On testing plausible threats to construct validity
Schouwstra, S.J.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
CHAPTER 7.

THE NUTRITION EFFICACY SCALE: DEVELOPMENT AND CONSTRUCT VALIDATION USING THE DEDUCTIVE DESIGN

Abstract. - An overview of articles, dealing with the validation of self-efficacy scales, demonstrated that the emphasis is placed upon predictive relevance, rather than on showing that the test-score adequately reflects self-efficacy and nothing else. The aim of the present study was to support such a test score interpretation by theoretical rationales and empirical evidence. In order to achieve that aim, the Deductive Design was used to develop and validate a nutritional self-efficacy scale. The evaluation of the rationales and empirical evidence show that the interpretation of the level-score is trustworthy. In contrast, the interpretation of the strength-score is not supported, because the strength-score seems to arise from response tendencies. The results from this validation study demonstrate the relevance of addressing the two major validity concerns, construct representation and irrelevant variance, from the outset of test development. As such, this study underlines the potential value of the Deductive Design for construct validation.

INTRODUCTION

Within health psychology a multitude of health behavior models are used and often new models or model modifications are put forward (see Conner & Norman, 1996; Glanz, Lewis, & Rimer, 1990; Shumacker et al., 1990; Spruijt-Metz, 1995). Within these models one construct seems particularly popular: self-efficacy. Of course, self-efficacy’s popularity is far from limited to the health domain. Control-related constructs, such as self-efficacy, find their place within many research domains and many perspectives, e.g. educational psychology, organizational and occupational psychology, mental health psychology and military psychology, to mention a few.

The widespread use and popularity of this concept seems to indicate that consensus about self-efficacy has been reached. However, it is consensus about the relevance of self-efficacy that has been reached. But given the fact that self-efficacy finds its place within many different models, no consensus seems to exist about the boundaries and nomological network of self-efficacy. Because the nomological network provides a basis for construct validation, one starts to
wonder about the construct-validity of the various scales measuring self-efficacy.

**Evaluation of operationalization and validation of self-efficacy scales**

Despite self-efficacy’s popularity, relatively few test-validation studies have been published (of the 2430 articles that have “self-efficacy” as a descriptor, 107 also have “test-validity” as a descriptor in PsychLIT). And, more interestingly, the number of different self-efficacy scales mentioned in psychological abstracts seems to exceed the number of validation studies. In order to get an inkling of the construct validation of self-efficacy scales, 20 publications were examined more closely (see Appendix D).

First, the publications were examined upon rationales underlying the test score interpretation, that is the operationalization. The operationalization differed to quite some extent (see Appendix D). In total 21 self-efficacy scales were developed. Although self-efficacy beliefs should be measured in terms of judgements of capability (Bandura, 1997, p. 42), 4 scales (19%) failed to do so (Davis et al., 1994; Galavotti et al., 1995; Glynn & Ruderman, 1986; Young, Tian Oei, & Crook, 1990). In the remaining 17 scales (81%) judgements of confidence in the ability to perform a behavior (14) or confidence in performing the behavior were obtained (3). In all studies, only one self-efficacy score was obtained, instead of two scores, that is one score for the level of self-efficacy and one for strength of self-efficacy.

Furthermore, Bandura (1997, p. 42) states that self-efficacy beliefs should be measured under different levels of task demands within a given domain, and under different situational circumstances. Of the 17 scales measuring self-efficacy as defined, 11 scales (52%) met that demand. In 7 scales the difficulty or attractiveness of the items was varied by mentioning impeding situations and in 4 scales by mentioning levels of performance.

Next, the publications were scrutinized for empirical evidence dealing with the two major concerns of construct validity (Messick, 1995): construct representation and absence of irrelevant variance (see Appendix E). Construct representation refers to the evidence that the test-score reflects the whole of the construct as defined. Absence of irrelevant variance refers to evidence that the test-score does not contain excess reliable variance associated with other distinct constructs or method variance such as response sets or guessing propensities (Messick, 1995, p. 742).
Chapter 7

Reliability analyses were reported in 16 studies (80%). Evidence concerning the structure of the self-efficacy scales was presented in 11 articles (52%), in the form of exploratory factor analyses (8 studies) or confirmatory factor analyses (3 studies). One study also reported the correlation with another self-efficacy scale (Clark, Abrams, Niaura, Eaton, & Rossi, 1991).

Almost all studies primarily focus at prediction as a form of validity evidence (95%). Only 1 article (Perkel, 1992) did not mention any evidence related to prediction of treatment outcomes or prediction of behavior. The influence of other psychological constructs is assessed also in 14 studies, 12 (60%) of which report the association of self-efficacy with those psychological constructs. The included psychological constructs are very diverse, e.g. personality, social support, repression, locus of control, self-esteem and perceived barriers to mention a few. Only 2 studies (1%) had a clear theoretical underpinning for the included psychological constructs (Galavotti, et al., 1995; Sonstroem, Harlow, & Josephs, 1994).

Only 1 study tested for gender differences (Glynn & Ruderman, 1986), and in another 7 studies (35%) the association of self-efficacy with demographic variables was assessed. None of the studies incorporated evidence concerning irrelevant variance directly related to the measurement method, such as irrelevant variance caused by social desirability, response tendencies or interviewer effects.

These lines of evidence demonstrate that in validation the emphasis is placed upon the relevance of self-efficacy beliefs for prediction. To summarize, evidence concerning prediction is presented in 95% of the studies, reliability evidence in 80%, evidence concerning the structure in 55%, and 50% of the studies report the association with other constructs. But only 5 articles (25%) present all four lines of evidence (concerning reliability, structure, association with other constructs, and prediction). Furthermore, in only 3 (15%) of those 5 articles the item wording was consistent with the definition of self-efficacy (Talbot, Nouwen, Gingras, Gosselin, & Audet, 1997; Sonstroem, Harlow, & Josephs, 1994; Sullivan, LaCroix, Russo, & Katon, 1998). Sonstroem, Harlow, and Josephs (1994) also measured self-efficacy beliefs under different situational circumstances and they presented a clear theoretical underpinning for the nomological network. But, like all other studies, they did not assess both the level of self-efficacy and the strength of self-efficacy, nor did they investigate irrelevant variance associated with the measurement method.
Concluding, the concern of irrelevant variance is ignored both in the empirical validation and in the operationalization. The 20 studies do support their test score interpretation by theoretical rationales. These rationales, however, only explicitly concern the construct of interest, and not the nomological network. Furthermore, the empirical evidence presented primarily focuses at the relevance for prediction, rather than on demonstrating that the whole of the construct is reflected in the test-score and that no irrelevant variance is present.

Focus of the present study

The aim of the present study is to employ a construct validation procedure, in which the emphasis is placed upon demonstrating that the whole of the construct is reflected in the test-score and that no irrelevant variance is present. Second, this demonstration will not only incorporate empirical evidence, but will also incorporate the theoretical rationales underlying the test-score interpretation. In order to realize those two aims a self-efficacy scale for healthy eating will be developed and validated using the Deductive Design for construct validation.

The Deductive Design

Validity is built into the test from the outset of test-development (Anastasi, 1986), therefore, the Deductive Design starts with specifying the rationales underlying the test score interpretation. Three kinds of rationales underlie the test-score interpretation, the first of which obviously deals with the theoretical ideas about the construct of interest. So, the theoretical notions about the construct of interest have to be formulated first (see Figure 7.1). Next, these theoretical notions have to be translated into test content. Third, models have to be presented that depict how the test-score is assumed to reflect the construct of interest. These three rationales (formulation, translation and modeling) provide a basis for collecting and analyzing empirical data. Because construct validity is an evaluative judgement of the trustworthiness of the test-score interpretation, the last step within the Deductive Design is the evaluation of the three rationales and empirical evidence.

The focus of this whole process from formulation to evaluation is excluding two global threats to construct validity. In construct validation we seek to insure and demonstrate that the test score reflects all important dimensions or
<table>
<thead>
<tr>
<th>Major concerns</th>
<th>Scientific arguments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rationales</td>
</tr>
<tr>
<td></td>
<td>Formulation</td>
</tr>
<tr>
<td>Construct representation</td>
<td>Explicit definition (Attributes of interest and structure)</td>
</tr>
<tr>
<td></td>
<td>Implicit definition (Related constructs and phenomena)</td>
</tr>
<tr>
<td>Irrelevant variance</td>
<td>Purpose and conditions</td>
</tr>
<tr>
<td></td>
<td>Skills involved in response generation</td>
</tr>
</tbody>
</table>
The Nutritional Efficacy Scale

facets of the construct. Furthermore, we should minimalize the occurrence of test score variance that is irrelevant to the construct of interest. Within the Deductive Design the aim is to exclude plausible sources of irrelevant variance, that can be anticipated or derived from the particular task or measurement mode. Those two issues, construct representation and irrelevant variance, are addressed from the outset of formulating the theoretical notions underlying the test score interpretation.

RATIONALES

FORMULATION

The first rationale, underlying the test-score interpretation, consists of the theoretical ideas concerning the construct of interest. The formulation of those theoretical ideas should provide an answer to four questions:

1) To what group of attributes does the construct refer and how does the construct exert its influence (the explicit construct definition)?
2) Which phenomena does the construct effect, in interplay with what other constructs (the implicit construct definition)?
3) Why is the construct defined and for whom, where, and when (the purpose and the conditions)?
4) What are the skills and abilities, that come into play in response generation, given the selected task?

