

On the Subjective–Objective Distinction for Measures of Memory and Cognition

Theoretical and Methodological Issues in
Questionnaire Development and Validation

Peter Vestergren



Department of Applied Educational Science
Educational Measurement
Umeå University
No. 7 · 2011

Department of Applied Educational Science
Educational Measurement
Umeå University
Doctoral thesis 2011

Cover art and design by Björn Sigurdsson

Printed by Print & Media, Umeå University
Umeå, Sweden
September 2011

© Peter Vestergren

ISSN 1652-9650
ISBN 978-91-7459-271-9

"Forty-two!" yelled Loonquawl. "Is that all you've got to show for seven and a half million years' work?"

"I checked it very thoroughly," said the computer, "and that quite definitely is the answer. I think the problem, to be quite honest with you, is that you've never actually known what the question is."

"The Hitchhiker's Guide to the Galaxy" by Douglas Adams

Abstract

The aim of this thesis was to develop a questionnaire for cognitive functioning, which could possibly be used as a screening instrument for early signs of dementia in the future. The introduction discusses the often made distinction between subjective and objective measures. A background to the four articles is provided, focussing on findings of weak relationships between self-report- and laboratory measures of memory/cognition. Studies I and II provided results and conclusions that guided instrument development and validation in Studies III and IV. All studies were based on data from participants in the Betula Prospective Cohort Study. Study I investigated predictors of scores on an established self-report instrument for memory failures (PRMQ). Candidate predictors were memory performance on laboratory tests, age, depressive symptoms, and personality traits. There was no relation to age, and test performance did not predict self-reported memory, but depressive symptoms and personality did. Given the finding of a lack of a relation to age, and a bulk of research articles claiming that memory complaints are common in the elderly or increase with age, Study II used a global rating of problems with memory, and reports of perceived causes. In contrast to Study I, problems ratings were related to age, such that increasing age meant higher severity of problems. Furthermore, perceived causes of memory problems differed across age. The elderly reported aging while the young reported stress and multitasking as primary causes. With these results as a background, the purpose of Study III was to develop a new instrument (the Cognitive Dysfunction Questionnaire - CDQ) with the explicit aim that scores should be related to laboratory test performance. A global construct of cognitive functioning with an emphasis on memory systems was adopted, and an item pool was generated. Based on exploratory principal components analysis and correlations with criterion measures (laboratory test performance), twenty items in six domains were selected. Preliminary psychometric evidence showed that the CDQ was reliable, and related to age and objective measures, but not to depressive symptoms. In Study IV, twenty additional items were constructed, and the CDQ was responded to by participants in independent samples. Confirmatory factor analysis was used to test the factor structure derived from Study III, and refinement was undertaken by collapse of two domains and exclusion of items. The final factor structure was cross-validated. Competing models and measurement invariance across age and sex was tested. Psychometric properties were investigated for the final 20-item version.

Keywords: cognitive dysfunction, measurement, memory complaints, self report, subjective memory, subjective–objective

Acknowledgements

Several people have been involved in the process that has resulted in this thesis. First, I would like to thank my supervisors at the Department of Applied Educational Science, Widar Henriksson and Christina Wikström. I am especially thankful for contributions “above and beyond the call of duty” of my third supervisor, Michael Rönnlund at the Department of Psychology. Michael is also my co-author on some of the articles along with Timo Mäntylä, Lars Nyberg and Lars-Göran Nilsson. I am in special debt to Lars-Göran and Lars for letting me work within the Betula study. Lars is also the origin of the idea of developing a new questionnaire for memory/cognitive functioning. I would also like to thank the Betula staff, the participants in the Betula study, and all those who have helped me with the construction and evaluation of items for the questionnaire. Thanks also to my colleagues at the department, especially my former and present fellow doctoral students: Gunilla Näsström, Per-Erik Lyrén, Anna Sundström, Tova Stenlund, Anna Lind-Pantzare, Daniel Arnesson and Marcus Strömbäck Hjärne. Marcus also read and commented on a draft version of the thesis – some additional thanks to you!

Finally, I think it is only fair that I should thank (and curse) my Freudian Superego. I guess this would not have been possible without you, but please give us a break once in a while... (the Id sends its worst by the way).

Peter Vestergren

Degermyr, August 2011

Articles

This thesis is based on the following articles:

- I. Rönnlund, M., Vestergren, P., Mäntylä, T., & Nilsson, L.-G. (2011). Predictors of self-reported prospective and retrospective memory in a population-based sample of older adults. *The Journal of Genetic Psychology, 172*(3), 266-284. doi: 10.1080/00221325.2010.538450
- II. Vestergren, P., & Nilsson, L.-G. (2011). Perceived causes of everyday memory problems in a population-based sample aged 39–99. *Applied Cognitive Psychology, 25*, 641-646. doi: 10.1002/acp.1734
- III. Vestergren, P., Rönnlund, M., Nyberg, L., & Nilsson, L.-G. (2011). Development of the Cognitive Dysfunction Questionnaire (CDQ) in a population based sample. *Scandinavian Journal of Psychology, 52*(3), 218-228. doi: 10.1111/j.1467-9450.2010.00861.x
- IV. Vestergren, P., Rönnlund, M., Nyberg, L., & Nilsson, L.-G. (2011). *Multigroup confirmatory factor analysis of the Cognitive Dysfunction Questionnaire: Instrument refinement and measurement invariance across age and sex*. Manuscript submitted for publication.

Referencing will be to “Article” or “Study” with the enumeration above. All published articles are reprinted with permission of the respective copyright holders.

Contents

1. Introduction	1
Disposition of the thesis	3
Some philosophy of science	3
Terminology	4
2. The subjective–objective distinction	5
The objective test in test theory	6
Concluding remarks	8
3. Cognitive functions	9
Memory	9
<i>Memory systems</i>	10
Cognitive functions in healthy aging and Alzheimer’s disease	11
Concluding remarks	12
4. Subjective memory/cognition	12
Measures of subjective memory	12
What do memory questionnaires measure?	13
The relation between subjective and objective measures of memory	15
Measuring subjective memory: Three paradigms	16
<i>Memory complaints paradigm</i>	16
<i>Everyday memory paradigm</i>	17
<i>Metamemory paradigm</i>	17
5. Validity and validation	18
Validity in an assessment perspective	18
<i>The classical view of validity</i>	18
<i>The recent view of validity</i>	18
Validity in a research perspective	19
A comment on validity	19
The view of validity for the thesis	21
6. Synthesis	23
Subjective/objective methods and constructs	23
Development and validation	24
A schema for the subjective–objective distinction	24
<i>The theoretical domain</i>	25
<i>The construct</i>	26
<i>The stimulus</i>	26
<i>The response</i>	26
<i>The scoring rule and the response indicator</i>	26
<i>Scaling</i>	27
Implications for the development and validation of instruments in the subjective memory domain	27
<i>Memory complaints paradigm</i>	27

<i>Everyday memory paradigm</i>	28
<i>Metamemory paradigm</i>	29
Concluding remarks	29
7. Development of the Cognitive Dysfunction Questionnaire	30
The Betula Prospective Cohort Study	31
8. Statistical methods	32
Classical test theory	32
Exploratory and confirmatory factor analysis	33
Additional statistical methods	33
9. Summary of the studies	34
Study I	34
Study II	35
Study III	35
Study IV	36
10. Discussion	37
Main results and conclusions from the studies	38
Limitations	39
Future research	40
Concluding remarks	41
References	42
Appendix	51
Abbreviations and acronyms	51
Statistical notation	51
Errata	52
<i>Article III</i>	52

1. Introduction

The study and measurement of human cognitive functions is mainly based on methods whereby stimulus elicited behavior is observed and recorded through standardized protocols. There is control of the setting, time, duration and type of stimulus to be responded to in a specific task. The nature and possible range of responses are restricted and often predetermined. These protocols are usually called *objective tests* and constitute a gold standard for basic research and clinical assessment. A second set of methods, used less frequently, are realized with self-report instruments such as questionnaires and self-rating scales. These, commonly denoted *subjective* methods, are often also in a standardized format where the stimulus, nature and possible range of responses are restricted and predetermined. The theoretical models and statistical methods available to those that develop objective tests and self-report instruments are also to a great extent the same. Despite these striking similarities, self-report instruments do not challenge objective tests as a gold standard, neither in research, nor in clinical assessment. There appears to be some fundamental difference between the two methods. In this thesis I will discuss and try to pinpoint this difference.

All articles in the thesis are based on data from participants in the *Betula Prospective Cohort Study* (Nilsson, et al., 2004; Nilsson, et al., 2000; Nilsson, et al., 1997), an ongoing research project that started in 1988. The main aims of the project are to study how memory functions change during adulthood and old age, to identify risk factors for dementia, and to identify early preclinical signs of dementia. Measurement by means of objective memory tests is an important and central part of the methodological arsenal of the Betula study, but various self-report, or subjective, measures are also used. One such measure is the *Prospective and Retrospective Memory Questionnaire* (PRMQ; Rönnlund, Mäntylä, & Nilsson, 2008; G. Smith, Della Sala, Logie, & Maylor, 2000). Apparently, the majority of tests in the study concern memory, and so does the questionnaire. Hence, it seems natural to expect that they should be related, and show similar relations to variables of interest, particularly age. However, from the initial analyses there appeared to be no relation between the PRMQ and memory test performance, and no relation to age. A closer look at empirical research studies with other instruments revealed an equally confusing picture. These were results that raised many questions, for example:

- *What do we mean by subjective and objective measures?*
- *When should we expect a relationship between them?*

- *How do the results with scores from the PRMQ compare to results with scores from other instruments in other studies?*

...and finally

- *Is it possible to develop a questionnaire that when used results in scores that are related to memory and cognitive performance on laboratory tests?*

Trying to find answers to these questions introduces the two main aims of the thesis. The first aim is addressed in the four articles, and is certainly of outmost importance for society and individuals. It can be formulated as the aim of finding and developing methods for detecting early signs of dementia disorders, such as *Alzheimer's disease* (AD) - a neurodegenerative disorder for which cognitive dysfunction, especially impairment in episodic memory, is a prominent feature. Self-report instruments have been suggested as one possible method for detecting these early signs. The rationale is that the affected individual has a privileged status bearing on first-hand experience with subtle cognitive dysfunction early on in the disease. This could thereby make the affected individual an important source of information that is possible to obtain by means of a questionnaire. The thesis is concerned with the development of such a questionnaire - the *Cognitive Dysfunction Questionnaire* (CDQ). The main reason for developing the CDQ was that it could possibly be used as a screening instrument for early signs of dementia in the future.

The second aim of the thesis, mainly covered in the introductory part, is to address theoretical and methodological issues when developing this type of instruments. This is the general *subjective-objective* distinction that is often made in research articles in various disciplines. Fulfillment of this aim entails a review of some of the writings on the issue, and an outline of a theoretical framework in regard to development and validation of instruments viewed from the subjective-objective perspective.

