of Three Multidimensional Profiles for Assessing Consumer Reactions to Advertisements," *Journal of Advertising,* (4), 6-14.

Zinkhan, George M. and Claes Fornell (1985), "A Test of Two Consumer Response Scales in Advertising," *Journal of Marketing Research,* (4), 447-452.

EFI
The Economic Research Institute

Reports since 1995

Published in the language indicated by the title

2000

Bornefalk, A., Essays on Social Conflict and Reform.
Karlsson Stider, A., Familjen och firman

1999

Adler, N., Managing Complex Product Development.
Allgulin, M., Supervision and Monetary Incentives.
Andersson, P., Experto Credite: Three Papers on Experienced Decision Makers.
Ekman, G., Från text till batong – Om poliser, busar och svennar.
Eliasson, A-C., Smooth Transitions in Macroeconomic Relationships.
Flink, H., Gunnarsson, J., Wahlund, R., Svenska hushållens sparande och skuldsättning– ett konsumentbeteende-perspektiv
Gunnarsson, J., Portfolio-Based Segmentation and Consumer Behavior: Empirical Evidence and Methodological Issues.
Hamrefors, S., Spontaneous Environmental Scanning.
Helgesson, C-F., Making a Natural Monopoly: The Configuration of a Techno-Economic Order in Swedish Telecommunications.
Japanese Production Management in Sunrise or Sunset. Christer Karlsson (editor).
Jönsson, B., Jönsson, L., Kobelt, G., Modelling Disease Progression And the Effect of Treatment in Secondary Progressive MS. Research Report.
Lindé, J., Essays on the Effects of Fiscal and Monetary Policy.
Ljunggren, U., Indikatorer i grundskolan i Stockholms stad före stadsdels-nämndsreformen – en kartläggning.
Ljunggren, U., En utvärdering av metoder för att mäta produktivitet och effektivitet i skolan – med tillämpning i Stockholms stads grundskolor
Lundbergh, S., Modelling Economic High-Frequency Time Series.
Mägi, A., Store Loyalty? An Empirical Study of Grocery Shopping.
Mölleryd, B.G., Entrepreneurship in Technological Systems – the Development of Mobile Telephony in Sweden.
Nilsson, K., Ledtider för ledningsinformation.
Rognes, J., Telecommuting – Organisational Impact of Home Based – Telecommuting.
Sandström, M., Evaluating the Benefits and Effectiveness of Public Policy.
Sjöstrand, S-E., Sandberg, J., Tyrstrup, M., (red) Osynlig Företagsledning.
Skalin, J., Modelling Macroeconomic Time Series with Smooth Transition Autoregressions.
Spagnolo, G., Essays on Managerial Incentives and Product-Market Competition.
Strauss, T., Governance and Structural Adjustment Programs: Effects on Investment, Growth and Income Distribution.
Svedberg Nilsson, K., Effektiva företag? En studie av hur privatiserade organisationer konstrueras.
Söderström, U., Monetary Policy under Uncertainty.
Werr, A., The Language of Change The Roles of Methods in the Work of Management Consultants.

Wijkström, F., Svenskt organisationsliv – framväxten av en ideell sektor.