Purpose and conditions

The purpose of this study is to develop a self-efficacy scale for healthy eating behavior. Although many self-efficacy scales exist in this domain, these scales usually focus on one particular behavior, such as overeating. The aim is to develop a scale that covers several eating behaviors, which are required to achieve healthy nutrition.

The population of interest consists of young adults. The change from adolescence to adulthood is characterized by important role transitions and during this transitional period often unhealthy eating patterns are adopted, which are difficult to change later in adulthood. Bandura (1997, p. 184) strongly emphasizes the importance of self-efficacy beliefs in this period of life:
Chapter 7

“Those who enter adulthood poorly equipped with skills and plagued by nagging doubts about their capabilities find many aspects of their adult life aversive, full of hardships, and depressing”.

Explicit construct definition

Bandura (1997) defines self-efficacy as “the beliefs in one’s capabilities to organize and execute the courses of action required to produce given attainments” (p. 3). Self-efficacy beliefs can vary on three dimensions: level, strength and generality. The level of a self-efficacy belief reflects the difficulty of the action a person deems him/herself capable of organizing and executing. Strength refers to how certain a person is that he/she is capable of executing a given task. Moreover, people can differ with respect to the range of activities one thinks he/she can execute, which is the generality of self-efficacy (Bandura, 1992, p.42-43). Summarizing,

Self-efficacy is a person’s certainty and the perceived level of difficulty at which a person thinks, that he/she is capable of organizing and executing a range of actions required to produce an attainment.

This definition has to be narrowed down to fit our research question. The attainment we are focussing on is healthy nutrition. A healthy nutrition requires several eating behaviors, or a particular food-intake. On the basis of the food-intake behaviors, identified by the World Health Organization (1998) as enhancing health, we selected four food-intake behaviors, that also frequently occur in health literature: eating fruit and vegetables, eating little fat, not overeating and not eating snacks. Obviously these behaviors are not difficult to perform, but the frequency or regularity with which these behaviors are performed makes them more or less difficult (see Bandura, p. 43). The definition can thus be narrowed down to:

Healthy eating self-efficacy is a person’s certainty and the perceived frequency at which a person thinks, that he/she is capable of the execution and organization of eating fruit and vegetables, eating little fat, not overeating and not eating snacks, required to produce healthy nutrition.

Implicit construct definition

Self-efficacy finds its place in many different theories or theoretical models. In all models self-efficacy influences behavior directly and in interplay with goals or intentions. In the different theories a host of other constructs are postulated through which self-efficacy exerts its influence on behavior, such as
values, attributions, motivation, and outcome expectancies. But in the Social Cognitive Theory (Bandura, 1986), the theory about self-efficacy, these other constructs are only potential antecedents of self-efficacy. Put differently, these constructs have an impact on behavior only through their influence on self-efficacy beliefs and not vice versa. In formulating the implicit definition we will adhere to the Social Cognitive Theory, also because that is the part of the nomological network that is equivalent in all theories. Furthermore, the implicit definition is formulated in terms of outcomes and not in terms of processes. This latter choice is prompted by the multitude of processes through which self-efficacy beliefs are thought to exert influence (see Bandura, 1997).

Self-efficacy beliefs are beliefs that shape the goals (or intentions) one sets for oneself. Directly and in interplay with those intentions self-efficacy influences the execution of behaviors (the performance attainments).

Skills involved in response generation

Self-efficacy will be measured by presenting statements to subjects, which require a typical response. Typical response tasks (Cronbach, 1990) are tasks with which we seek to determine a person's typical performance, i.e., what that person is likely to do, feel, or think in a given situation. Tourangeau (1984, 1987) gives a global description of the response process involved. First, respondents have to interpret the statement, which requires verbal skills. Next, relevant information has to be retrieved from memory. In case the statements refer to rare behaviors, feelings or thoughts, this phase can cause irrelevant variance associated with individual differences in memory retrieval. Because the focus of the present research is a frequently occurring behavior (eating behavior), we do expect no irrelevant variance caused by such differences in memory retrieval capacity.

Third, on the basis of the retrieved information a judgement has to be formed. In this third phase respondents have to reflect upon the relevant information, which requires self-reflective skills. Consequently, the responses might be influenced by self-deception in this phase (Metcalf, 1998; Robinson & Ryff, 1999). Finally, the judgement has to be reported using the response scale provided. During this last phase the respondent might be induced to polish his or her response conform the social norms. This last phase can thus give rise to impression management.
Chapter 7

To summarize, verbal skills are required for response generation. Furthermore, self-deception and impression management might influence the responses.

TRANSLATION

The second rationale underlying a test score is the way in which the formulated theoretical ideas are translated into test content. The entire process of translation is directed at translating the construct definition into observable variables in such a way that all potential respondents can give a response as intended. This entails that the attractiveness of items has to reflect the various aspects and levels of the construct as defined. Furthermore, the reflection or influence of the other constructs or concepts associated with the task and administration method should be minimized or excluded. Therefore the translation involves four steps.

First, in order to prevent irrelevant variance due to the skills involved in response generation, guidelines for item wording are formulated. Second, the boundaries of item wording are delineated to prevent that the test-score reflects other constructs from the implicit construct definition. Third, the explicit construct definition is translated into item content and finally, the method of administration is determined (see Figure 7.1).

Guidelines

In developing statements we seek insurance that the individual differences in verbal skills, self-deception or impression management do not influence the variation in test-scores. Verbal skills are involved in the first two phases of the response process, the interpretation of the statements and the retrieval of relevant information. To prevent influence of verbal skills, responding to the statements should require a minimal level of verbal skills and the statements should require the same level of verbal skills. When the statements do not require the same level of verbal skills, differences in item attractiveness reflect differences in the difficulty of item wording, rather than differences in self-efficacy. Therefore the aim is to develop items that are of equal length and that contain only short, easy and well-known words.

After the interpretation, information is retrieved from memory. Because the presented statements always contain words that do not refer to the construct of interest, also irrelevant information is retrieved. Not only do we want to limit
the amount of irrelevant information retrieved, but we want also to limit any variation between items in this amount of irrelevant information. For that reason, the item wording has to be kept as constant as possible, except for words referring to different aspects of the construct of interest.

Self-deception and impression management might influence the third and fourth phase of the response process, forming and reporting a judgement. Forming and reporting a judgement requires a reflective process, which is usually not an aspect of the construct of interest and which can thus confound the responses. This reflective process will be more influential when more information is retrieved. Therefore short statements have to be developed, that can be processed very rapidly and automatically.

Summarizing, in order to prevent irrelevant variance the following guidelines for translation were formulated:
1) The statements should be of equal length.
2) The statements should contain only short, easy and well-known words.
3) The wording has to be kept as constant as possible except for words referring to different aspects of the construct of interest.
4) The statements have to be short.

Boundaries of item wording
As Bandura (p. 45) points out, the item content of self-efficacy scales must represent beliefs about personal abilities to produce specified levels of performance and must not include other characteristics. The items should not make any reference to the constructs of the implicit definition, intentions and behavior. The items should, therefore, not be phrased in terms of “I will”. Also words referring to actually performing the eating behaviors should be avoided. For example, the statement “I am usually able to eat low-fat foods” should not be used, because it refers to actually performing the behavior in the past at a certain frequency rather than a self-efficacy belief.

Translation into item content and into a response scale
According to Bandura (1997), the items should be phrased in terms of “I can do”. Bandura proposed two different response formats: the dual-judgement format and the single-judgment format. In both formats the generality of the self-efficacy belief is revealed by the response patterns. In the single judgement
format respondents only indicate their certainty. In the dual format judgement (see below) a direct estimate of strength and of level is obtained.

I can perform [a given action] at a [given frequency] [Yes/No][Uncertain ... Certain]

Three problems are associated with the suggested formats. First of all, the single judgement format yields no direct estimate of self-efficacy level. Consequently, the dual judgement format is required for validation of the relevancy of the separate dimensions.

The second problem is that, in both formats, the assessment does not match the definition exactly. The formats yield an assessment of a person's certainty that he or she is or isn’t capable of executing and organizing the four eating behaviors at a range of frequencies. However, according to the explicit construct definition the formats should yield an assessment of a person’s certainty and the perceived frequency at which a person thinks, that he or she is capable of executing and organizing the four eating behaviors. These problems are overcome when the following format is used:

I can perform [a given action] [Never... Always] [Uncertain ... Certain]

The scale was therefore structured according to two facets: a behavior facet (4 eating behaviors) and a judgement facet (level and strength). The items measuring the beliefs regarding each behavior should make reference to situational conditions under which the behaviors have to be performed (Bandura, 1997, p. 39 - 42). In order to keep the wording as constant as possible the following phrasing was used for all items:

Even when [a given condition], I can perform [a given action]

The conditions can be of a personal nature or of a social nature. According to Bandura, the personal conditions that can impede performance consist of negative evaluative cognitions, negative affect and a high arousal. Furthermore, the social environment can exert pressure to behave in a particular way (Bandura, 1997). On the basis of conditions mentioned in existing scales 10 different conditions were selected (see Table 7.1). The item wording and the response labels were formulated in such a way the item scores rise with increasing self-efficacy, yielding dominance data. Level is measured with a response scale varying from never (0) to daily (7) and strength is measured with a response-scale varying from completely uncertain (0) to completely certain (10). The entire scale (see Table 7.1) consisted of 4 behaviors x 10 conditions = 40 statements, each presented with two response-scales.
### Table 7.1: Items of the Nutrition Efficacy Scale

<table>
<thead>
<tr>
<th>Even when</th>
<th>I feel unfit</th>
<th>I feel tense</th>
<th>I feel down (or sad)</th>
<th>I feel uncertain</th>
<th>I am angry (or frustrated)</th>
<th>I am nervous</th>
<th>I am very busy</th>
<th>I think it's boring</th>
<th>I think it's unpleasant</th>
<th>Others do the opposite</th>
</tr>
</thead>
<tbody>
<tr>
<td>I can make sure that I</td>
<td>don't overeat</td>
<td>eat few fats</td>
<td>eat fruit and vegetables</td>
<td>don't eat snacks</td>
<td>Times a week</td>
<td>0 1 2 3 4 5 6 7</td>
<td>Certainty</td>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Administration**

A written questionnaire is used as measurement method. The items are preceded by a short instruction telling the respondents that they will be presented statements. The respondents are instructed to indicate next to each statement how often they are able to perform the behavior under the conditions mentioned and how certain they are they can do so. The instruction is followed by an example of how to fill in the responses.