Even though the topic of the thesis is wide, I will make some restrictions; self-reported information regarding factual conditions can be obtained by various types of instruments, such as questionnaires, inventories and surveys. The information may, for example, concern whether one owns a car, is employed, and so forth. The purpose with these types of instruments is usually not to measure some theoretical construct. For this thesis, we will be confined to those instruments that purport to measure some domain of behavior or attribute of an individual's psyche, or physiology via the psyche. This usually implies instruments with multiple items that are aggregated into scales. The reader should note that there are methodological issues with relevance to both purposes, and that it is not always easy to draw a definite

line between the two when one goes beyond the most common textbook examples.

Disposition of the thesis

The thesis consists of four research articles and an extensive introductory and summarizing section.

Chapter 2 takes a broader view of the subjective–objective distinction as it is used in research articles, and reviews relevant terms and definitions in the literature.

Chapter 3 briefly describes theory and empirically based knowledge of the structure and function of memory and cognition, and how they are affected by healthy aging and Alzheimer’s disease.

Chapter 4 discusses the nature of subjective measures of memory/cognition, the terms that are used, and the relationship between subjective and objective measures. Finally, three major paradigms for research on subjective memory are suggested and described.

Chapter 5 introduces and discusses the concepts of validity and validation.

Chapter 6 attempts to provide a synthesis of the previous chapters in regard to development and validation of subjective and objective measures.

Chapter 7 describes the process and rationale for the development of the CDQ in four major steps. Next is a description of the aims and design of the Betula Prospective Cohort Study in more detail.

Chapter 8 presents the major statistical methods and models used in the articles.

In Chapter 9, the four articles in the thesis are summarized.

Chapter 10 is a general discussion of the results, conclusions and limitations of the studies. Some future directions for research are also suggested.

Finally, the four articles are presented in numerical order.

Some philosophy of science

The intricate topics of this thesis make it necessary to comment on some basic philosophical issues. As an individual ruminator, I have difficulty in transgressing the arguments of global skepticism without resorting to assumptions. However, being involved in science, it then appears especially suitable to assume that there is an external world, existing independently of me, and that I can attain knowledge of. My current philosophy of science is perhaps best described as one of *scientific realism* (for an overview, see e.g., Chakravartty, 2011). With regard to unobservable, or *latent* variables, I believe that posing them require substantial evidence from observable events. Unfortunately, these latent variables or *constructs* sometimes appear to be disengaged from what *can* be observed, and in a sense have a genesis and life of their own in people’s (including scientists’) heads. I find it strange

not to assume that the constructs “exist beyond the measures proposed to describe them” (a view adopted by e.g., Nunnally & Bernstein, 1994, p. 104). As I see it, the problem with unobservable variables is that it is easy to confuse *what cannot be observed because it does not exist* with *what does exist but cannot be directly observed*. My goal with the research is to describe and understand “stuff” that exist, may be it that it is difficult to obtain arguments for or against their existence, but in my view, that is one of the aims of science.

Terminology

There are numerous terms in this thesis and within the broader fields of psychometrics, research on cognitive functions and clinical assessment. The definitions are not always clear and several terms may be used interchangeably. Hence, terms such as *test*, *instrument*, *scale*, *questionnaire* and *measure* may be used interchangeably. This is also to some extent the case for cognitive- *function*, *functioning*, *ability*, *dysfunction*, *impairment* and *performance*.

The terms *measurement* and *assessment* are often used interchangeably in other contexts. However, I strongly believe that they should be kept separate. Observation and measurement are fundamental to many scientific disciplines, including behavioral science. I am hesitant to call all activities that involve assigning numerals to observations, or for that matter all activities that involve ranking or sorting people, measurement. By this I do not mean that they are without utility or lacking in scientific rigor, but I question the label in some cases. Nevertheless, the word measurement is used frequently in the field, and will be used also in this thesis.

General measurement theory is a complicated subject that is beyond the scope of this thesis (for reviews, see e.g., Díez, 1997a, 1997b; Mari, 2003; Michell, 1999). Suffice it to say that in psychometrics, measurement involves rules for assigning symbols to objects so as to represent quantities of attributes numerically, and classification: defining whether objects fall in the same or different categories with respect to a given attribute (Nunnally & Bernstein, 1994).

Assessment commonly refers to a process that integrates test information with information from other sources (American Educational Research Association [AERA]; American Psychological Association [APA]; & National Council on Measurement in Education [NCME], 1999). However, a fundamental addition to this description of assessment is that the object for assessment is usually a *specific individual*, and that the assessment is performed with a *specific purpose* that is the basis for *decision-making*. This purpose may be diagnosis, treatment, selection, placement, and so forth. Tests and measurement are tools that *may* be used for assessment. This distinction has far reaching consequences for many issues in this thesis,

most notably regarding the concept of *validity*. I primarily use the word assessment to refer to the focus on specific individuals and a practical purpose, and not so much the integration of test information with information from other sources.

As already touched upon, there are many references to *constructs* in the thesis. Nunnally and Bernstein (1994) describe a construct as an abstract, unobservable (latent) variable. It reflects a hypothesis that a variety of behaviors will correlate with one another in studies of individual differences and/or will be similarly affected by experimental manipulations. The use of the term construct can, among other sources, be traced to MacCorquodale and Meehl (1948), who made a distinction between *intervening variables* and *hypothetical constructs*. Upon closer inspection of the article, the overarching term appears to be *concept*, of which there are two subtypes; *abstractive* and *hypothetical*. Statements about abstractive concepts (intervening variables) only *abstract* specific empirical relationships and do not contain words that are not reducible to them, that is, they are *operationally defined*. Formulation of a hypothetical concept (construct), on the other hand, contains words that cannot be completely reduced to the empirical relationships.

In the influential article on *construct validity*, Cronbach and Meehl (1955) state that “A construct is some postulated attribute of people, assumed to be reflected in test performance” (p. 283). They also state that “Construct validation is involved whenever a test is to be interpreted as a measure of some attribute or quality which is not “operationally defined”.” (p. 282). Hence, assuming that abstractive concepts (intervening variables) are still viable in science, they are not constructs, and construct validation is perhaps not implied. *Reaction time* or *speed* on a test, are possibly examples of abstractive concepts. Both are important variables in research on cognition, but there appears to be little surplus meaning in the concepts.

It should also be noted that the tradition of referring to constructs, albeit common in social science, is not as common in other disciplines. Furthermore, when the research is in the borderland of psychology, biology and medicine it is not always immediately clear whether a concept or variable should be regarded as being a construct. It is possible that at least some of the concepts are more abstractive than hypothetical.

2. The subjective–objective distinction

Many research articles in health sciences and psychology make a subjective–objective distinction. It can be in terms of a reference to subjective/objective *methods*, *assessment*, *measures*, or as an adjective modifying a concept. Such concepts could, for example, be (subjective/objective) *health* (e.g.,

Cleary, 1997; Johnston, Propper, & Shields, 2009), *sleep* (e.g., Armitage, Trivedi, Hoffmann, & Rush, 1997; Young, Rabago, Zgierska, Austin, & Laurel, 2003), *stress* (e.g., Cole, Tucker, & Friedman, 1986; Solomon, Mikulincer, & Hobfoll, 1987), and *memory* (e.g., Lincoln & Tinson, 1989; Valentijn, et al., 2005).

Upon closer inspection, the definition and delineation of what is considered subjective and objective is not straightforward. At one level the distinction appears to be purely methodological. Subjective methods here refer to procedures for obtaining information from individuals, usually by questionnaires or self-rating scales, while objective methods refer to procedures based on directly observable behavior, physiological or physical variables. Subjective and objective methods may, but do not need to be concerned with the same subject matter or theoretical framework. A typical study may refer to questionnaires or self-rating scales of, for example, depressive symptoms or stress as subjective measures, whereas physiological variables, such as blood pressure or hormone concentration may be referred to as objective measures. See, for example, Hindmarch and Parrott (1977) as an example in pharmacological research.

At a second level, a general concept or construct may be adopted, for example, *health*, and the measures may be those of subjective health and objective health. The global or fuzzy character of the general construct permits alternative views and instrumentations. Sometimes, the terms subjective and objective *indicators* are used. See, for example, Danna and Griffin (1999) for a review in regard to health and well-being.

At a third level, a specific construct or attribute within a theoretical framework is measured by alternative methods that are subjective and objective, and the results are expected to converge. The objective methods are usually considered trustworthy and valid. A lack of convergence of results is usually attributed to methodological limitations or invalidity of subjective measures. See, for example, Spector (1994) and Howard (1994) for reviews of the use of self-report questionnaires in research on organizational behavior.

The objective test in test theory

Given the frequent use of the subjective–objective distinction in research articles, we now turn to the literature on test theory and psychometrics with hope of a clearer picture.

Choppin (1997) defines objective methods of observation as “those in which any observer who follows the prescribed rules will assign the same values or categories to the events being observed as would another observer” (p. 771). McDonald (1999) states that a test is objective if the items in it does not leave any room for judgment in the scoring of the responses. From this focus on scoring, Rothstein (1989) mainly defines the subjective–objective

distinction in terms of reliability and error. Because reliability estimates indicate error of measurement, subjectivity represents error. However, he also mentions that the adjectives subjective–objective may be applied to the phenomenon being measured.

Adopting the terminology of Cattell, Kline (2000) defines objective tests as “tests whose purport is hidden from subjects, thus making deliberate distortion (but not sabotage) difficult, and whose scoring is objective”(p. 291). Hence, the procedure for scoring is a shared defining feature among the authors, but Kline and Rothstein present additional characteristics.

Kerlinger (1964) refers to tests and scales as measurement instruments, and states that tests are scales, but scales are not tests. His emphasis is on the fact that scales do not ordinarily have the meanings of competition and success and failure that tests do. Furthermore, he addresses the distinction between “measures responded to by the subject being measured” and rating scales that are “measures of individuals and their reactions, characteristics and behaviors by observers”. He further states “the contrast, then, is between the subject as he sees himself and the subject as others see him” (p. 514). It should be noted that *scales* belong to the category “measures responded to by the subject being measured”. Sometimes, in other circumstances, the term *self-rating scale* is used. My interpretation is that Kerlinger would categorize this as a scale.

Netemeyer, Bearden, and Sharma (2003) also make a distinction between scales and tests. Furthermore, they contrast constructs that are *perceptual*, *subjective* or *opinion-based* with mental/ability testing and classification measurement for clinical diagnosis. Note that the two latter categories seem refer to the methods or procedures rather than the constructs.

Other related distinctions can be found in the use of the terms *accuracy* versus *semantic* indicators of perception (Goldiamond, 1958), *optimal* versus *typical* performance (e.g., Crocker & Algina, 1986), *cognitive* versus *non-cognitive* variables (e.g., Sedlacek, 2004), *direct* versus *indirect* personality tests (e.g., Scott & Johnson, 1972), *ability* versus *personality* tests (e.g., Anastasi & Urbina, 1997).