1998

Andersson, M., On Testing and Forecasting in Fractionally Integrated Time Series Models.
Berg-Suurwee, U., Styrning av kultur- och fritidsförvaltning innan stadsdelsnämnds-reformen
Berg-Suurwee, U., Nyckeltal avseende kultur- och fritidsförvaltning innan stadsdels-nämndsreformen.
Bergström, F., Essays on the Political Economy of Industrial Policy.
Bild, M., Valuation of Takeovers.
Charpentier, C., Samuelson, L.A., Effekter av en sjukvårdsreform – en analys av Stockholmsmodellen.
Eriksson-Skoog, G., The Soft Budget Constraint: The Emergence, Persistence and Logic of an Institution. The Case of Tanzania 1967-1992.
Gredenhoff, M., Bootstrap Inference in Time Series Econometrics.
Ioannidis, D., I nationens tjänst? Strategisk handling i politisk miljö – en nationell teleoperatörs interorganisatoriska, strategiska utveckling.
Johansson, S., Savings Investment, and Economic Reforms in Developing Countries.
Levin, J., Essays in Company Valuation.
Ljunggren, U., Styrning av grundskolan i Stockholms stad innan stadsdelsnämnds-reformen.
Mattsson, S., Från stat till marknad – effekter på nätverksrelationer vid en bolagise-ringsreform.
Nyberg, A., Innovation in Distribution Channels – An Evolutionary Approach.
Olsson, P., Studies in Company Valuation.
Reneby, J., Pricing Corporate Securities.
Roszbach, K., Essays on Banking Credit and Interest Rates.
Runsten, M., The Association Between Accounting Information and Stock Prices. Model development and empirical tests based on Swedish Data.
Segendorff, B., Essays on Bargaining and Delegation.
Sjöberg, L., Bagozzi, R., Ingvar, D.H., Will and Economic Behavior.
Sjögren, A., Perspectives on Human Capital: Economic Growth, Occupational Choice and Intergenerational Mobility.
Studier i kostnadsintäktsanalys, red Jennergren, P.
Söderholm, J., Målstyrning av decentraliserade organisationer. Styrning mot finansiella och icke-finansiella mål.
Thorburn, K., Cash Auction Bankruptcy and Corporate Restructuring
Wijkström, F., Different Faces of Civil Society.
Zethraeus, N., Essays on Economic Evaluation in Health Care. Evaluation of Hormone Replacement Therapy and Uncertainty in Economic Evaluations.

1997

Alexius, A., Essays on Exchange Rates, Prices and Interest Rates.
Andersson, B., Essays on the Swedish Electricity Market.
Berggren, N., Essays in Constitutional Economics.
Ericsson, J., Credit Risk in Corporate Securities and Derivatives. Valuation and Optimal Capital Structure Choice.
Charpentier, C., Budgeteringens roller, aktörer och effekter. En studie av budget-processerna i en offentlig organisation.
De Geer, H., Silfverberg, G., Citizens'Trust and Authorities' Choices, A Report from The Fourth International Conference on Ethics in the Public Service, Stockholm June 15-18, 1994.
Friberg, R., Prices, Profits and Exchange Rates.

Från optionsprissättning till konkurslagstiftning. red. Bergström, C., Björk, T.
Hagerud, G.E., A New Non-Linear GARCH Model.
Haksar, A., Environmental Effects of Economywide Policies: Case Studies of Costa Rica and Sri Lanka.
He, C., Statistical Properties of Garch Processes.
Holmgren, M., Datorbaserat kontrollrum inom processindustrin; erfarenheter i ett tidsperspektiv.
Jennergren, P., Tutorial on the McKinsey Model for Valuation of Companies.
Lagerlöf, J., Essays on Political Economy, Information, and Welfare.
Lange, F., Wahlund, R., Planerade och oplanerade köp - Konsumenternas planering och köp av dagligvaror.
Löthgren, M., Essays on Efficiency and Productivity; Contributions on Bootstrap, DEA and Stochastic Frontier Models.
Nilsson, H.E., Using Conceptual Data Modelling in Management Accounting: A Case Study.
Sjöberg, L., Ramsberg, J., En analys av en samhällsekonomisk bedömning av ändrade säkerhetsföreskrifter rörande heta arbeten.
Säfvenblad, P., Price Formation in Multi-Asset Securities Markets.
Sällström, S., On the Dynamics of Price Quality.
Södergren, B., På väg mot en horisontell organisation? Erfarenheter från näringslivet av decentralisering och därefter.
Tambour, M., Essays on Performance Measurement in Health Care.
Thorén, B., Berg-Surwee, U., Områdesarbete i Östra Hökarängen - ett försök att studera effekter av decentralisering.
Zhang Gang., Chinese Rural Enterprises Between Plan and Market.
Åhlström, P., Sequences in the Profess of Adopting Lean Production.
Åkesson, G., Företagsledning i strategiskt vakuum. Om aktörer och förändrings-processer.
Åsbrink, S., Nonlinearities and Regime Shifts in Financial Time Series.