Bandura (1997, p. 47) proposed to present the items in ascending order of difficulty or in a random order, but the difficulty of the items is not known. The job of filling out the scale might be rather tedious, given that 40 statements have to be judged twice. To limit the chances of response tendencies as a consequence of fatigue, the ordering should be as easy as possible. Therefore, all items concerning one behavior are presented on the same page and thus the variance in item wording between subsequent items is minimized. First the ten items related to overeating are presented, then the 10 items related to eating little fat, next those related to eating fruit and vegetables and last the items about not eating snacks. Each page starts with restating the instruction and the behavior in question is printed in bold. For each behavior the ten items are presented in the same order (see the order of conditions in Table 7.1).
Chapter 7

MODELS

Finally, the researcher has to specify how the score reflects self-efficacy in a measurement model. Also, a structural model has to be presented in which is specified how the test-score is related to the other construct of the implicit construct definition and to the concepts required for generating a response.

Measurement model

Bandura proposes to use the (average) sum-score for the two judgments, yielding level-score and a strength-score. Self-efficacy level is defined as the sum of the frequency-judgements divided by the total number of items. Similarly, self-efficacy strength\(^1\) is defined as the sum of the certainty-judgements divided by the total number of items. The Social Cognitive Theory (Bandura, 1986) does not specify the exact nature of the relationship between self-efficacy level and self-efficacy strength. We will assume that both dimensions are linearly related.

Path model of implicit construct definition

Both the level-score and the strength-score are positively related to intentions and behavior (see Figure 7.2). Intentions are also positively related to behavior. We will assume all relationships to be linear.

---

\(^1\) This is the score Bandura proposed for the single judgement format. For the dual judgement format Bandura proposed a different strength-score: The sum of the certainty-judgements of the items that had a frequency-judgement equals 1 (was answered with “yes”) divided by the total number of items. This score is, in fact, a combined score of the certainty-judgements and the frequency judgements. This score is only equivalent to the strength-score of the single judgement task (the average sum score), when all items have a frequency response equal 1.
Associations with skills

When the potentially confounding concepts, verbal ability, self-deception and impressions management, influence the test-scores, they are assumed to have a linear relationship with self-efficacy level and self-efficacy strength. The hypothesis is that no linear relationship exists between the confounding concepts and self-efficacy level and self-efficacy strength.

EVIDENCE

METHOD

Overview

After the presentation of the three rationales, empirical evidence has to be presented. The empirical evidence should also cover the two basic issues of construct validity: construct representation and irrelevant variance.

The empirical evidence should support, first, the interpretation that the test score reflects the whole of the construct as defined explicitly and implicitly. The explicit construct representation was examined in three ways. The association with other scores reflecting eating self-efficacy was assessed, confirmation of the implemented structure was sought, and, third, the attractiveness of the level- and strength-scores was examined. The implicit construct definition was examined by testing whether the predicted relationships with other constructs and phenomena were found. Furthermore, the relevance of the level- and strength-scores for prediction was examined.

The evidence should also support the interpretation that nothing else is reflected. More specifically, evidence is required that no variance is caused by either the skills or traits involved in the response process or by the particular measurement method. In two ways concept-related irrelevant variance was examined. Several item-wording characteristics were evaluated to check whether the formulated guidelines were met. Second, we tested whether no association occurred between the self-efficacy scores and verbal ability, self-deception or impression management. As for method-related irrelevant variance, response tendencies might have occurred as the result of the administration method (see Administration, p. 89). Therefore, the responses were examined upon the occurrence of response tendencies.
Chapter 7

Measures

Explicit construct representation

Eating self-efficacy was measured using the newly developed Nutrition Efficacy Scale and using three other instruments. Sheeska, Woolcott and McKinnon (1993) developed the first additional instrument for measuring self-efficacy. The Healthy Eating self-efficacy scale was developed in order to test a new theoretical model concerning intentions to adopt healthy eating behaviors. The instrument (α = 0.87) consisted of 23 items with an 11-point rating scale, varying from not at all confident (0) to extremely confident (10). The items referred to the ability to buy, choose, eat, drink or prepare healthy food (respectively, 3, 5, 9, 3 and 3 items) and referred to two different kinds of healthy food: low-fat (11 items) and high-fiber (11 items and 1 item unspecified). The attractiveness of the items was varied by referring to the frequency of behavior and/or by referring to a particular situation (for example under social pressure, at work or eating outdoors).

The second additional instrument was the Weight Efficacy Life-Style Questionnaire developed by Clark, Abrams, Niaura, Eaton, and Rossi (1991). The items of this scale agree well with Bandura’s ideas, furthermore the explicit construct representation of the scale was well validated. This instrument consisted of 20 items that had to be answered on a 10-point confidence scale. The items referred to the ability to resist overeating when food was available (4 items, α = .76), under social pressure (4 items, α = .90), during positive activities (4 items, α = .70), when experiencing physical discomfort (4 items, α = .82) and when experiencing negative emotions (4 items, α = .87).

The third instrument was a short version of the Nutrition Efficacy Scale consisting of 8 items. The Short Nutrition Efficacy Scale was identical to the Nutrition Efficacy Scale, except no conditions (impediments to the behaviors) are mentioned in the items. For example, ‘I can make sure that I eat enough fruit and vegetables’. This instrument was administered to test the hypothesis that the mention of impeding conditions makes the items more difficult or less attractive.

Implicit construct representation

For the implicit construct representation eating intentions and eating behavior were measured. Eating intentions were measured by 4 items referring to the 4 different behaviors mentioned in the Nutrition Efficacy Scale. For example, ‘I intend to eat enough fruit and vegetables next week’. Subjects had
to indicate how often they intended to perform the behavior on an 8-point scale ranging from never (0) to every day (7).

*Eating behavior* was also measured by self-report, using 4 items referring to the same 4 behaviors. For example, ‘Past week I ate enough fruit and vegetables’. Subjects had to indicate on an 8-point scale ranging from never (0) to every day (7) how often they had performed the behavior past week.

**Method-related irrelevant variance**

The finding of an association between two test-scores that represent unrelated constructs, but that are obtained with almost identical methods, is a strong indication of the occurrence of response tendencies. Therefore, a *Control Scale* was administered. The item wording and response scales were identical to the Nutrition Efficacy Scale, except that the items did not refer to eating behavior but to falling asleep fast. The assessment of eating behavior and sleeping behavior prior to the validation study showed that falling asleep fast was uncorrelated with the four eating behaviors. Similarly response tendencies for the Short Nutrition Efficacy Scale were measured using the *Short Control Scale*.

**Concept-related irrelevant variance**

In order to check whether the formulated guidelines were met, the wording characteristics of the four self-efficacy scales were compared. Four measures were calculated for each item of the self-efficacy scales: Item length, Word length, Unique words, and Unique Word ordering. Item length was the number of words per item and Word length the average number of letters per word. To check whether the item wording was as constant as possible, for each item the number of words were calculated, that did not occur in any of the other items of the same self-efficacy scale (Unique Words). Furthermore, the number of words that did not occur at the same position from the beginning of the sentence or clause within the other items of the scale was calculated (Unique word ordering).

To test the association of the test-scores with the potentially confounding concepts, four scales were used. *Self-deception* and *impression management* were measured with two scales from the Performance Motivation Questionnaire (Vorst, 2000). Both scales consisted of 6 indicative and 6 contra indicative items. ‘I’m always very pleasant to people’ is an example of an indicative Impression Management item. An example of an indicative Self-deception item is ‘It
doesn't bother me when people do not trust me'. Subjects had to indicate on a 5-point scale how strongly each statement applied to them.

Verbal Ability was measured using two instruments. Both verbal tests consisted of 40 items and subjects were given 10 minutes to complete each test. The first instrument was the *Verbal Analogy Test* (Elshout, 1976). Each item consisted of three words and four response alternatives. The first two words described a relation. Subjects had to choose the alternative that described the same relation with the third word offered in the item. For example,

```
milk : cow = sweater : a. shoe  
b. sheep  
c. knitting  
d. meat
```

The second instrument was the *Vocabulary Test* (Elshout, 1976). Each item consisted of a word and four response alternatives. Subjects had to choose the right meaning of the word from the response alternatives. For example,

```
apocryphal a. sectarian  
b. vague  
c. spurious  
d. concealed
```

**Sample and procedure**

First year psychology students (167 men, 326 women, and 16 of unknown gender) had to fill out 78 different questionnaires as a course requirement on five separate sessions. All sessions were exactly a week apart.