Nunnally and Bernstein (1994) discuss the very basic distinction between *Judgment* and *Sentiment*, where the former has a correct response, while the latter has not. To exemplify, an item for the first category could be in question form, for example, “What is 1 + 2?” for which there is an answer that is correct or incorrect. The item for a sentiment may also be in question form, for example, “Do you like Swedish meatballs?”, but there is no correct answer (at least that is directly available to an external observer).

Possibly the most relevant attempt at teasing out the subjective–objective distinction when it comes to measurement can be found in a discussion paper on “subjective measures of well-being” (Veenhoven, 2004). The author makes a distinction between *substance* matter and *assessment* that can both

be of the subjective or objective type. Objective substance matter is concerned with things that exist independent of subjective awareness, for example, a *cancer tumor*. Subjective substance matter, as I understand it, does not exist independent of subjective awareness, for example, *happiness*. Objective assessment (measurement) is based on explicit criteria and performed by external observers. Subjective assessment involves self reports based on implicit criteria. This categorization results in several possible combinations of substance-assessment, ranging from objective-objective to subjective-subjective. The author admits that there are no clear dichotomies in practice, but appears to leave out the possibility of an external observer utilizing implicit criteria. This alternative is a possibility that should be noted, especially as a contrast to objective assessment. However, we will not be concerned with this type.

Annett (2002) argues that “all empirical observations, including those conventionally labeled as ‘objective’ are unavoidably subjective” (p. 966). He claims that it is rather an issue of shared meaning between observers, *intersubjectivity*, that is the key criterion of scientific probity. Building on Weber’s distinction between subjective and objective *perceptions*, he states that “The distinction is between judgements of objects and events in the external world, that is potentially public information, and judgements of internal sensations and feelings obtained by introspection which are essentially private” (p. 968).

A similar description is provided by Moore (1995) in a reference to “folk psychology” and “traditional psychology”. He introduces the domain concept, where the *subjective domain* is that of “unobservable immediate experience, personal consciousness, and the self as an agent”. The objective domain, where behavior takes place, “is the domain of the material, physical, and publicly observable” (p. 33).

Concluding remarks

This chapter has reviewed some of the writings on the subjective–objective distinction as it applies to tests, scales and measurement. The intent was to illuminate the heterogeneity in the writings, and the lack of consensus on a somewhat neglected issue. The reasons could possibly be that it is a complicated issue that has philosophical repercussions. Nevertheless, some major themes have appeared. The differences appear to lie in types of constructs, instruments, items, and responses. The two major categories apparently concern *what* is measured and *how* it is measured.

Despite the term “behavioral science” with its roots in behaviorism (the rejection of unobservable internal mental states as a subject for scientific study), many areas of current psychology and other disciplines appear to deal with the subjective domain as described above. The main method for external observers to gain access to the domain is by means of language, and

specifically so by the semantic *meaning* of language. What is measured, that is, the constructs within the subjective domain, may be those of beliefs, preferences, attitudes, and so forth. These constructs are in a sense defined by the personal experience. Having abandoned a strict behaviorist approach as many disciplines seem to have done, it appears counterintuitive to question their accuracy or validity with reference to events in the objective domain.

I will return to the subject and outline a schema for categorizing the different steps in the development and validation of instruments in standardized format.

3. Cognitive functions

Cognition is a broad term that is typically used to refer to mental processes, such as attention, memory, learning, language and reasoning. There are many aspects of cognition that can be addressed, but for this thesis we will focus on basic cognitive functions that enable people to live their everyday lives. We will be particularly concerned with those instances when cognition does not work in a functional manner, that is, *cognitive dysfunction*.

Human cognition is often described in analogue to a computer. This rests on the notion that both the computer and the brain handle information in various forms. In the context of neuropsychological assessment, Lezak (2004) presents four major classes of cognitive functions corresponding to computer *input*, *storage*, *processing* and *output* of information. Input is realized by means of *receptive functions* that involve the ability to select, acquire, classify and integrate information. *Memory and learning* are storage and retrieval functions. *Thinking* is processing through mental organization and reorganization of information. The *expressive functions* serve communication and action based upon information.

Memory

Memory is a fundamental cognitive function and can be defined as the capacity to *encode*, *store* and *retrieve* information (Tulving, 2000). Even though memory is a single term, it refers to a multitude of capacities (Roediger, Marsh, & Lee, 2002).

At a general level, there are several terms describing different *forms* or *kinds* of memory. As such, a distinction is often made between *short-* and *long-term* memory. Short-term memory refers to the retention of small amounts of information over periods of a few seconds, whereas long-term memory refers to storage of information over longer time periods (minutes to years). With regard to long-term memory, a distinction is also often made between *declarative* and *non-declarative* memory. Declarative memory

deals with propositional knowledge that is under conscious control – “knowing that”, and non-declarative memory is memory for nonverbal behaviors and skills – “knowing how” (e.g., Squire, Knowlton, & Musen, 1993).

A more recent distinction is that of *prospective* versus *retrospective* memory. Prospective memory (or remembering) is a hypothetical cognitive ability or set of abilities that enable the carrying out of intentions at a future point in time, and retrospective memory is memory for past events (Burgess & Shallice, 1997).

Memory systems

Theories of memory generally concern the way in which memory is organized, that is, the *architecture* and the activities therein, that is, the *processes* (e.g., Eysenck & Keane, 2005). The view that human memory may be described in terms of multiple systems has been influential. Three main criteria have been proposed for identifying a memory system (Schacter & Tulving, 1994):

(1) *Class inclusion operations*. These refer to the class of information that the system operates on, and are inclusive in the sense that the system can process a wide range of information of a particular kind, regardless of the content.

(2) *Properties and relations*. The properties concern rules of operation, kind of information, neural substrates, and statements of what biological function the system fulfills. The relations concern relations between the system and other accepted systems in terms of the properties.

(3) *Convergent dissociations*. A memory system should differ clearly in several ways from other memory systems. These differences are convergent in the sense that they are of different *kinds*, observed with different *kinds of tasks*, in different *populations*, and with different *techniques*.

Schacter and Tulving (1994) proposed that there are five major memory systems; *working memory*, *semantic memory*, *episodic memory*, *procedural memory* and *the perceptual representation system*. Working memory is a system that deals with temporary storage and maintenance of information in the form of internal representations. It also functions as to mediate controlled manipulation of these representations. The information in store is short lived and working memory operates at a conscious level (Nilsson, 2003).

Semantic memory is a declarative long-term memory system that is the basis for an individual’s general knowledge about the world in terms of facts, concepts and vocabulary. There is, furthermore, evidence of two subcomponents of semantic memory; *knowledge* and *fluency* (e.g., Nyberg, Maitland, et al., 2003).

Episodic memory is also a declarative long-term memory system that enables acquisition and retrieval of personal experiences that occur at a particular time and place. For example, remembering what one had for dinner yesterday would require the use of episodic memory. There is also evidence of two subcomponents of episodic memory; *recall* and *recognition* (e.g., Nyberg, Maitland, et al., 2003).

Procedural memory, in turn, is a non-declarative long-term memory system that is the basis for the learning of motor and cognitive skills, for example, learning to ride a bike, learning to read, and so forth.

Finally, the perceptual representation system is a non-declarative memory system that operates on perceptual information about the form and structure of words and objects (Schacter, Wagner, & Buckner, 2000).

Support for the memory systems model comes from several sources and disciplines. The similarities and dissimilarities in behavior of people with amnesia or brain damage have been particularly important and informative (see, e.g., Schacter & Tulving, 1994; Schacter, et al., 2000; Squire, 2004).

Cognitive functions in healthy aging and Alzheimer's disease

A large number of studies have found age-related declines in many cognitive domains, for example, speed, working memory, episodic memory, spatial ability, reasoning, executive function, and so forth (e.g., Allen, et al., 2001; Baltes & Lindenberger, 1997; Byrne, 1998; Park, et al., 2002; Salthouse & Ferrer-Caja, 2003; Verhaeghen & Salthouse, 1997).

Episodic memory function, in particular, appears to be highly sensitive to aging, with a pronounced and linear decrease in episodic memory performance as a function of age in cross-sectional studies (for reviews, see e.g., Bäckman, Small, & Wahlin, 2001; Craik, 2000). Judging from longitudinal analyses, episodic memory function is relatively stable up to middle age, followed by a sharp decline in old age (e.g., Rönnlund, Nyberg, Bäckman, & Nilsson, 2005). The evidence also suggests that working memory function is affected in a detrimental way by aging (e.g., Dobbs & Rule, 1989; Salthouse, 1994).

Semantic memory, the perceptual representation system, and procedural memory all appear to be relatively stable across the adult life span (e.g., Nilsson, 2003).

Most cognitive functions are severely impaired in Alzheimer's disease. There is also evidence of a preclinical phase with multiple deficits in cognitive function (e.g., Bäckman, Small, & Fratiglioni, 2001). Even though the impairment in cognitive functions appears to be global (possibly with the exception of short-term memory), episodic memory function seems especially impaired (Bäckman, Jones, Berger, Laukka, & Small, 2005).

Furthermore, the impairment appears in all stages of episodic memory processing, that is, encoding, storing and retrieving information.

Given the impairments in various cognitive functions with aging, a common explanatory cognitive mechanism has been sought for. Processing speed and working memory in particular appear to explain much of the age-related variance in other cognitive functions (e.g., Verhaeghen & Salthouse, 1997).

Concluding remarks

The chapter on cognitive functions has been very brief considering the vastness of the subject. However, the main take-home message is that many cognitive functions, especially episodic memory, are affected by both aging and dementia of the Alzheimer type. This is particularly important to remember as when people in general refer to “memory”, episodic memory is probably what they have in mind.

4. Subjective memory/cognition

The subjective–objective distinction as described previously is also often made in research on memory and cognitive functions. The majority of empirical studies and instruments have focused specifically on memory. From now on, the review will mainly be concerned with, and refer to subjective memory. As mentioned previously, memory may be considered as an aspect of the more general term cognition. Hence, the term subjective memory could be replaced with subjective cognition.

Measures of subjective memory

Empirical studies of subjective memory have used a variety of measures, scores from questionnaires in both un-standardized and standardized format, subsets of items, or single items. A number of reviews of available memory questionnaires have been published during the years (e.g., Gilewski & Zelinski, 1986; Herrmann, 1982; Perez Garcia, Godoy Garcia, Vera Guerrero, Laserna Triguero, & Puente, 1998). The questionnaires vary in many ways in regard to scope, item content and dimensionality. The scope or content of items may, for example, be in terms of *forgetting*, *remembering*, *memory quality*, *memory change*, *memory use* and *attitudes about memory* (Herrmann, 1982), *perceived differences between age groups*, *seriousness of memory failures*, *strategies and mnemonics usage*, *overall judgment of memory functioning*, *memory knowledge*, *demands on memory in daily life*, *memory for past events*, *effort when forgetting occurs*, or *relationship of memory to personality traits* (Gilewski & Zelinski, 1986). Apparently, the common denominator is the *word* memory, but it is questionable whether all

should be regarded constructs within a theoretical framework of memory. The heterogeneity makes a structured review very difficult. Henceforth, the focus will be on questionnaires where a main research question is the relation to laboratory test performance or diagnoses with cognitive dysfunction as a feature.