1996

Advancing your Business. People and Information Systems in Concert. red. Lundeberg, M., Sundgren, B.
Att föra verksamheten framåt. Människor och informationssystem i samverkan. red. Lundeberg, M., Sundgren, B.
Andersson, P., Concurrence, Transition and Evolution - Perspectives of Industrial Marketing Change Processes.
Andersson, P., The Emergence and Change of Pharmacia Biotech 1959-1995. The Power of the Slow Flow and the Drama of Great Events.
Asplund, M., Essays in Industrial Economics.
Delmar, F., Entrepreneurial Behavior & Business Performance.
Edlund, L., The Marriage Market: How Do You Compare?
Gunnarsson, J., Three Studies of Financial Behavior. Research Report.
Hedborg, A., Studies of Framing, Judgment and Choice.
Holgersson, C., Höök, P., Ledarutveckling för kvinnor - Uppföljning av en satsning på Volvo - Research Report.
Holmberg, C., Stores and Consumers - Two Perspectives on Food Purchasing.
Håkansson, P., Wahlund, R., Varumärken. Från teori till praktik.
Karlsson, A., The Family Business as an Heirloom. Research Report.
Linghag, S., Man är handelsstudent. Research Report.
Molin, J., Essays on Corporate Finance and Governance.
Mägi, A., The French Food Retailing Industry - A Descriptive Study.
Mölleryd, B., Så byggdes en världsindustri - Entreprenörskapets betydelse för svensk mobiltelefoni. Research Report.

Nielsen, S., Omkostningskalkulation for avancerede produktions-omgivelser - en sammenligning af stokastiske og deterministiske omkost-ningskalkulationsmodeller.
Normark, P., Danielsson, L., Larsson, A., Lundblad, P., Kooperativa nyckeltal. Research Report.
Sandin, R., Heterogeneity in Oligopoly: Theories and Tests.
Sandén-Håkansson, U., Från kampanjmål till mediemix - en studie av samarbete mellan annonsörer, reklambyråer och mediebyråer. Research Report.
Stein, J., Söderlund, M., Framgång i arbetet, strategier för att utföra arbetsuppgifterna, arbetsuppgifternas karaktär och utbildningskvalitet. En empirisk studie av civil-ekonomer. Research Report.
Strömberg, P., Thorburn, K., An Empirical Investigation of Swedish Corporations in Liquidation Bankruptcy. Research Report.
Söderlund, M., Och ge oss den nöjda kunden. En studie av kundtillfredsställelse och dess orsaker och effekter. Research Report.
Thodenius, B., Användningen av ledningsinformationssystem i Sverige: Lägesbild 1995. Research Report.
Ulfsdotter, U., Internationalisering för expansion eller hemmamarknadsförsvar? De nordiska marknaderna för fruktyoghurt 1982-1994.
Westelius, A., A Study of Patterns of Communication in Management Accounting and Control Projects.
Wijkström, F., Den svenska ideella sektorn och pengarna. Research Report.
Örtendahl, M., Health and Time - A Problem of Discounting.

1995

Becker, T., Essays on Stochastic Fiscal Policy, Public Debt and Private Consumption.
Blomberg, J., Ordning och kaos i projektsamarbete - en socialfenomenologisk upplösning av en organisationsteoretisk paradox.
Brodin, B., Lundkvist, L., Sjöstrand, S-E., Östman, L., Styrelsearbete i koncerner
Brännström, T., Bias Approximation and Reduction in Vector Autoregressive Models. Research Report.
Ekonomisk politik i omvandling. red. Jonung, L.
Gunnarsson, J., Wahlund, R., Hushållens finansiella strategier. En explorativ studie.
Höök, P., Chefsutveckling ur könsperspektiv - Mentorskap och nätverk på Vattenfall - Research Report.
Levin, J., Olsson, P., Looking Beyond the Horizon and Other Issues in Company Valuation. Research Report.
Mägi, A., Customer Statisfaction in a Store Performance Framework. Research Report.
Persson, P-G., Modeling the Impact of Sales Promotion on Store Profits.
Roman, L., Institutions in Transition. A Study of Vietnamese Banking.
Sandberg, J., Statistisk metod - dess vetenskapliga hemvist, grundläggande principer och möjligheter inom samhällsvetenskapen. Research Report.
Sandberg, J., How Do We Justify Knowledge Produced by Interpretative Approaches? Research Report.
Schuster, W., Redovisning av konvertibla skuldebrev och konvertibla vinstandelsbevis - klassificering och värdering.
Söderberg, K., Farmartjänsten - Ny kooperation inom lantbruket. Research Report.
Söderqvist, T., Benefit Estimation in the Case of Nonmarket Goods. Four Essays on Reductions of Health Risks Due to Residential Radon Radiation.
Thorén, B., Användning av information vid ekonomisk styrning - månadsrapporter och andra informationskällor.