The Vocabulary Test, the Short Nutrition Efficacy Scale, the Nutrition Efficacy Scale, the two Control Scales and the Intention Scale were administered the first session. A week later, during the second session, the Behavior Scale was administered. During the last session, the Verbal Analogy Test was administered. The Nutrition Efficacy Scale was filled out by 233 subjects (72 men, 154 women and 7 of unknown gender, $M_{age} = 20.88$, $SD_{age} = 4.56$).

**Analyses**

**Explicit construct representation**

*A multi-method comparison* was performed to test the hypothesis, that the Nutrition Efficacy scores had positive linear associations with other self-efficacy measures. The correlations between the scores of the four efficacy
scales were calculated and it was tested whether these correlations were significantly higher than 0.

Second, a confirmatory factor analysis with LISREL 8.30 was performed to test the hypothesis that the structure of the responses corresponds to the theoretical facet structure. Two theoretical facets were imbedded in the efficacy scale (see Translation, p 86), a behavior facet and a judgement facet. Each behavior was measured with 10 items that refer to different situations. Strictly speaking, these situations are another facet, but this facet is not an aspect of self-efficacy. To test the theoretical facet structure, for each of the four behaviors a level- and strength-score was calculated, the structuple scores. The covariance matrix of the structuple scores was used, because much less responses are required to yield stable parameter estimates when modeling the 8 structuple-scores than when modeling the 80 item-scores. Furthermore, the structuple scores will not have the strong skewness and kurtosis, which are often present in the item responses to Likert scales. Strong kurtosis and skewness violate the assumption of multivariate normality underlying many of the analysis techniques.

The confirmatory factor model is displayed in Figure 7.3. Six facet elements (factors) were modeled, two judgement facet elements (self-efficacy level and self-efficacy strength) and four behavior facet elements (Overeating, Low fat, Fruit, and Snacks). Each structuple-score loaded on one of two correlated factors, which represented the facet-elements of the judgement facet (efficacy level or efficacy strength). The behavior-facet was modeled in the error-matrix: The two structuples of the same element of the behavior-facet (Overeating, Low fat, Fruit and Snacks) were allowed to have correlated errors (see Figure 7.3).

Restricted factor analysis was used (Oort, 1992) to assess whether one of the structuple-scores was more strongly related with the test-scores of the Weight Efficacy Scale and the Healthy Eating Efficacy Scale. Because the Weight Efficacy Scale only measured self-efficacy beliefs regarding overeating, expectations were that the “Overeating” structuple scores were more strongly related with Weight Efficacy than the structuple-scores of the other three behaviors. In contrast, the Healthy Eating Efficacy Scale did not measure self-efficacy beliefs regarding overeating, therefore we expected that the “Overeating”-scores had a weaker association with Healthy Eating Efficacy than the structuple scores of the other three behaviors. In the first model fitted (the Null model), only the level-factor and strength-factor were correlated with the test-scores of the Weight Efficacy Scale and the Healthy Eating Efficacy
Scale. A structuple-score was more strongly related with the other self-efficacy measures, when the model in which that structuple-score loaded on Weight Efficacy or Healthy Eating Efficacy fitted significantly better than the model without that factor loading. An extra factor loading between a structuple and the other self-efficacy measures will result in a significant improvement of model fit, when the modification index exceeds the Adjusted Critical Value, \( AC = (\chi^2/(C + df - 1)) \times C \). The Adjusted Critical Value (AC) is a correction of the Critical Value (C) in order to prevent chance capitalization (Oort, 1992).

Third, a repeated measures analysis was performed in SPSS to test the hypothesis that the impeding conditions lower the attractiveness of the Nutrition Efficacy Scale. The conditions that impede performance were not mentioned in the Short Nutrition Efficacy Scale, hence the statements of the Short Nutrition Scale should be more attractive and have a higher mean than the Nutrition Efficacy Scale. Also, we wanted to test whether the self-efficacy beliefs regarding the four behaviors differed in attractiveness (or difficulty). Three within-subjects factors were specified in the repeated measures analysis: scale type (Nutrition Efficacy versus Short Nutrition Efficacy), judgement-facet (Strength versus Level) and the behavior-facet (Overeating, Low fat, Fruit, and Snacks. The mean of the Nutrition Efficacy Scale was compared with the mean of the Short Nutrition Scale and the mean of Self-efficacy Strength was compare with the mean of Self-efficacy level (simple contrasts). Each mean of the behavior-facet elements was compared with the mean of the previous elements of the behavior-facet (a difference contrast, also known as a reverse Helmert contrast).

**Implicit construct representation**

Three aspects were examined: the association of level and strength with intention and behavior, the predictive relevance of the Nutrition Efficacy-scores was compared with the predictive relevance of the other efficacy scores and the predictive relevance of the structuple score was examined.

A path analysis (see Jöreskog & Sörbom, 1993) was performed using LISREL8W to test whether self-efficacy level and self-efficacy strength were related to intentions and behavior as predicted (see Figure 7.2). A (saturated) recursive regression model was fitted to the covariance structure of the test-scores of the Nutrition Efficacy Scale, the Behavior Scale and the Intention Scale. In a recursive regression model the independent variables are causal to the dependent variables, but a dependent variables can also be causal to the
other dependent variables. The level- and strength-scores were the independent variables and intention and behavior the dependent variables. The regression coefficients of the independent variables are the $\gamma$-parameters. The regression coefficients of the dependent variables (intention) upon other dependent variables (behavior) are the $\beta$-parameters. A regression coefficient, which is not significant, indicates that the variable is irrelevant for prediction. Similarly, path analyses were performed using the test-scores of the Weight Efficacy Scale and the Healthy Eating Efficacy Scale.

To test whether the Nutrition Efficacy Scale had a better predictive relevance than the other three efficacy scales a stepwise regression was performed. In the regression model all efficacy-scores were entered as independent variables and behavior and intention were entered as dependent variables. In the first model all regression coefficients ($\beta$- and $\gamma$-parameters) were fixed at zero. In each successive model the regression coefficient with the highest modification index was set free to be estimated.

To test whether each structuple score had predictive relevance a path analysis was performed using the structuple scores. A recursive regression model was fitted to the covariance structure of the structuple-scores of the Nutrition Efficacy Scale and the item-scores of the Behavior Scale and the Intention Scale. The structuple-scores were only causally related with the corresponding intention-item and behavior-item (see Figure 7.7). For example, the Strength-Overeating and Level-Overeating structuples were only causal to overeating-intentions and overeating-behavior not to the other three intentions and other three behaviors. Likewise, each intention-item was only causally related to the corresponding behavior and unrelated to the other three behaviors.

Method-related irrelevant variance

To establish whether no response tendencies had occurred, the correlations between the test-scores of the Nutrition Efficacy Scale and the test-scores of Control Scale were calculated and tested for significance. To test whether response bias occurred at structuple-level restricted factor analysis was used. In the Null Model the structuples do not differ in their functioning towards the potential violator (response tendencies). In this Null Model only the level-factor and strength-factor were allowed to correlate with the test-scores of the Control Scale. A structuple differs in its functioning, when allowing the structuple-score to load on the potential violator results in a significant
Chapter 7

TABLE 7.2  EXPECTATIONS AND ANALYSES USED

<table>
<thead>
<tr>
<th>Expectations</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explicit construct representation</td>
<td>multi-method comparison</td>
</tr>
<tr>
<td>1 The Nutrition Efficacy scores have positive linear associations with other self-efficacy measures</td>
<td></td>
</tr>
<tr>
<td>2 The structure of the responses corresponds to the theoretical facet structure</td>
<td>confirmatory factor analysis</td>
</tr>
<tr>
<td>3 The “Overeating” structuples have a weaker association with Healthy Eating Efficacy and a stronger association with Weight Efficacy than the other structuples</td>
<td>restricted factor analysis</td>
</tr>
<tr>
<td>4 The impeding conditions lower the attractiveness of the Nutrition Efficacy Scale</td>
<td>repeated measures analysis</td>
</tr>
<tr>
<td>Implicit construct representation</td>
<td></td>
</tr>
<tr>
<td>5 Self-efficacy level and self-efficacy strength are linearly related to intentions and behavior</td>
<td>path analysis with test-scores</td>
</tr>
<tr>
<td>6 The Nutrition Efficacy scores have better predictive relevance than the other efficacy scores</td>
<td>stepwise regression analysis</td>
</tr>
<tr>
<td>7 Each structuple score has predictive relevance</td>
<td>path analysis of structuple-scores</td>
</tr>
<tr>
<td>Method-related irrelevant variance</td>
<td></td>
</tr>
<tr>
<td>8 No response tendencies occur</td>
<td>bivariate test of correlations</td>
</tr>
<tr>
<td>9 No bias occurs in the structuples due to response tendencies</td>
<td>restricted factor analysis</td>
</tr>
<tr>
<td>Concept-related irrelevant variance</td>
<td></td>
</tr>
<tr>
<td>10 The guidelines for item wording were met at test level and at structuple level</td>
<td>test for homogeneous variances</td>
</tr>
<tr>
<td>11 No association occurs between the Nutrition Efficacy scores and impression management, self-deception and verbal ability scores</td>
<td>test for equality of means</td>
</tr>
<tr>
<td>12 No bias occurs in the structuples due to impression management, self-deception and verbal ability</td>
<td>bivariate test of correlations</td>
</tr>
</tbody>
</table>

TABLE 7.3  MULTI-METHOD COMPARISON

<table>
<thead>
<tr>
<th>Nutrition Level</th>
<th>Short Nutrition Level</th>
<th>Healthy Eating</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nutrition Level</td>
<td>1</td>
<td>Strength</td>
<td>0.41 **</td>
</tr>
<tr>
<td>Short Nutrition Level</td>
<td>0.60 **</td>
<td>0.25 **</td>
<td>1</td>
</tr>
<tr>
<td>Healthy Eating Level</td>
<td>0.41 **</td>
<td>0.20 *</td>
<td>0.52 **</td>
</tr>
<tr>
<td>Weight Level</td>
<td>0.57 **</td>
<td>0.34 **</td>
<td>0.51 **</td>
</tr>
</tbody>
</table>

Note. * p < .05. ** p < .01.
improvement of model fit. We tested only for strong bias (1% significance), because a simulation study of Oort (1996) demonstrated that a 5% significance level might yield false positives.