What do memory questionnaires measure?

Many more or less different terms are used in the literature regarding subjective memory, for example, *memory problems* (e.g., G. E. Smith, Petersen, Ivnik, Malec, & Tangalos, 1996), *subjective memory beliefs* (e.g., Cook & Marsiske, 2006), *metamemory* (e.g., McDonald-Miszczak, Gould, & Tychynski, 1999), *forgetfulness* (e.g., Mol, van Boxtel, Willems, & Jolles, 2006), *self-perceptions of memory* (e.g., Podewils, McLay, Rebok, & Lyketsos, 2003), and *subjective memory loss* (e.g., Clarnette, Almeida, Forstl, Paton, & Martins, 2001). This is likely as a consequence of the various disciplines that have an interest in the subject (a quick search in *Google scholar* with the keywords “subjective memory” results in research articles in the disciplines of, for example, *Cognitive psychology, Gerontology, Clinical psychology, Neurology, Psychiatry, Neuropsychiatry, Clinical neurophysiology, and Cognitive neuroscience*). However, even within disciplines there is considerable heterogeneity in regard to terminology and theory.

Possibly, the most common terms in use are *Memory Complaints* and *Subjective Memory Complaints*. Their popularity is likely a result of the heightened interest in the “transitional period between normal ageing and the diagnosis of clinically probable very early AD” (Alzheimer’s Disease; Petersen, 2004, p. 183), usually termed *Mild Cognitive Impairment (MCI)*. “Mild cognitive impairment is a syndrome defined as cognitive decline greater than that expected for an individual’s age and education level but that does not interfere notably with activities of daily life” (Gauthier, et al., 2006, p. 1262). Subjective memory complaints or self reports are included in clinical staging systems for dementia such as the *Global Deterioration Scale (GDS; Reisberg, Ferris, de Leon, & Crook, 1982)* and the *Clinical Dementia Rating scale (CDR; Hughes, Berg, Danziger, Coben, & Martin, 1982)*. Diagnostic systems such as the *Diagnostic and Statistical Manual of Mental Disorders – DSM IV-TR (American Psychiatric Association, 2000)* and the *International Statistical Classification of Disease and Related Health Problems - ICD-10 (World Health Organization, 1992)* also refer to subjective memory complaints.

Common to the diagnoses of dementia and MCI is that the impairments in cognition should constitute a change from a previous level of functioning. Usually, in clinical practice a patient is not assessed by means of cognitive tests prior to the onset of the condition. Hence, reports (subjective memory

complaints) by the patient and/or people in their surrounding, usually family or relatives, are believed to constitute a source of information regarding previous levels of functioning. Petersen (2004) states that “This criterion is ‘soft’ and may be a challenge to implement, but without prior cognitive function testing, it is critical for the purpose of excluding individuals with lifelong static cognitive deficits” (p. 189).

Viewed from an instrument development and validation perspective a clear and unambiguous definition of the concept (i.e., construct) is of great importance. Unfortunately, adequate definitions of memory complaints seem to be lacking. Below are some excerpts of sentences, more or less explicitly stated as definitions:

are thought to reflect the memory problems that people experience in everyday life (Gilewski & Zelinski, 1986, p. 93).

questionnaires designed to measure person's assessment of his or her own memory functioning in everyday situations (Gilewski & Zelinski, 1986, p. 93).

refer to everyday concerns cited by people both with and without objective evidence of memory impairment (Mitchell, 2008, p. 497).

complaints may appear in the form of specific memory deficit and sometimes, within broader contexts, include affective aspects (Zandi, 2004, p. 353).

a broad definition that incorporates any self-report of difficulties with memory as assessed by questionnaire in a research study (Reid & MacLulich, 2006, p. 472).

self-reported problems remembering different types of information (Hertzog, Park, Morrell, & Martin, 2000, p. 257).

a generic description of subjects who seek medical aid on their own initiative and who do not represent any categorical or diagnostic attribution (Guarch, Marcos, Salamero, & Blesa, 2004, p. 352).

As memory complaints are usually closely related to clinical assessment and diagnosis, we turn to the DSM IV-TR Glossary (American Psychiatric Association, 2000). Howsepian (2008) has noted that problems occur in the communication between clinicians and researchers due to limitations pertaining to clarity, logical coherence, factual content, redundancy, and ambiguity in the glossary. The DSM makes a distinction between *signs* and *symptoms*:

Sign: An objective manifestation of a pathological condition. Signs are observed by the examiner rather than reported by the affected individual (as cited in Howsepian, 2008, p. 33).

Symptom: A subjective manifestation of a pathological condition. Symptoms are reported by the affected individual rather than observed by the examiner (as cited in Howsepian, 2008, p. 33).

These definitions shed some light on the issue. It appears that memory complaints are symptoms. However, the reference to “A subjective manifestation of a pathological condition” presupposes the presence of pathology, which seems like begging the question when it comes to memory complaints. In connection with insomnia diagnoses, Howsepian (2008) comments that there are no complaints that are not subjective, so the adjective may be redundant and confusing. Nevertheless, from the attempts at definitions it seems that these memory complaints are *verbal behavior* (or in some cases even the more physical behavior of seeking medical consultation). The semantic content of this verbal behavior apparently refers to memory. In the case of MCI, the examiner wants *signs* of previous levels of cognitive functioning, but in their absence use reports (subjective complaints). From this it should be evident that the concept “subjective memory complaints” has clear limits of application for research purposes. In the case of MCI, the examiner wants a proxy to prior cognitive function testing. “Subjective memory complaints” appears to be an unsuitable concept for such a proxy.

The relation between subjective and objective measures of memory

A major issue when it comes to subjective measures of memory is the relation to objective memory performance on laboratory tests. This hypothesized relation may be posed as a research question (are there relations between variables or constructs?), or in terms of validation of questionnaires (investigating external, concurrent or predictive validity). Unfortunately, these two objectives are often confused. The distinction between the two becomes clearer and especially important when dealing with instruments in standardized format that are used repeatedly. Regardless of the aims of the studies and the types of subjective measures, the evidence for a substantial and reliable relation between subjective and objective measures of memory is weak (for reviews, see e.g., Gilewski & Zelinski, 1986; Herrmann, 1982; Perez Garcia, et al., 1998; Reid & MacLulich, 2006). The most common finding is that of a relationship with depressive symptoms, rather than objective memory measures (e.g., Bolla, Lindgren, Bonaccorsy, & Bleecker, 1991; Zandi, 2004). A relationship with personality traits is also often found (e.g., Hänninen, et al., 1994; Pearman & Storandt, 2004, 2005; Ponds & Jolles, 1996).

Measuring subjective memory: Three paradigms

Upon reading the research articles on subjective memory and reviews of memory questionnaires, I have found that thinking in the terms of paradigms is useful. This is because of the differences in what subjective measures are claimed to reflect (i.e., constructs). Sometimes, the same standardized questionnaire is used in different studies, but the constructs claimed to be measured differ. Furthermore, virtually the same results may lead to conclusions that are rather conflicting viewed across different paradigms. The third meaning of paradigm in the *Merriam-Webster online dictionary* reads “a philosophical and theoretical framework of a scientific school or discipline within which theories, laws, and generalizations and the experiments performed in support of them are formulated” (paradigm, 2011). Summarizing Kuhn, Shuttlesworth (2008) describes that a paradigm dictates:

- *What is studied and researched*
- *The type of questions that are asked*
- *The exact structure and nature of the questions*
- *How the results of any research are interpreted*

These sentences capture very well my reasons for adopting the concept in regard to subjective memory. Below follow the three major paradigms as I conceive them. In practice, the borders are not clear, and aspects of the paradigms may influence the same research agenda.

Memory complaints paradigm

Clinicians are often confronted with patients who say that they have problems with their memory. These memory complaints need to be further investigated in the course of diagnosis. A number of conditions or diagnoses are implicated (or none at all). A diagnosis of memory impairment is mainly based on assessment by means of neuropsychological tests and norms for the healthy population. Importantly, other conditions must be excluded as a cause (e.g., depression), and the clinician needs some estimate of previous memory/cognitive function to make a diagnosis of, for example, dementia. Hence, the verbal memory complaint is not necessarily accompanied by memory impairment as defined by neuropsychological tests.

The main questions for research in this paradigm are whether memory complaints are indicative of, or predict future, memory impairment as defined by neuropsychological tests, and if not, what other causes there may be. Elderly and clinical populations are the majority of those researched. The main audience for the publications consists of clinicians who may benefit from the results in their practice. An epidemiological perspective is often taken, where the prevalence of memory complaints is investigated, much like

for physical diseases or related symptoms (see e.g., Jonker, Geerlings, & Schmaud, 2000 for a review). Even though there are multi-item questionnaires developed within this paradigm, the most indicative studies usually utilize a few, or single items, similar to “Do You have problems with your memory?” (e.g., Bassett & Folstein, 1993).

Everyday memory paradigm

Research on everyday memory stresses the study of memory and cognition in everyday settings or tasks that closely resemble everyday situations. Laboratory tests are often seen as artificial and of questionable “ecological validity” (see e.g., Bower, 2000). Hence, everyday memory may be seen as an alternative paradigm to research in the laboratory. The questionnaire is one of the methods employed. The populations studied are not confined to, or even mainly the elderly or clinical. The questionnaire is seen as a method by which people can report potentially public events. “Memory failures” is therefore a common concept for questionnaires in this paradigm. The audience for publications is mainly other researchers, but these may also be involved in clinical assessment. Examples of instruments developed within this paradigm are the *Cognitive Failures Questionnaire* (Broadbent, Cooper, FitzGerald, & Parkes, 1982) and the *Everyday Memory Questionnaire* (Sunderland, 1983).

Metamemory paradigm

Metamemory refers to a collection of different types of cognitive processes and information. Three general categories have been proposed (Hertzog & Dixon, 1994, pp. 227-251):

1. *Declarative knowledge about memory tasks and memory processes – defined as knowledge about how memory functions and the viability of strategic behaviours for tasks requiring memory processes*
2. *Memory monitoring – defined as the awareness of the current state of one’s memory system*
3. *Self referent beliefs about memory.*

The reference or standard for metamemory research is usually objective laboratory test performance. The cognitive processes of metamemory are studied in relation to this test performance. Hence, terms such as *accuracy*, *awareness* and *insight* are often used. The last category, *beliefs*, appears to be a subjective construct. The populations studied are the elderly, clinical and the general population. The audiences for the publications may be basic researchers, clinical researchers and clinicians. An example of a questionnaire developed in this paradigm is the *Metamemory in Adulthood Questionnaire* (Dixon, Hultsch, & Hertzog, 1988).

5. Validity and validation

Validity is a fundamental issue in research, measurement, and assessment. The concept is perhaps most often associated with the applied use of tests for assessment purposes, that is, *test validity*. The typology and focus does however differ somewhat depending on the discipline. When tests and instruments are used in basic research, their validity rather becomes as an issue within the validity of the research process as a whole. This is perhaps clarified by the related concept *validation* - the process whereby evidence of validity is gathered (Urbina, 2004). Many aspects of validity are the same in the use of tests and instruments for assessment and research. However, assessment has additional practical aspects that are increasingly being subsumed under the umbrella term “validity”, but the relevance of these to the research process appears limited in many cases. Below is a division of validity into an assessment perspective and a research perspective. Unfortunately, the assessment perspective, often signified by the terms *testing* and *test validity*, is seldom sufficiently explicit in acknowledging that by testing is meant *the use of tests for assessment purposes*.

Validity in an assessment perspective

The classical view of validity

Traditionally, a test or instrument has been considered valid if it measures what it purports to measure (e.g., Kelley, 1927). Three major meanings have been given the concept: criterion-related validity, content validity and construct validity.

Criterion-related validity involves the correspondence between an instrument score and a criterion variable that is external to the instrument. If the criterion is obtained at the same time it is called *concurrent validity*, and in the future *predictive validity*. Correlation and regression are the main methods used for gathering evidence of criterion-related validity.

Content validity involves showing that the items in, or behaviors elicited by an instrument are a representative sample from the intended domain of content or behavior.

Construct validity involves showing that the instrument measures the (theoretical) construct intended.

The recent view of validity

The *Standards for Educational and Psychological Testing* state that “Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests” (AERA, APA, & NCME, 1999, p. 9). This definition reflects the idea of *construct*

validity as a the overarching unified concept in the world of validity (Lissitz & Samuelsen, 2007). Most influential in the transition to what is often called “modern validity theory” (e.g., Cizek, Rosenberg, & Koons, 2008) are probably S. Messick (e.g., Messick, 1989) and M.T. Kane (e.g., Kane, 2001). These theorists have heralded the broad unitary validity concept, where even *social consequences* of testing have been included. Furthermore, validity is seen as a matter of degree, and not a property of tests, but refer to the *appropriateness, meaningfulness* and *usefulness* of score-based inferences (e.g., Messick, 1995).

The current conception appears to be accepted by many in the test user community, but has its critics. Borsboom, Mellenbergh and van Heerden (2004) state that “A theory of validity that leaves one with the feeling that every single concern about psychological testing is relevant, important, and should be addressed in psychological testing cannot offer a sense of direction to the working researcher” (p. 1061). The inclusion of consequences of testing in the validity concept may be the most controversial, and has also been questioned by others (e.g., Cizek, et al., 2008; Lissitz & Samuelsen, 2007; Mehrens, 1997; Popham, 1997; Sechrest, 2005). Sechrest (2005) even states that the standards “do not necessarily make much sense once one gets beyond the realm of commercially marketed instruments” (p. 1585).

Validity in a research perspective

In their book on experimental and quasi-experimental designs, Shadish, Cook and Campbell (2002) use the term validity to refer to the “approximate truth of an inference” (p. 34). They present four types of validity; *statistical conclusion, internal, construct* and *external* validity.

Statistical conclusion validity concerns the approximate truth of inferences as to (1) whether a cause and effect covary, and (2) how strong this covariation is.

Internal validity concerns the approximate truth of inferences about whether an observed covariation between two variables, A and B, reflects a causal relationship from A to B in the form in which the variables were manipulated or measured.

Construct validity refers to the approximate truth of inferences concerning the higher order constructs that represent sampling particulars (persons, settings, treatments and outcomes).

External validity refers to the approximate truth of inferences concerning whether a cause-effect relationship holds over variation in persons, settings, treatment variables, and measurement variables.

A comment on validity

Psychological (and educational) tests are primarily used for making decisions about a person, a group, or a program. This decision making

should always take place in the context of (psychological) assessment (Urbina, 2004). Tests are also often used in research, but there are usually no direct practical consequences. For basic psychological research, the focus on construct validity and the unitary view of validity appears well suited. The problem is that there are many practical aspects in the primary use of tests that do not readily fit into the unitary view. Neither do these practical aspects have relevance for construct validity in basic research. Ethics and feasibility of a research agenda should always be considered, but it is questionable whether these should be regarded as aspects of validity.

In my view, the broad and unitary validity concept as reflected in the *Standards* (AERA, APA, & NCME, 1999) is almost impossible to comprehend and apply for research purposes. There are certainly many aspects of the use of tests for assessment that are important to consider, but to include them under the heading of validity only obfuscates the concept, or as Mehrens (1997) puts it, "If validity is everything, then validity is nothing" (p. 18).

The view that validity is about the *appropriateness, meaningfulness* and *usefulness* of score-based inferences (e.g., Messick, 1995) appears to have almost completely detached from the empirical domain and any notions of truth or existence of what is measured. The classical types of validity are often claimed to have been introduced because of shortcomings in solely relying on one definition, and the need for other standards for different kinds of instruments. Another and perhaps more illuminating explanation is a change in the ruling paradigms of philosophy of science and (psychological and educational) science. A shift can be traced from a predominately *positivist* and *behaviorist* approach (criterion validity), via *scientific realism* (early conceptions of construct validity), to the current *instrumentalism* or *pragmatism*. See, for example, Anastasi (1950) and (Bechtoldt, 1959) for early and particularly illuminating articles.

In an understandable and sound reaction to the current conception of validity, Borsboom, et al. (2004) have put forward their own conception: "a test is valid for measuring an attribute if (a) the attribute exists and (b) variations in the attribute causally produce variation in the measurement outcomes" (p. 1061). Surprisingly, this recent and critical conception of validity have similarities with the description of construct validity made by Loevinger (1957). According to her, "The basic concept is that of the construct validity of the test, the degree to which it measures some trait which really exists in some sense" (p. 685).

The dividing line in the conception of validity appears to be between the uses of instruments for practical assessment and (basic) research. It is possible that the application of the validity concept is, or should be, different depending on the use. However, in my view this is not satisfactory and only leads to confusion. Many of the problems with the application of the validity

concept could be resolved by explicitly acknowledging that assessment in practice has a number of aspects that are important, but really not aspects of a scientific conception of validity. Introduction of an additional concept, such as *utility* could be fruitful. Lissitz and Samuelsen (2007) have proposed the use of the term mainly as substitute for “criterion related validity evidence”. However, their focus is mainly on practical testing in educational settings and put emphasis on content rather than theory. Kendell and Jablensky (2003) have also made a distinction between validity and utility, but their focus is on psychiatric diagnoses. They view many diagnostic categories as having utility in providing information about outcome, treatment response, and aetiology. But the authors claim that they are often not valid as discrete entities with natural boundaries to other disorders.

In my variant, the validity concept should be reserved for construct validity much in line with the unitary view. The utility then refers to usability for some purpose and practical assessment issues, such as consequences (for a similar view, see Cizek, Bowen, & Church, 2010). The utility concept should also apply to tests that are constructed for some practical, a-theoretical purpose, such as to predict some variable or outcome, or those dominated by the content aspects of validity. This does not mean that the claims for, and evidence of utility should be less stringent. I think this would serve both those involved in practical assessment and those who use instruments for research purposes.

The view of validity for the thesis

A variant of the classical definition of construct validity as described in, for example, Netemeyer, et al. (2003), appear to be the most relevant for the issues and the instrument developed in this thesis. According to them “*Construct validity refers to how well a measure actually measures the construct it is intended to measure*” (p. 11). In lieu with the *Standards* (AERA, APA, & NCME, 1999), the typology of validity is not to be interpreted as meaning that there are different types of (construct) validity that are optional or interchangeable. Nevertheless, the terms are useful in describing particular sources of validity evidence. Therefore, traditional nomenclature is used with the addition of the word “evidence”. The nomenclature and descriptions are based in part on Netemeyer, et al. (2003), but there are my own additions that appear particularly important for delineating the types. Factorial evidence is not included as a type of validity in Netemeyer, et al. (2003). The *Standards* (AERA, APA, & NCME, 1999) refer to evidence based on internal structure.

Evidence of construct validity:

- Content evidence
- Factorial evidence

- Criterion-related evidence
 - Predictive evidence
 - Concurrent evidence
 - Convergent evidence
 - Discriminant evidence
 - Known-group evidence

Content evidence is essentially classical content validity, that is, it involves showing that the items in, or behaviors elicited by, an instrument constitute a representative sample from the intended domain of content or behavior.

Factorial evidence, or evidence based on internal structure, is based on statistical analyses of the structure of the instrument.

Among the sources of *criterion-related evidence*, *predictive evidence* refers to the shown ability of a measure to predict a subsequent and temporally ordered criterion variable. In adopting the label, it seems important that this criterion is not a measure of the same construct, only obtained at a latter point in time. Preferably, the criterion should be some future behavior or condition that is theoretically linked to the predictor construct. In the case of instruments such as that developed in this thesis, this criterion could be cognitive impairment, that is, a future trajectory of decreasing cognitive performance, or a future diagnosis of AD or indicators thereof (e.g., pathological changes in the brain such as beta-amyloid plaques and neurofibrillary tangles).

Concurrent evidence refers to the degree to which a measure correlate with a criterion variable that is collected simultaneously. Like for predictive evidence, the criterion should not be a measure of the same construct, but a behavior or condition that is theoretically linked to the construct. In the current setting, the criterion could be some everyday behavior indicative of cognitive dysfunction, or a concurrent AD-diagnosis or indicators thereof.

Convergent evidence refers to the degree a pair of measures that are designed to measure the same construct correlate. By utilizing measures of the same construct, convergent evidence is delineated from predictive and concurrent evidence. Albeit often obtained simultaneously, the proximity in time should not be a prerequisite if the construct is stable over time in individuals.

Discriminant evidence refers to the degree to which a pair of measures designed to measure conceptually different, but similar, constructs correlate. In the current setting, depressive symptoms and stress appear to be particularly relevant constructs.

Known-group evidence refers to the ability of a measure for distinguishing reliably between groups that should differ in levels of the construct. In the current setting these groups may be based on different

diagnostic categories with cognitive dysfunction as features, or perhaps groups based on age.

6. Synthesis

This chapter attempts a synthesis of some key issues in the previous chapters by explicating the subjective–objective distinction in the development and validation of measures of memory and cognition.