Concept-related irrelevant variance
As for the concept-related irrelevant variance, first the wording characteristics were checked. To assess whether the guidelines (p. 86) were met, the wording characteristics of the self-efficacy scales were compared. The equality of the means and the homogeneity of variances of the four wording characteristics were tested in SPSS. Similarly, we tested the differences between the structuples on those four measures.

Furthermore, we tested whether the correlations between the test-scores of the efficacy scales and the test-scores of impression management, self-deception and verbal ability scales were not significant. Restricted factor analysis was used to test for bias at structuple-level.

For convenience, all expectations and analysis techniques are listed in Table 7.2). The fit of the various LISREL-models will be evaluated using the Scaled Chi-square statistics\(^2\) (Satorra & Bentler, 1994), the NFI and the NNFI. When the model fits the covariance structure the Scaled Chi-square statistic will be non-significant. Values over .90 of the NFI and the NNFI are indicative of a fitting model\(^3\) (Bentler, 1992).

RESULTS

Explicit construct representation
The Nutrition Efficacy Scale had a high internal consistency (\(\alpha = .96\) for level and \(\alpha = .97\) for strength), like the other efficacy scales (see Appendix F). The multi-method comparison showed that all self-efficacy scales had significant correlations with one another. The level of Nutrition Self-efficacy was strongly correlated with the scores of the Healthy Eating Efficacy Scale, the Weight Efficacy Scale and level of the Short Nutrition Efficacy Scale. The strength of Nutrition Self-efficacy, on the other hand, had only mediocre correlations with these scales (see Table 7.3). It is remarkable that the level-

---

2 The Scaled Chi-square statistic is used, because the data of the Nutrition Efficacy Scale have multivariate skewness \((z = 25.39)\) and kurtosis \((z = 13.02)\).

3 Very recently, Hu and Bentler (1999) proposed more stringent cutoff criteria: \(NNFI \geq .95\). This cutoff criterion results in low Type I and Type II error rates, even when the assumptions regarding the distribution and independence are violated.
Chapter 7

**Figure 7.3** The Confirmatory Factor Model of the Nutrition Efficacy Scale

![Diagram of the Confirmatory Factor Model showing relationships between levels (Overeating, Low fat, Fruit, Snacks) and strengths (Overeating, Low fat, Fruit, Snacks) with self-efficacy level and strength.]

Note. ns = non-significant, for all other parameters $p < .01$

**Table 7.4** Fit of the Models of the Structures Scores

<table>
<thead>
<tr>
<th>Model</th>
<th>$N$</th>
<th>df</th>
<th>Scaled $\chi^2$</th>
<th>$p$</th>
<th>NFI</th>
<th>NNFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>233</td>
<td>15</td>
<td>28.88</td>
<td>0.02</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>2a</td>
<td>230</td>
<td>27</td>
<td>76.52</td>
<td>0.00</td>
<td>0.90</td>
<td>0.87</td>
</tr>
<tr>
<td>2b</td>
<td>230</td>
<td>23</td>
<td>39.84</td>
<td>0.02</td>
<td>0.95</td>
<td>0.94</td>
</tr>
<tr>
<td>3a</td>
<td>121</td>
<td>27</td>
<td>34.94</td>
<td>0.14</td>
<td>0.91</td>
<td>0.92</td>
</tr>
<tr>
<td>3b</td>
<td>121</td>
<td>20</td>
<td>20.29</td>
<td>0.44</td>
<td>0.94</td>
<td>0.97</td>
</tr>
<tr>
<td>4</td>
<td>215</td>
<td>39</td>
<td>62.58</td>
<td>0.01</td>
<td>0.92</td>
<td>0.94</td>
</tr>
</tbody>
</table>

| Difference Test | | | | | |
| Model 1 and 2a  | 12 | 47.64 | 0.00 |
| Model 1 and 2b  | 8  | 10.96 | 0.20 |
| Model 1 and 3a  | 12 | 6.06  | 0.91 |
| Model 1 and 3b  | 5  | 8.59  | 0.13 |
| Model 1 and 4   | 24 | 33.7  | 0.09 |

Note. CFA = Confirmatory factor model; RFA = Restricted Factor model

100
scores, instead of the strength-scores, correlate more strongly with Weight efficacy and Healthy Eating efficacy, because both Weight efficacy and the Health Eating efficacy are strength-scores.

Each set of 10 items measuring the same structuple showed high internal consistency (.91 ≤ α ≤ .95; see Appendix F). The hypothesized structure of the structuple-scores is confirmed in both the Nutrition Efficacy Scale and the Short Nutrition Efficacy Scale. The confirmatory factor model fitted on the covariance structure of the 8 structuples, Scaled $\chi^2 (15, N = 233) = 28.88, p = .02; NFI = 0.95; NNFI = 0.95$. The confirmatory factor model is presented in Figure 7.3. This model fitted equally well on the covariance structure of the scores of the Short Nutrition Efficacy Scale, Scaled $\chi^2 (15, N = 228) = 24.55, p = .06; NFI = 0.92; NNFI = 0.91$.

Next, the Weight Efficacy and the Healthy Eating Efficacy scores were included in this measurement model to test whether one of the structuple-scores was more strongly related with Weight Efficacy and Healthy Eating Efficacy. The Null model in which Weight Efficacy and Healthy Eating Efficacy were only correlated with the level-factor and the strength-factor did not fit (see Table 7.4, model 2a). Stepwise inclusion of factor loadings, for which the modification index exceeded the Adjusted Critical Value, yielded a fitting model (Table 7.4, model 2b). In this model “Overeating”-level and “Overeating”-strength had a positive factor loading with the Weight Efficacy-factor and a negative factor loading with the Healthy Eating Efficacy-factor (see Figure 7.4). Thus, the restricted factor analysis showed that, consistent with expectations, the “Overeating” structuples were more strongly related with Weight Efficacy and weaker with Healthy Eating Efficacy than the structuples regarding the other three behaviors.

The hypothesis regarding the attractiveness of the structuple-scores was tested with a repeated measures analysis. The analysis showed that the means of the Nutrition Scale were not lower than the means of the Short Nutrition Scale (see Table 7.5), because no significant main effect of the scale occurred. But a significant interaction between behavior and the scale did occur, indicating that for some behaviors a difference in the mean of the two scales (with and without impeding conditions) occurred. Furthermore, main effects of judgement and behavior occurred and a significant interaction between judgement and behavior. Tests of the within subjects contrasts showed that the contrast between “Snacks” and the other three behaviors was significantly
Figure 7.4 Results of the Restricted Factor Model of the Nutrition Efficacy Scale and the Other Efficacy Scores

Note. ns = non-significant, for all other parameters p < .01.
larger for the level-strength scores than for the strength-scores, \( F(1) = 47.51, p < .01 \) (see Figure 7.5). Furthermore, this contrast between “Snacks” and the other three behaviors was larger in the Nutrition Scale than in the Short Nutrition Scale, \( F(1) = 41.07, p < .01 \). As can be seen in Figure 7.5, the scale had only an effect upon the mean of “Snacks”. The means of “Overeating”, “Low fat” and “Fruit”-structuples were equal in both scales, with and without impeding conditions. Clearly, the hypothesis that the impeding conditions would make the items less attractive was not confirmed for the “Overeating”, “Low fat” and “Fruit” structuples. The inclusion of conditions into the items did have an effect upon the attractiveness of the “Snacks” structuple, but this effect was in the opposite direction. The mean was higher in the Nutrition Efficacy Scale (with conditions) than in the Short Nutrition Scale (without conditions). The inclusions of conditions made the “Snacks”-structuple more attractive, rather than less attractive.

Summarizing, the confirmation of the implemented structure and the association of the test-scores and the structuple-scores with other efficacy measures support the explicit construct representation. The empirical findings support the adequacy of the explicit construct representation of the level-scores of the Nutrition Efficacy Scale more strongly than of the strength-scores. The necessity of including impeding conditions into the items was not supported.

**Implicit construct representation**

The path analysis of the test-scores showed that self-efficacy level predicted behavior \( R^2 = .36 \) rather well and also predicted intentions \( R^2 = .11 \), see Figure 7.6). The strength of self-efficacy seemed to be irrelevant, because self-efficacy strength was not significantly associated with either intentions or behavior. The use of the Short Nutrition Efficacy Scale yielded very similar parameter estimates. Again self-efficacy level was significantly associated with intentions and behavior, but self-efficacy strength was not (see Figure 7.6).

The path model using the structuple-scores fitted well (Scaled \( \chi^2(60, N = 215) = 78.41, p = 0.06; NFI = .94; NNFI = .96 \)). The results demonstrate the relevancy of all four level-scores. Each level-score (see Figure 7.7) predicts the corresponding behavior and intention. But of the four strength-scores, only the strength-score of “Overeating” was a significant predictor of the corresponding behavior.
### Table 7.5: Repeated Measures Analysis

<table>
<thead>
<tr>
<th>Effect</th>
<th>$\Lambda_w$</th>
<th>$df_{hyp}$</th>
<th>$df_{error}$</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale</td>
<td>0.98</td>
<td>1</td>
<td>227</td>
<td>3.78</td>
</tr>
<tr>
<td>Judgement</td>
<td>0.12</td>
<td>1</td>
<td>227</td>
<td>1677.36 **</td>
</tr>
<tr>
<td>Behavior</td>
<td>0.55</td>
<td>3</td>
<td>225</td>
<td>60.97 **</td>
</tr>
<tr>
<td>Scale * Judgement</td>
<td>0.99</td>
<td>1</td>
<td>227</td>
<td>1.77</td>
</tr>
<tr>
<td>Scale * Behavior</td>
<td>0.84</td>
<td>3</td>
<td>225</td>
<td>14.11 **</td>
</tr>
<tr>
<td>Judgement * Behavior</td>
<td>0.81</td>
<td>3</td>
<td>225</td>
<td>17.09 **</td>
</tr>
<tr>
<td>Scale * Judgement * Behavior</td>
<td>0.97</td>
<td>3</td>
<td>225</td>
<td>2.25</td>
</tr>
</tbody>
</table>

*Note.* *p* < .05. **p** < .01

### Figure 7.5: Interaction Effects Between Scale and Behavior-Facet and Between Judgement-Facet and Behavior-Facet

#### Scale * Behavior interaction

![Graph showing the interaction between Scale and Behavior](image)

#### Judgement * Behavior interaction

![Graph showing the interaction between Judgement and Behavior](image)

104
The Nutritional Efficacy Scale

**Figure 7.6** Path Model of Self-efficacy Scores, Intention Score and Behavior Score

![Path Model Diagram]

Note. **p < .01. *p < .05. ns = non-significant.

Nutrition Efficacy Scale: parameters in bold typeface, Short Nutrition Scale: parameters between brackets.

**Figure 7.7** Path Model of Structure Scores and the Item Scores of Intention and Behavior

![Path Model Diagram]

Note. **p < .01. *p < .05. ns = non-significant
To test whether the Nutrition Efficacy Scale had a better predictive relevance than the other three efficacy scales, a stepwise regression was performed. The forward regression analysis using the test-scores yielded a model in which behavior (R² = .36) was only predicted by the level-scores of Nutrition Efficacy Scale and of the Short Nutrition Scale, Scaled \( \chi^2(7, N = 217) = 3.51, p = .83 \); NFI = 0.99; NNFI = 0.99. Intentions (R² = .30) were predicted by both level-scores and by the self-efficacy scores from both the Healthy Eating Efficacy Scale and the Weight Efficacy Scale. The regression model is depicted in Figure 7.8. In this model intentions were not associated with behavior. This analysis demonstrated that the Nutrition Efficacy Scale and the Short Nutrition Efficacy Scale performed better than the Weight and Healthy Eating Efficacy Scales in terms of predictive relevance for behavior.

The separate path analyses with the self-efficacy scores from the Weight Efficacy Scale and the Healthy Eating Efficacy Scale demonstrated that these scores were less strongly related to behavior (respectively, \( \gamma = .35, R^2 = .24 \), and, \( \gamma = .24, R^2 = .17 \)) than the level-scores of the Nutrition Efficacy Scale. The Weight Efficacy-scores were also less strongly related to intentions (\( \gamma = .14, R^2 = .02 \)), but the Healthy Eating Efficacy-scores (\( \gamma = .46, R^2 = .21 \)) were more strongly related to intention than the level-score of the Nutrition Efficacy Scale.

Concluding, the empirical results support the implicit construct representation of the test scores and structuple scores of the Nutrition Efficacy Scale. Furthermore, the Nutrition Efficacy test-scores perform better than the test-scores from the existing scales. The results do not support the implicit construct representation of the strength-scores of either version.

*Method-related irrelevant variance*

The strength-score of the Nutrition Efficacy Scale was significantly correlated with the strength-score of the Control Scale (see Table 7.6). Similarly, the strength-score of the Short Nutrition Scale was significantly correlated with the strength-score of the Short Control Scale. Clearly, response tendencies occurred in the strength-scores. The correlations of the strength-scores with the control scales even exceed the correlations with the other Self-efficacy scales (in Appendix G the complete MultiTrait-Multimethod Matrix is presented). Therefore, the desideratum, formulated by Campbell and Fiske (1959), that a variable correlates higher with an independent effort to measure the same trait than with measures designed to get at different traits which happen to employ the same method, was not met with the strength-scores.
Figure 7.8 Stepwise Regression Analysis

![Diagram of the Nutritional Efficacy Scale](image)

Note. All parameters \( p < .01 \).
The correlations between the efficacy scores are not depicted.

Table 7.6 Correlations Between Test-scores of the Nutrition Efficacy Scales and Plausible Sources of Irrelevant Variance

<table>
<thead>
<tr>
<th>Method-related irrelevant variance</th>
<th>Nutrition Level</th>
<th>Nutrition Strength</th>
<th>Short Nutrition Level</th>
<th>Short Nutrition Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control Scale</td>
<td>-0.08</td>
<td>-0.06</td>
<td>-0.10</td>
<td>-0.03</td>
</tr>
<tr>
<td>Strength</td>
<td>0.08</td>
<td>0.48 **</td>
<td>0.00</td>
<td>0.38 **</td>
</tr>
<tr>
<td>Short Control Scale</td>
<td>0.12</td>
<td>0.01</td>
<td>0.12</td>
<td>0.01</td>
</tr>
<tr>
<td>Strength</td>
<td>0.12</td>
<td>0.37 **</td>
<td>-0.02</td>
<td>0.22 *</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Concept-related irrelevant variance</th>
<th>Nutrition Level</th>
<th>Nutrition Strength</th>
<th>Short Nutrition Level</th>
<th>Short Nutrition Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impression Management</td>
<td>0.18 *</td>
<td>0.11</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>Self-deception</td>
<td>0.19 *</td>
<td>0.09</td>
<td>0.08</td>
<td>-0.03</td>
</tr>
<tr>
<td>Verbal Analogies</td>
<td>0.08</td>
<td>-0.04</td>
<td>0.03</td>
<td>0.00</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>0.16</td>
<td>-0.02</td>
<td>0.17</td>
<td>-0.05</td>
</tr>
</tbody>
</table>

Note. ** \( p < .01 \), * \( p < .05 \).
The restricted factor analysis showed that the Null Model in which the response tendency factor was only correlated with the level-factor and the strength-factor did not fit significantly worse (see Table 7.4, model 3a), than the measurement model without the potential violator. The strength-structuple of "Snacks" had a modification index that exceeded the Adjusted Critical Value (see Table 7.7). After removal of this structuple from the analyses (see Table 7.4, model 3b), self-efficacy strength was still strongly ($\phi = .44$) related to response tendencies, but none of the modification indices exceeded the Adjusted Critical Value. This last finding indicates that strong response tendencies occurred in all four strength-structuples, but more strongly in the strength-structuple of "Snacks". No evidence of method-related irrelevant variance in the level-scores was found.

**Concept-related irrelevant variance**

According to the guidelines the item wording should be kept as constant as possible. This guideline was well met, because the items of the Nutrition Efficacy Scale did not vary much in word order and contained very few unique words (see Table 7.9). The Nutrition Efficacy Scale had significantly less Unique Word Order and Unique Words and less variation in these wording characteristics than the Healthy Eating Efficacy and the Weight Efficacy Scale. The Nutrition Efficacy Scale had also significant less Unique Word Order and less variation in Unique Words than the Short Nutrition Scale (see Table 7.8).

A further requisite was that the items should be short, of equal length and should contain short words. The words in the Nutrition Efficacy Scale consisted on average of 4 characters, which was significantly shorter than the words in the Healthy Eating Scale. The items of the Nutrition Efficacy Scale varied less in length than the items of the Healthy Eating Scale, but the items were rather long (on average 16 words). The Short Nutrition Scale and the Weight Efficacy Scale had significantly shorter items (see Table 7.8). The wording characteristics of the four structuples hardly differed. The only difference found was that the "Overeating"- and Low fat"-structuples contained significantly shorter words than the "Fruit"-structuples, $t (18) = -4.77, p < .001$ and $t (18) = -3.19, p < .01$, and Snacks"- structuples, $t (18) = -6.37, p < .001$ and $t (18) = -4.84, p < .001$.