Subjective/objective methods and constructs

Research on memory and cognition is often quantitative. Thus, objective scoring of overt responses is implicitly implied, and not usually the reason for making the subjective–objective distinction. In general, the verb *subjective* refers to the method for obtaining data, that is, by means of questionnaires, self reports, and self-rating scales. This methodological aspect of the subjective objective distinction is however not sufficient. The distinction may also apply to the phenomenon or construct the method is supposed to measure. Objective constructs are usually defined by, or derived from publicly observable events, whereas subjective constructs refer to those derived from personal inner states and communicated by language to other individuals. Subsequently, it is possible to measure a predominately objective construct (e.g., memory) with a subjective method (e.g., self report), and a subjective construct (e.g., memory beliefs) with a subjective method (self report). It is however not possible to measure a subjective construct with an objective method. Objective scoring of the overt responses is fully possible, but the means of obtaining the observation of the inner state is only available to the individual. To provide a blunt example, imagine a questionnaire about “health” issues. One item in such a questionnaire could query “Has your left arm been surgically amputated above the elbow?” with the response alternatives “yes” and “no”. Whether the arm has been amputated or not is an issue that is in theory and practice possible to investigate by external observers. The phenomenon is thereby in the objective domain, but the method for obtaining the information is subjective. It is fully possible, but perhaps unlikely, that the individual provides an incorrect answer. The correctness of the answer is possible to ascertain by physical examination. The key issue is that the item is formulated such that the state of affairs *can* be observed by external individuals, and not whether it is observed or not.

A second question could be “Do you feel content with your body appearance”, again with the response alternatives “yes” and “no”. The phenomenon is in the subjective domain, not accessible for external

observers, *and* the method is subjective. One could hypothesize a relationship between the responses, but the phenomena are independent.

Development and validation

Reflection upon the domain and nature of a construct is recommended before adoption or explication. The nature of the construct will govern what types of stimuli and possible responses are used in the instrument. If a construct is within the objective domain, the stimuli should be chosen so as to avoid contamination from the subjective domain. Hence, inexact words such as *problems* or *difficulties* should be avoided (consider the possible responses of an amputee to the question “Do you have problems with your arms?”).

The nature of the construct will also govern the adequacy of different types of validity evidence. Laboratory test performance as validity evidence is only adequate if the construct is in the objective domain. It is possible and completely adequate to study relations between constructs from different domains (subjective and objective), or to use one for predicting the other, but the validity concept should be reserved for evidence obtained within the domain. Hence, (construct) validity of *memory beliefs* or *memory complaints*, among others, cannot be inferred from correlations with laboratory test performance.

If attempts are made at measuring an objective construct by means of a subjective method there may be methodological limitations as to what degree of accuracy can be achieved. However, the instrument developer also has a responsibility to take care in the construction so as to fully exploit the method’s potential. In practice it will be difficult to separate methodological limitations from flaws in the construction of the instrument.

It could be argued that the subjective–objective distinction is not a methodological issue. That there are really only constructs and that what signifies the objective test is objective scoring of responses. The subtleties of any difference could be referred to the theoretical framework surrounding the construct. Although being a tempting argument, I think such reasoning is a mistake.

A schema for the subjective–objective distinction

Here I present a schema for the subjective–objective distinction when it comes to standardized measurement instruments. It specifically applies to ability tests and questionnaires or self-report scales. In the test theory tradition (e.g., McDonald, 1999), these would all be considered objective tests by virtue of objective scoring. The schema is the result of my attempt at achieving a structure for thinking about the subjective–objective distinction. It is at best a simplification of, and an approximation to, reality. The subjective–objective distinction is adopted at several steps and is a matter of

degree rather than clear cut dichotomies. The schema is inspired by Veenhoven (2004) and particularly the distinction between subjective- and accuracy-type response indicators made by Goldiamond (1958).

To explain the parts of the schema I will use two prototypical examples of instruments measuring constructs from a predominately objective or subjective domain. Both utilize a dichotomous response-format that can be objectively scored. Several aspects (e.g., guessing) have been omitted to promote clarity. Scaling is performed according to the linear (summative, centroid) model (Nunnally & Bernstein, 1994)

As the example of an objective construct, consider *memory* as measured by a simple memory test. The participant is presented with a written list of five common words for two minutes (stimuli). Two minutes later, the participant is presented with a written list of ten words (of which five were presented previously), and is to indicate whether each word was on the previous list by indicating “yes” or “no” (responses).

As the example of a subjective construct, consider *pain* as measured by a questionnaire. The participant is presented with a list of ten body parts, for example, left leg, neck, and so forth (stimuli), and is to indicate whether he or she is experiencing pain in respective body part by indicating “yes” or “no” (responses).

Table 1. A schema of the Subjective-Objective distinction

Theoretical domain:	Objective	↔	Subjective
Construct:	Objective	↔	Subjective
Stimulus:	Physical/Semantic content	↔	Subjective/Semantic content
Response:	Behavioral/Semantic content	↔	Semantic meaning
Indicator rule:	Explicit/Axiomatic	↔	Implicit
Response indicator:	Objective (Accuracy)	↔	Subjective
Scoring:	Objective		Objective
Scaling :	Stimulus centered / Subject centered		Stimulus centered / Subject centered

The theoretical domain

The first step in the development of instruments concern whether the theoretical domain is predominately subjective or objective. In the case of a weak theoretical framework without an explicitly formulated construct, this step should at least be possible. I will adhere to the definitions by Moore

(1995) of the subjective domain as that of “unobservable immediate experience, personal consciousness, and the self as an agent”, and the objective domain as “the domain of the material, physical, and publicly observable” (p. 33). Even though memory can be studied at an experiential level, much of our knowledge and theory of memory are based on behavior and publicly observable events. Hence, the scientific concept memory, essentially derive from the objective domain. In a related vein, pain can be inferred from observation of behavior or physiological variables, but is essentially defined by the unobservable immediate experience, and is therefore within the subjective domain.

The construct

In the second step the construct is explicated. An objective construct is usually defined by directly observable behavior. A subjective construct would be defined with a reference to a personal inner state, usually only directly accessible to the individual, and where the state is inferred, or even in some sense defined by language. To attain the status of a construct, it should also constitute a certain degree of theoretical abstraction from behavior rather than being a label for it. In case of the latter, it is questionable whether it is to be regarded as a construct. Validity in the sense of construct validity is thereby hard to adopt. Evidence of content relevance (i.e., content validity) should probably take precedence. In our examples, the memory construct is objective and the pain construct is subjective.

The stimulus

In both these examples, the stimuli are in the form of written items. With regard to the memory items, it is the explicit *semantic content*, represented by written words that exhibit objective properties. For the pain items, it is the implicit *semantic meaning* that exhibit subjective properties. In general, the presence or absence of a stimulus for an objective construct is usually publicly observable and physical. The presence or absence of the referent of the semantic meaning (i.e., pain) is neither.

The response

In both the examples, the overt responses are indications of “yes” or “no”. In general, objective constructs tend to utilize response repertoires that are publicly observable behavior.

The scoring rule and the response indicator

There is one main reason for making the subjective–objective distinction in regard to the two instruments in the examples. It concerns the rules for assigning numerals to the responses. We could call both *rules of presence*. For memory items, the rule of presence is explicit and axiomatic, based on

the correspondence or non-correspondence between the stimulus and the response alternative (semantic content), that is, there is a correct response. The presence of the correct response is coded “1” and the absence “0”. The response indicator is of the *objective* or *accuracy* type.

In the case of pain, the rule is implicit and there is no correct response. The rule of presence is based on the correspondence or non-correspondence between the inner state and the semantic meaning of the response alternative. It furthermore rests on the assumption that the participant can observe the presence of pain (coded “1”), or its absence (coded “0”). The observation may be faulty, but we have no obvious means of determining whether it is so. The response indicator is of the *subjective* type.

Scaling

Scaling is the assignment of numbers to objects to represent quantities of attributes (Nunnally & Bernstein, 1994). These “objects” can either be stimuli or individuals. There are generally two types of scaling: *stimulus centered* and *subject centered*. The subject centered methods deal with individual differences among individuals, whereas the stimulus centered methods deal with differences between objects used as stimuli. The latter is most related to *psychophysics*, that is, “the study of the relation between variation in physical dimensions of stimuli and their associated responses” (Nunnally & Bernstein, 1994, p. 33).

Generally, the thesis has focused on differences between individuals and thereby the subject centered methods. Stimulus centered methods are generally used in the development of instruments to derive item statistics.

In psychometrics, both for objective and subjective constructs, the properties of the items (e.g., difficulty) are derived from the responses. When the properties of the stimuli (items) are inferred rather than known beforehand (as they are in psychophysics) there is a conflation between the magnitude of the attribute (construct) and the magnitude of the stimulus. Although, some methods are claimed to not suffer from this (e.g., Item response theory models). In the case of a subjective construct, the overt response has to be used as a response indicator, but it is not necessarily correct, and the implicit indicator rule is unknown. This causes a higher degree of uncertainty than for objective response indicators. The consequences of this should be explored.

Implications for the development and validation of instruments in the subjective memory domain

Memory complaints paradigm

First of all, the concept of memory complaints needs to be scrutinized. It appears that the abstraction from behavior and theoretical framework is very

limited. Memory complaint is rather a descriptive label for a predominately verbal behavior observed in clinical situations. This causes a clash with the definition of a construct as an abstract unobservable latent variable. The verbal behavior is in theory possible to record for external observers, that is, it is in the objective domain, but it appears not to be a construct. It would be more suitable to regard it as a concept that describes a domain of behavior, without reference to its hypothesized cause or causes (latent variables). As a consequence, construct validity is not implied, and content validity (or utility) aspects take precedence. Subsequently, in developing an instrument for measuring memory complaints, the item or items should elicit a behavior or sample of behaviors representative of the domain of complaining behavior. Assuming that memory complaints are verbal behavior, the semantic content of items (stimuli) should be representative of the semantic content observed in the clinic (or wherever it is expressed). The response indicator would be of the objective type, but not an indicator of accuracy.

An alternative view, perhaps better described by the term “memory problems” than “memory complaints” could regard it as a subjective construct. This would imply a unique personal meaning as to what memory is and what constitutes a problem. The *semantic content* of the items should then elicit the *semantic meaning* for the individual. The response indicator is of the subjective type.

If, on the other hand, the objective is to construct an instrument that will measure memory/memory impairment as defined by neuropsychological tests (i.e., an objective construct) the label memory complaints is no longer suitable. Furthermore, then the responsibility of the instrument developer becomes constructing items that elicit the relevant responses within the confines of the method. The correlation between scores from an instrument measuring memory complaints and scores from neuropsychological tests cannot be used to evaluate its validity. It may very well be that there is a relation whereby one can predict the other to some extent, and it may be useful in both practical assessment situations and research, but it does not say anything about validity. Predictive utility would be the preferred term.

If memory complaints, despite the terminology is regarded as a subjective construct the same reasoning apply. The validity of a subjective construct cannot be evaluated or questioned with reference to the objective domain.

Everyday memory paradigm

As an alternative to laboratory based research, everyday memory research should have the possibility of adopting the same theoretical frameworks (e.g., memory systems). Consequently, the domain and its constructs are objective. It is possible that the behavioral indicators differ, but they are still in the objective domain. Hence, contamination from the subjective domain should be avoided in the development of instruments. Assuming that the

constructs are the same for measures in laboratory and everyday memory research, it is possible to evaluate aspects of construct validity of either by means of correlation. It is also possible to evaluate the construct validity of an everyday memory questionnaire (a subjective method), either by designing an objective method that capture the same behavioral indicators in everyday situations, or correlating its scores with laboratory measures of the same construct.