None of the self-efficacy scores were significantly correlated with the two measures of Verbal Ability (see Table 7.6). The complete MultiTrait-MultiMethod matrix is presented in Appendix G. The level-score of the
The Nutritional Efficacy Scale

**TABLE 7.7 MODIFICATION INDICES (MI) AND EXPECTED PARAMETER CHANGE (EPC) FOR THE STRUCTUPLE-SCORES**

<table>
<thead>
<tr>
<th>Level</th>
<th>Control MI EPC</th>
<th>Control Strength MI EPC</th>
<th>Impression MI EPC</th>
<th>Self-Management MI EPC</th>
<th>Deception MI EPC</th>
<th>Verbal Ability MI EPC</th>
<th>Vocabulary MI EPC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overeating</td>
<td>1.21 -0.05 0.80 0.06</td>
<td>0.46 -0.01 4.46 0.02 4.68 0.04</td>
<td>1.21 0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low fat</td>
<td>0.91 -0.05 0.02 0.01</td>
<td>0.22 -0.01 2.67 -0.02 0.67 -0.02</td>
<td>2.67 -0.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fruit</td>
<td>0.18 -0.02 3.84 0.16</td>
<td>2.00 0.02 1.72 -0.01 0.16 -0.01</td>
<td>1.50 0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Snacks</td>
<td>0.10 0.02 6.63 -0.21</td>
<td>0.15 0.01 0.79 0.01 0.77 -0.02</td>
<td>0.03 0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**AC (1%) = 7.10**

**TABLE 7.8 PAIRWISE COMPARISON OF THE WORDING CHARACTERISTICS OF THE FOUR EFFICACY SCALES**

<table>
<thead>
<tr>
<th>Equality of variance</th>
<th>Equality of means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Short Nutrition</td>
</tr>
<tr>
<td></td>
<td>F</td>
</tr>
<tr>
<td>Unique Word Order</td>
<td></td>
</tr>
<tr>
<td>Short Nutrition</td>
<td>3.53</td>
</tr>
<tr>
<td>Healthy Eating</td>
<td>48.11 **</td>
</tr>
<tr>
<td>Weight</td>
<td>52.79 **</td>
</tr>
<tr>
<td>Unique Words</td>
<td></td>
</tr>
<tr>
<td>Short Nutrition</td>
<td>54.43 **</td>
</tr>
<tr>
<td>Healthy Eating</td>
<td>33.72 **</td>
</tr>
<tr>
<td>Weight</td>
<td>108.24 **</td>
</tr>
<tr>
<td>Item length</td>
<td></td>
</tr>
<tr>
<td>Short Nutrition</td>
<td>0.90</td>
</tr>
<tr>
<td>Healthy Eating</td>
<td>21.67 **</td>
</tr>
<tr>
<td>Weight</td>
<td>1.80</td>
</tr>
<tr>
<td>Word length</td>
<td></td>
</tr>
<tr>
<td>Short Nutrition</td>
<td>1.01</td>
</tr>
<tr>
<td>Healthy Eating</td>
<td>1.35</td>
</tr>
<tr>
<td>Weight</td>
<td>0.24</td>
</tr>
</tbody>
</table>

Note. **p < .001. *p < .01.**

109
Chapter 7

Table 7.9  Wording Characteristics of the Four Efficacy Scales and the Structuples

<table>
<thead>
<tr>
<th>Efficacy Scale</th>
<th>Item length</th>
<th>Word length</th>
<th>Unique words</th>
<th>Unique word ordering</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N_i$</td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
</tr>
<tr>
<td>Nutrition</td>
<td>40</td>
<td>16.68</td>
<td>1.94</td>
<td>4.09</td>
</tr>
<tr>
<td>Short Nutrition</td>
<td>4</td>
<td>9.75</td>
<td>0.96</td>
<td>4.10</td>
</tr>
<tr>
<td>Healthy Eating</td>
<td>23</td>
<td>15.70</td>
<td>4.30</td>
<td>4.79</td>
</tr>
<tr>
<td>Weight</td>
<td>20</td>
<td>12.75</td>
<td>2.45</td>
<td>4.09</td>
</tr>
<tr>
<td>Structuple</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overeating</td>
<td>10</td>
<td>16.90</td>
<td>1.66</td>
<td>3.70</td>
</tr>
<tr>
<td>Low fat</td>
<td>10</td>
<td>15.90</td>
<td>1.66</td>
<td>3.88</td>
</tr>
<tr>
<td>Fruit</td>
<td>10</td>
<td>18.10</td>
<td>2.23</td>
<td>4.27</td>
</tr>
<tr>
<td>Snacks</td>
<td>10</td>
<td>15.80</td>
<td>1.40</td>
<td>4.51</td>
</tr>
</tbody>
</table>

Nutrition Efficacy Scale had a small, but significant association with Impression Management ($r = .18$) and Self-Deception ($r = .19$). The Weight Efficacy score was also significantly associated with Self-Deception ($r = .19$). To test whether the structuples were biased, the potential violators were included in the measurement model (see Table 7.4, model 4). None of the structuple-scores had a modification index that was larger than the Adjusted Critical Value (see Table 7.7). In other words, the model fit would not improve, when one of the structuples would be allowed to load on one of the violator factors. This finding means that none of the structuple-scores functioned differently with respect to Verbal Ability, Self-deception or Impression Management.

EVALUATION

The last step of the Deductive Design is evaluating whether the rationales and the empirical evidence support the test score interpretation, that the whole of the construct is represented and nothing else. The evaluation should focus at three characteristics of the rationales and of the empirical evidence. First of all, the ambit of support has to be evaluated. The rationales and evidence can support the test score interpretation completely, or only partially. Partial support occurs, for example, when both rationales and empirical evidence support the explicit construct representation, but do not exclude plausible method-related irrelevant variance.
Second, we have to evaluate how strongly these two arguments (rationales and empirical evidence) support the interpretation. When very specific or restrictive models and hypotheses are formulated on the basis of the underlying theory and these hypotheses or models are empirically confirmed, the support for the interpretation is strong. If many assumptions or relatively arbitrary choices were made, the support will be rather weak. The stronger the links between rationales and evidence, the stronger the support can be. Third, we have to evaluate whether the rationales and empirical evidence limit the extent to which the test-score interpretation can be generalized.

Each of the four issues, addressed throughout the presentation of rationales and empirical evidence, will be, first, summarized shortly and then discussed.

Summary

Explicit construct representation

The construct of self-efficacy was defined and this definition was narrowed down to fit our research question. Consistent with Bandura’s directions, this definition was translated into a general item wording. On the basis of literature review, we selected 10 conditions under which the 4 actions can occur for the items. In order to follow the definition as closely as possible, a polytomous response scale for level was used, rather than the usual dichotomous response scale. In the modeling phase, the average sum scores were used as an assessment of the level and strength of self-efficacy. The implemented theoretical structure was confirmed empirically. Furthermore, the association of the two test-scores with other measures of self-efficacy support the interpretations of the level scores, and to a lesser extent the interpretation of the strength-scores. The hypothesis that the inclusion of conditions into the items would make the items less attractive was not confirmed.

Implicit construct representation

An implicit construct definition was formulated, in which intentions and behavior defined self-efficacy. This definition was used to set boundaries for the item wording in the phase of translation. Next, the implicit construct definition was represented in a path model. In this model the assumption was made, that the constructs and behavior were linearly related. The empirical evidence supported the implicit construct representation of efficacy level, but not the implicit construct representation of strength. The strength-scores of self-efficacy seemed irrelevant for the prediction of behavior and intentions. In
terms of the implicit construct representation the level-scores of the Nutrition Efficacy Scale outperformed the other two existing self-efficacy scales. Both the rationales and the empirical evidence, therefore, strongly support the implicit construct representation of self-efficacy level. However, the empirical evidence does not support the implicit construct representation of self-efficacy strength.

**Method-related irrelevant variance**

The purpose of this study was to develop a self-efficacy scale for healthy eating behavior of young adults. A written questionnaire was used as measurement method. This method has the advantage of a high level of anonymity and the absence of interviewer effects. In order to prevent too much variance in item wording between subsequent items, all items concerning one behavior were presented after one another.

The influence of the measurement method upon the test scores was empirically tested. The self-efficacy scores were compared with scores obtained with a similar measurement instrument assessing another trait. The empirical evidence showed strong response tendencies occurred in the strength-responses, but not in the level responses. Furthermore, we cannot exclude influences of the sampling method upon the test scores, because the scale was administered only in a sample of psychology students and drop-out occurred.

**Concept-related irrelevant variance**

Self-efficacy was measured by presenting subjects statements requiring a typical response. Verbal skills were required for response generation. Furthermore, self-deception and impression management could influence the responses. In order to prevent these anticipated constructs to influence the test score, four guidelines for item wording were formulated. The item wording characteristics of the Nutrition Scale were good, especially in comparison with the Healthy Eating Efficacy Scale. The only disadvantage was that the items were rather long. The scores were not correlated with verbal skills, but the level-scores were weakly correlated with self-deception and impression management.

**Discussion**

The interpretation, that the level-scores implicitly and explicitly represent self-efficacy level was well supported by theoretical rationales and empirical evidence. The generality of the test-score interpretation is limited to the four selected actions and the particular sample. Therefore, replication studies are
required to establish whether the test-score interpretation extends to other nutritional actions and whether the interpretation is valid for the population of young adults. Furthermore, both rationales and empirical evidence support that the particular method and verbal skills do not influence the level scores.