Metamemory paradigm

Metamemory appears to consist of three major sub-constructs; *knowledge*, *monitoring* and *beliefs*. The first two seem to be objective constructs, and the last a subjective construct. Construct validity should be possible to evaluate according to regular procedures for knowledge and monitoring, and objective/accuracy response indicators are implied. A somewhat difficult task is to outline the domain of knowledge and monitoring. Furthermore, it should be addressed whether knowledge and monitoring are individual-difference constructs or not.

Being in the subjective domain, the validity of the construct beliefs cannot be evaluated with reference to the objective domain. The response indicators are of the subjective type. Given that the apparent semantic content is in agreement with knowledge or monitoring constructs, it is possible to study their interrelations. One could possibly then refer to the accuracy of beliefs, but it should be remembered that this does not apply to construct validity.

A complication with the metamemory construct is that beliefs are sometimes defined as incorrect knowledge or monitoring, (see, e.g., Herrmann, 1982; Herrmann, Sheets, Gruneberg, & Torres, 2005).

Concluding remarks

In this section, I have proposed some implications for the development and validation of instruments in the subjective memory domain with a basis in the subjective-objective distinction. The schema provided was an attempt to explicate the differences and similarities between subjective and objective measures in terms of *what* is measured and *how* it is measured. It is, however, a somewhat artificial description. In reality, the different steps are intertwined. Objective and subjective constructs are in a sense defined by the methods used to measure them, and the methods we use influence the constructs.

In regard to the different paradigms, the picture is complicated by a blending in some cases. Furthermore, some of the terms used in the subjective memory domain have questionable status as constructs.

Nevertheless, this theoretical framework may be helpful by initiating reflection on the use of the subjective-objective distinction, particularly in regard to memory and cognition. It is my attempt at explicating the thought

processes that were at work before, during, and after the development of the Cognitive Dysfunction Questionnaire.

7. Development of the Cognitive Dysfunction Questionnaire

Several texts describe the steps involved in the development of instruments (e.g., Clark & Watson, 1995; Nunnally & Bernstein, 1994; Wilson, 2005). One of course aims to develop a valid instrument from the start, so development and validation are in a sense intertwined. The development of the CDQ is extensively described in Studies III and IV, and generally followed four major steps as summarized by Netemeyer, et al (2003);

(1) *Construct definition and content domain*; a literature review, theory, a clear construct definition and identification of the content domain are stressed in the first step. The dimensionality of the construct also needs to be addressed.

As described in previous chapters, the literature review of subjective memory measures revealed a highly heterogeneous field with few clear theoretical frameworks and constructs. This is likely not a result of a lack of theories and models of memory and cognition, but the fact that these belong to the objective domain and are based on laboratory research and objective tests. Studies I and II provided some general results as to the relations between variants of subjective memory measures and other important variables, such as objective memory test performance, age, depressive symptoms, personality and perceived causes of “memory problems”. This information, along with suggestions from the literature was used when constructing the CDQ.

The decision was to adopt a global construct of cognitive functioning, with the memory systems model as the main theoretical framework. In terms of the paradigms as described previously, the everyday memory paradigm is the best description of the approach. Cognitive functioning was regarded as an objective construct and the method for measuring the construct subjective.

The global construct implies multiple related dimensions (factors) and a higher-order general dimension (factor).

(2) *Generating and judging measurement items*; in this second step, a pool of items is constructed and evaluated. Furthermore, the response-format is determined and item content is related to theoretical dimensionality. The judging of items can be performed by both laypersons and experts.

A comprehensive pool of items was constructed. The general strategy was that the content should describe events that are potentially publicly observable, and similar to events recorded by objective tests in the

laboratory. The main difficulty was to translate theoretical descriptions and definitions into events that could be observed and reported by the individuals in everyday life. A five-point scale for relative frequency of occurrence was chosen as the response format. Finally, the items were reviewed by several people and revised prior to pilot testing.

(3) *Designing and conducting studies to develop and refine the scale*; in this step the item pool is pilot tested on a relevant sample, exploratory factor analysis and item analyses are performed, initial estimates of internal consistency (reliability) and validity are evaluated. Finally, items with desirable properties are selected.

The pilot testing was performed in a sample of adults that participated in the Betula study some years previously. The main method for item selection was correlations with criterion measures derived from laboratory tests of memory and cognition. Twenty items in six domains were chosen. The results from early attempts at validation were encouraging as to internal consistency and relations to relevant variables.

The first part of Study IV may also be seen as involving this step. Twenty additional items were constructed to provide an equal number of items for the domains, and enable exclusion of items with similar wording. This refinement is further described in step four.

(4) *Finalizing the scale*; the final step may be seen as validation and involve several issues and analyzes from the previous steps. Several samples are utilized to test the consistency of factor structure and psychometric properties. Furthermore, norms across studies can be established.

The beginning of this step is described in Study IV. A final refinement of the CDQ was undertaken by means of confirmatory factor analysis. Two domains were collapsed and items were excluded, resulting in an instrument with five domains with four items each. The factor structure was cross-validated on an independent sample, and competing models were tested. Multi-group confirmatory factor analysis was performed to investigate measurement invariance across age and sex. Finally, psychometric properties and some evidence of validity were presented.

The Betula Prospective Cohort Study

The articles in the thesis are based on data from participants in the Betula Prospective Cohort Study (Nilsson, et al., 2004; Nilsson, et al., 2000; Nilsson, et al., 1997), an ongoing research project that started in 1988. The main aims of the project are to study how memory functions change during adulthood and old age, to identify risk factors for dementia, and to identify early preclinical signs of dementia. The participants belong to random samples from the population in Umeå municipality, stratified by age cohort and sex. To this date, six samples (S1-S6) have been included, and there have been five occasions for data collection at different time points (T1-T5),

totaling circa 4500 persons who were 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, or 80 years of age at inclusion in the study. The participants are investigated at five year intervals by means of cognitive tests, interviews and medical examinations. The articles in the thesis are based on participants from S4 at T3, S5 at T4, and S1, S3 and S6 at T5.

Nilsson, et al. (1997) have described the battery of cognitive tests used in the study. The tests mainly target various aspects of episodic and semantic memory, but also the perceptual representation system (priming), general cognitive functioning, visuospatial constructional ability, and so forth. The specific tests used for the studies in the thesis are extensively described in the articles (for additional information on the design of the Betula study, see, Nilsson, et al., 2004; Nilsson, et al., 2000; Nilsson, et al., 1997).

8. Statistical methods

Classical test theory

The CDQ was developed within the framework of classical test theory, that is, scoring of tests as linear combinations of responses to individual items (Nunnally & Bernstein, 1994). The set of statistical methods in the thesis is based on the *true score model*. Briefly, the true score model assumes that a randomly drawn object of measurement (individual) has a true score on an attribute. An observed score is a random variable that contains error and thereby differs from the true score. The observed score is the sum of true score and a random error component. From the assumptions follow that (1) the mean of errors for a population is 0, (2) the correlation between true and error scores is 0, and (3) the correlation between errors on separate measurements is 0. The model can be expressed $Y = T + E$,

where Y = observed score for a randomly drawn individual

T = random true score for an individual

E = random error

From the model follows that T and E are measured on the scale of Y and bounded within the range of Y .

As T and E are uncorrelated, the following is true of their respective variances:

$$\sigma_Y^2 = \sigma_T^2 + \sigma_E^2 \text{ and } \sigma_T^2 \leq \sigma_Y^2 \text{ and } \sigma_E^2 \leq \sigma_Y^2$$

From this, the reliability coefficient of Y can be defined as

$$\rho_y = \sigma_T^2 / \sigma_Y^2 = \sigma_T^2 / \sigma_T^2 + \sigma_E^2$$

In essence, the reliability coefficient is the ratio of true score variance to observed score variance. In practical applications the reliability coefficient has to be estimated from actual test scores. For this there are three main strategies; *test-retest*, *parallel or alternate forms*, and *internal analysis*. The first two are based on the correlation between two total test scores, and the last is based on relations between items making up a single test score, that is, a composite.

Cronbach's coefficient α estimates the reliability of a composite expressed as a function of the variance of the composite score and the covariances of the items (or tests) that form the composite. It can be used to estimate the lower bound of the *coefficient of precision* of a composite. Its usefulness rests on the idea that any test can be regarded as a composite of items. What needs to be known for the computation of Cronbach's coefficient α is the number of items, the variance of the composite score, and the sum of all item covariances.

Exploratory and confirmatory factor analysis

The main methods used in the development and validation of the CDQ were exploratory (EFA) and confirmatory factor analysis (CFA). Both methods are based on analyses of associations between variables in a set, usually a correlation- or covariance matrix. As the name suggests, EFA is used to *explore* the set of variables for clusters, named factors or components depending on extraction method (how the matrix is analyzed). It is essentially a data driven method where factor structures are isolated without consideration of the theoretical expectations of the researcher (Thompson & Daniel, 1996). In contrast, CFA tests the fit of data to a pre-specified factor structure. Hence, CFA is essentially theory driven, and also provides the possibility to evaluate the adequacy of alternative models for the same set of data. EFA is predominately used in the development of instruments, and CFA in the refinement and evaluation.

Both methods require many choices, decisions and judgments on behalf of the researcher. For example, as to the meaning of factors, the number of substantial factors, the relations between factors, and so forth. Even though the methods have become increasingly sophisticated and complex, the interpretation of results rests on human judgment.

Additional statistical methods

In addition to the methods of classical test theory and factor analysis various other techniques were used. Descriptive data for background variables of participants were analyzed with *t*-tests, analysis of variance (ANOVA) and χ^2

– tests. Bivariate relationships were investigated with the Pearson product-moment correlation coefficient (r) and Spearman's rank correlation coefficient (r_s). The partial correlation coefficient was used when statistically controlling for the influence of other variables on bivariate relationships. Furthermore, multiple regression analysis was used.

9. Summary of the studies

This section summarizes the aims, methods and results for the studies included in the thesis. Studies I and II are empirically oriented and provide a background, results and conclusions that have guided instrument development and validation that is the focus of Studies III and IV. The questions that the results from Study I raised, together with a literature review, may be considered the starting point for the whole research agenda.