One demand was not met, though. The items were rather long, although they should have been short in order to prevent irrelevant variance due to self-deception or impression management. Indeed, a weak association between self-efficacy level and self-deception and impression management was found. The weak association of the level-scores with self-deception and impressions management might have been caused by the long items. But not all results support this hypothesis.

Such associations did not occur with the Healthy Eating Scale, which had equally long items, on the other hand they did occur with the shorter Weight Efficacy Scale. Generally, self-deception is viewed as a response bias, but it can also be viewed as an adaptive and protective judgement heuristic (Metcalf, 1998). This view implies that the association with self-deception might be a characteristic of self-efficacy beliefs rather than a response bias. The finding that the latent self-efficacy factor was associated with self-deception, but no additional bias occurred at the level of structuple-scores gives some support to this view.

The results concerning the Short Nutrition Scale, on the other hand, do support the hypothesis that the associations with self-deception and impression management were caused by the long items. Such associations were not found with the Short Nutrition Scale, which had much shorter items, because no impeding conditions were mentioned. Indeed, the necessity of mentioning situational conditions, under which the behaviors have to be performed, requires more research. The situational conditions did not have the anticipated impact on the item attractiveness.

Another set of conditions might yield results consistent with Bandura’s theoretical formulations, because the selection of the particular conditions was not strictly theory-driven, but based upon common practice in designing self-efficacy scales. On the other hand, the mention of impeding conditions might not be necessary at all, when dealing with nutritional behaviors. Although Bandura is of the opinion that decontextualized items obscure what is being measured and create problems of predictive relevance, this opinion is not supported by our research. The level-score of the Short Nutrition Efficacy Scale, which consists of such decontextualized items, clearly had predictive
relevance and performs as well as the Nutrition Efficacy Scale in terms of the explicit construct representation. Although the explicit construct representation, the implicit construct representation, and the exclusions of irrelevant variance due to the method and verbal skills strongly support the interpretation of the level-scores, the same does not hold for the strength-scores. Some support for the interpretation of the strength-score was found. The absence of concept-related irrelevant variance was confirmed and the explicit construct representation also gives some support to the interpretation of the strength-score. The strength-score was positively related with the other self-efficacy measures and one factor underlies the strength-scores. In contrast, the empirical evidence did not support the implicit construct representation and also revealed strong method-related irrelevant variance.

During test-development several assumptions and choices were made, that might have caused this lack of validity. All relationships were assumed to be linear, although Bandura (1997, p. 43) pointed out that strength of perceived self-efficacy is not necessarily linearly related to choice behavior. Furthermore, a sum score model for self-efficacy strength was used, as most studies do, but other choices could have been made. Lee and Bobko (1994) compared the performance of different measurement models for self-efficacy. They compared the sum scores of level and strength with, what they call a composite measure. The composite measure was defined as the sum of the certainty responses for which the level-response was “yes”. The composite measure had slightly higher correlations with goals and outcome variables, than the level-score or the strength-score alone. But Lee and Bobko did not determine whether the use of single composite score yields better results than the use of both a level-score and a strength-score. Furthermore, several problems are associated with the analysis of such composite scores (Evans, 1991; Mellenbergh, Molendijk, De Haan, & Ter Horst, 1990).

Given the evidence for strong response tendencies, the most likely cause for lack of validity are the response scales. Each statement had to be judged twice, using two polytomous response scales, which is a rather difficult task. The inability of respondents to perform that dual judgement task might in fact be the latent factor underlying the strength-scores.

To conclude, the interpretation of the level-score as representing self-efficacy level is strongly supported and, hence, trustworthy. But the interpretation of the strength-score as reflecting self-efficacy strength is invalid.
CONCLUSION

A short overview of validation studies revealed several shortcomings in the validation procedures, generally employed. The rationales, underlying the test score interpretation, usually concerned the explicit construct definition only, but not the implicit construct definition. In contrast, the evidence focused at the implicit construct definition, and more specifically at the predictive relevance of the test score. The issue of irrelevant variance was ignored in both rationales and empirical evidence.

The aim of the present study was to employ a construct validation procedure, in which such shortcomings were prevented. Therefore the Deductive Design was used as a framework for validation. The Deductive Design requires that both issues of construct validity (construct representation and irrelevant variance) are addressed from the outset of test-development and that both rationales and empirical evidence can support the test score interpretation. The value of the Deductive Design validation procedure can be illustrated by discussing what would have happened if we had not employed that procedure.

As already noted, usually the concern of irrelevant variance is ignored in validation studies. Had only the construct representation been addressed in the present study, the interpretation of the strength score would have been considered quite valid. The association with the other self-efficacy scales and the confirmation of the structure support the test-score interpretation. In the light of that support, lack of support for the implicit construct representation would not have seemed too serious a threat. Especially, since the Social Cognitive Theory gives an explanation for such a finding: perceived self-efficacy is not necessarily linearly related to choice behavior. Furthermore, if the other proposed measurement model for self-efficacy strength had been used (irrespective of the methodological problems associated with this measure), no problems would have arisen with the implicit construct representation. The other measure for self-efficacy strength is the sum of the product of each strength- and level-score (see p. 90-90) and is therefore, in fact, a combined measure. As the level-score was associated with behavior and intention, the combined measure would also have been associated with behavior and intention.

The interpretation of the strength-score would also have been considered quite valid, had we used the validation procedure proposed by Campbell and
Chapter 7

Fiske (1959). They suggest a sequential validation procedure. Before testing the relationships between traits, a multitrait-multimethod comparison should give some confidence in one’s measures of a trait. In the multitrait-multimethod procedure the correlation between measures having the trait in common and measures having the method, but not the trait, in common are compared (either using observed scores or latent factor scores). Thus, first the explicit construct representation and the concept-related irrelevant variance are addressed. If these results are satisfactory, one turns to the implicit construct representation (the relationships between traits). Both the explicit construct representation and the absence of concept-related variance support validity. Given this initial support for validity, the subsequent failure to confirm the relationships with other traits would not have resulted in the conclusion that our test score interpretation is invalid. Once again, the possibility of formulating a theoretical explanation for the failure (non-linear relationships) would encourage us even more not to question the validity.

Chances are that, had we used a sequential procedure, we would never have raised the question of method-related irrelevant variance. And it is the evidence for response tendencies that throws doubt on validity and that places the other results in a different light. In the light of that evidence, the failure to confirm the implicit construct representation becomes a significant indication of invalidity. Furthermore, the high correlation between the strength-score and the score for response tendencies prompts us to consider the correlation between the strength-score and other efficacy scores as only moderate.

Our resistance to concluding that our measures are invalid and our inclination to search rather for theoretical explanations makes it important to address all four issues simultaneously and a priori. A theory-driven approach provides us with the rationales for a test score interpretation. Obviously, the trustworthiness of a test-score interpretation increases when both rationales and empirical evidence support that interpretation. Furthermore, a theory-

---

4 The multitrait comparison is less restrictive than the establishment of absence of concept-related irrelevant variance. The multitrait comparison demands only that the other traits be intended to differ. In contrast, the Deductive Design demands that the other traits form a threat to validity, given the specific task for measuring the trait of interest.

An example of a multitrait comparison that cannot be used to establish absence of concept-related irrelevant variance is a study of Embretson (1994). Embretson developed a test for spatial ability. Spatial ability is intended to differ from verbal ability. Therefore, as a multitrait comparison, the scores of the spatial ability test were compared with scores of a verbal ability test. However, the selected task for measuring spatial ability was non-verbal and, hence, verbal ability did not constitute a threat to validity.
driven approach forces us to inspect the underlying psychological theory very closely and might reveal where is room for improvement. But the a priori approach should not be limited to the construct representation. The fact, that it is relatively easy to come up with theoretical explanations for unexpected results, makes the a priori approach to irrelevant variance valuable and even necessary.

Again, the results concerning method-related irrelevant variance can serve as an example. The Social Cognitive Theory provides an (a posteriori) explanation for the correlation between the strength-scores and the scores of, what we have labeled as, response tendencies. We could have used another label for these scores, that is “sleeping self-efficacy”. According to the Social Cognitive Theory, efficacy beliefs can generalize to other behavioral domains, and consequently, we might have concluded that the strength of nutritional efficacy generalizes to sleeping self-efficacy.

However, generalization across behavioral domains can occur when the behaviors have something in common. Therefore, a behavior was selected that was not correlated with the nutritional behaviors. The a priori approach enabled us to exclude the theoretical explanation that nutritional efficacy generalized to sleeping efficacy. The simultaneous approach strengthens the conclusion that, in fact, we are dealing with response tendencies. None of the other nutritional efficacy measures were correlated with “sleeping efficacy” and the strength-score was more strongly correlated with “sleeping efficacy” than with the other nutritional efficacy measures. The a priori approach to all four concerns of construct validity enables us to determine the most likely cause of empirical results that do not match expectations.

To conclude, a construct validation procedure, as proposed in the Deductive Design, can give strong support to the trustworthiness of a test score interpretation and dictates future lines of validation research. More importantly, such a validation procedure enables us to reach the conclusion of invalidity and steers us away from the temptation to formulate alterations to the underlying psychological theory. The results of our study clearly demonstrate the potential value of the Deductive Design, in which the four issues of construct validity are addressed simultaneously from the outset of test-development.