Study I: Predictors of Self-reported Prospective and Retrospective Memory in a Population-based Sample of Older Adults

The background for this study focused on self-reports of memory in clinical settings as possible influences on the diagnoses and choices of treatments. The rationale was that it is therefore important to determine what individual differences in self-reported memory reflect, especially possible actual deficits in memory ability. The aim was to investigate predictors of scores on an established self-report measure for frequency of everyday memory failures - the Prospective and Retrospective Questionnaire (PRMQ, Smith et al., 2000). The participants constituted a population-based sample of 250 older adults, aged 60-90 years. Multiple regression analyses were performed with depressive symptoms, personality traits, and objective memory ability as the candidate predictors. Age, albeit being a highly relevant variable, was not included as a predictor because it was not significantly correlated with PRMQ-scores. The results showed that a higher frequency of reported memory failures (higher scores) was predicted by lower scores on the personality dimension of Self-Directedness as assessed by the Temperament and Character Inventory (TCI; Cloninger et al., 1993) and more depressive symptoms on the Center for Epidemiological Studies Depression Scale (CES-D; Radloff, 1977). However, PRMQ scores were not predicted by objective memory ability on a series of retrospective memory measures and one measure of prospective memory. Finally, univariate analyses of variance was performed to investigate whether scores on the PRMQ differed according to level of cognitive functioning (low, average and high) as measured by the Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1977). Mean scores on neither of the PRMQ subscales differed between the

three groups. In conclusion, the results indicated that within the older population, self-reported memory as measured by the PRMQ may reflect mood-state and personality factors rather than individual differences in memory and cognitive ability.

Study II: Perceived Causes of Everyday Memory Problems in a Population Based Sample Aged 39 – 99

Research studies usually find weak relationships between measures of memory complaints and laboratory memory performance. Furthermore, many research articles claim that memory complaints are common in the elderly or increase with age. Given the finding of a lack of a relationship between age and scores on the PRMQ in study I, the focus of Study II was on a global rating of *problems with memory*. This was considered the best way to operationalize the concept of memory complaints. The participants also reported what they perceived as causes of possible problems with memory, based on an open ended question. The prevalence, severity and perceived causes of the memory problems were investigated in a population-based sample of 361 adults ranging in age from 39 to 99 years. Nearly one third (30.2 %) of the participants reported memory complaints (a rating of at least moderate memory problems). In contrast to the results from study I, problems ratings were related to age, such that increasing age meant higher severity of problems, but the age-related differences were nevertheless small. The most frequent perceived causes were age/ageing, stress and multitasking (often formulated as “keeping many balls in the air”). Importantly, age/ageing as a cause was more frequent among older participants, and stress and multitasking were more frequent among middle-aged participants. The results suggested that everyday stress and level of engagement in multiple tasks or commitments, that place demands on cognitive resources, are important variables to consider when studying the relations between subjective everyday memory measures, age and memory performance on laboratory tests.

Study III: Development of the Cognitive Dysfunction Questionnaire in a Population-based Sample

With the results from Studies I and II as a background, the purpose of Study III was to develop a new questionnaire with the explicit aim that scores should be related to laboratory test performance. A thorough theoretical and methodological analysis was performed and a global construct of cognitive functioning was chosen for the new instrument. The main theoretical framework was that of memory systems (working memory, semantic memory, episodic memory, procedural memory and the perceptual representation system). Other cognitive domains were also included, such as temporal orientation and spatial navigation. Then, items were constructed

for each of the domains, resulting in a pool of 90. The items were included in a questionnaire that was sent to previous participants in the Betula Prospective Cohort Study. Background variables and cognitive test performance data were available for the participants, of which 370 (70 %) returned the questionnaire by mail, resulting in a sample with a mean age of 65 ± 15 years. A series of principal components analyses (exploratory factor analyses) were performed on the responses. Component scores were computed and correlated with age and cognitive test performance (MMSE, Block Design, semantic/episodic memory) that served as criteria for item selection. Twenty items loading on six positively correlated components constituted the final selection for the questionnaire that was named *the Cognitive Dysfunction Questionnaire (CDQ)*. The components were named *Procedural actions, Semantic word knowledge, Face recognition, Temporal orientation, Episodic memory for places, and Spatial navigation*. The final scale was internally consistent as reflected in a Cronbach's α of .90 for the total score. Evidence of construct validity was obtained by significant correlations with objective cognitive measures, and a subjective memory measure (PRMQ). Discriminant validity evidence was obtained by a low and non-significant correlation with depressive symptoms. Further evidence of construct validity was provided by correlations with age and educational attainment. The conclusion was that the CDQ is promising as a self-rating screening tool for cognitive dysfunction, and that it will be the subject of further development and validation.

Study IV: Multigroup Confirmatory Factor Analysis of the Cognitive Dysfunction Questionnaire: Instrument Refinement and Measurement Invariance across Age and Sex

For Study IV a total of twenty new items in the domains were constructed, in addition to the previous twenty. The new 40-item version was included in the fifth wave of data collection (T5) in the Betula study. The aims of Study IV were to refine the CDQ into a final version, to cross-validate its factor structure, test competing models, and to investigate measurement invariance across age and sex. The analyses were based on data for a total of 1115 participants (mean age: 63.0 ± 14.5 years, range: 25 - 95) randomly split into a refinement ($n = 569$) and a cross-validation ($n = 546$) sample. Sample A was used for the refinement of the CDQ into a 20-item instrument with 5 domains. To achieve this, confirmatory factor analysis was performed and items that contributed to model misfit were excluded. The final model for the factor structure of the CDQ consisted of five first-order factors and a second order general factor. This model was cross-validated by means of multigroup confirmatory factor analysis in samples A and B. The results from the analyses supported the equivalence of the model across the samples. Among

four competing models, all variants differing in parsimony, a correlated five factor model was found to provide the best statistical fit. However, the model with five first-order factors and one second-order factor was more parsimonious and comparable in fit, and was therefore selected as the final representation of the factor structure of the CDQ.

The analyses of measurement invariance were performed through a sequence of hierarchically nested models with increasing cross-group constraints on parameters. The results suggested that measurement invariance across age and sex was plausible up to the scalar level. This indicates that the same construct is measured across groups, and that the origins and units of measurement are the same. This is a prerequisite for studying relations between constructs and comparing group means. The final version of the CDQ consisted of 20 items in five domains (Procedural actions, Semantic word knowledge, Face recognition, Temporal orientation and Spatial navigation). The total score and subscale scores were found to be internally consistent with Cronbach's α ranging from .67 to .89. Finally, some evidence of construct validity was provided by showing that the CDQ total score was higher in older compared to younger groups, and higher in groups with higher levels of cognitive dysfunction as measured by the Minimal State Examination. The conclusion was that adoption of the final version of the CDQ appears to be a way of measuring cognitive dysfunction without administering formal cognitive tests. Furthermore, that future studies should apply it among clinical groups to further test its usefulness

10. Discussion

The first aim of this thesis was to develop a questionnaire for cognitive functioning. The rationale was that it could possibly be used as a screening instrument for early signs of AD in the future. This aim has been fulfilled insofar that the studies in the thesis describe the development process for the Cognitive Dysfunction Questionnaire. Furthermore, the results presented suggest that the instrument has two major prerequisites for showing utility as a screening instrument – adequate reliability and a relationship of the scores with scores from tests/instruments known to be sensitive to cognitive dysfunction. However, this evidence is not enough to claim clinical utility. Further evidence is needed in future studies.

The second aim of the thesis was to address theoretical and methodological issues when developing this type of instruments. This resulted in an attempt to discuss and explicate what is meant by the subjective–objective distinction, and an outline of a theoretical framework in regard to development and validation of instruments viewed from the subjective–objective perspective.

Main results and conclusions from the studies

Overall, the results from the studies have indicated that extreme care should be taken in the development of self-report measures of memory and cognition. As the results from Study I suggests, depressive symptoms and personality traits appear to affect responses to questions or statements about memory. However, these results can either be interpreted as, (1) an empirical finding as to the correlates of certain predominately verbal behaviors (e.g., memory complaints), (2) evidence of relationships between constructs (e.g., the relationship between personality traits and memory beliefs), (3) and/or an indication of construct irrelevant variance threatening validity (e.g., cognitive functioning as defined by objective tests). The interpretations are dependent upon the theoretical rationale and degree of construct explication.

The further finding in Study I of a lack of relationship between age and scores on the PRMQ is certainly at odds with much of the literature regarding memory complaints, where it is almost universally claimed that these are common in the elderly or increase with age (e.g., Geerlings, Jonker, Bouter, Ader, & Schmand, 1999; Jonker, et al., 2000; Kliegel & Zimprich, 2005; Reid & MacLulich, 2006). However, as I have discussed at length, memory complaints is an ill defined concept, of questionable status as a construct, and with clear limits of application. It is highly questionable whether scores on the PRMQ should be considered indicative of memory complaints. Support for this conclusion comes from Study II, where the single item “Do You experience problems with your memory?” was used. This was considered the best way to measure the concept. Indeed, there was a positive relationship between severity of memory problems, that is, memory complaints, and age. However, the relation was far from strong, so it raises some questions as to what “common in the elderly” means. Once again, it should be noted that an operationalization of memory complaints is difficult due to a lack of adequate definition. Furthermore, it is not clear whether the claim “common in the elderly” refers to the subset of individuals from the general population that seek medical consultation, the elderly as a clinical population compared to other clinical populations, or in the general population regardless of whether they seek medical consultation or not. In case of the first two, the results from Study II are not applicable, and the claim may indeed be valid.

The development of the CDQ as described in Study III posed many challenges. The main theoretical framework, that of memory systems, is essentially an objective construct. Hence, the theory is mainly based on inferences from observation of behavior under controlled laboratory conditions, in both healthy and brain damaged individuals. A major challenge was to translate theory into observable everyday events. Judging from the results of study III, the strategy was successful to some extent.

However, this conclusion rests on the assumption that the objective tests in the study measure what they purport to measure, that is, they are valid indicators of a global construct of memory/cognitive functioning. This assumption does indeed appear to be well-grounded given converging results from a bulk of studies utilizing similar or the same measures.

The results from Studies III and IV with the CDQ should be contrasted with the results from Studies I and III with the PRMQ. The results from Study I lent no support for the PRMQ as a measure of memory or cognitive functioning (objective constructs) in a sample of older adults (60-90 years old). We found no relation to objective memory/cognitive test performance, and no relation to age. Essentially the same results can be found in Study III (in a larger sample with a wider age range), albeit not explicated in the article (see table 2); the PRMQ was not significantly correlated with any objective measure or age. Hence, the findings of significant relations between the CDQ, objective memory/cognitive test performance and age, lend support for the CDQ as a measure of memory and cognitive functioning (an objective construct).

Limitations

Regarding the composition and factor structure of the CDQ, the memory systems model that served as the main theoretical framework is not easily identified among the factors. This is a limitation from the perspective of interpreting the factors. It is possible that the questions in the CDQ were not sufficiently specific to identify and delineate the different systems. From a different perspective it may be that people just do not have the ability of differentiating between memory failures of different origins (i.e., a semantic or episodic failure). Furthermore, it is likely that the memory systems share common processes, especially working, episodic and semantic memory (see, e.g., Nyberg, Marklund, et al., 2003), possibly making it much harder to identify the type of failure.

The five factors in the CDQ are all at least moderately correlated; alternatively there is a strong general second-order factor. From the view of factor analyses of various batteries of objective cognitive test, this is a common finding (see, e.g., Carroll, 1993). Given this strong general factor it can be questioned whether the factors contribute uniquely. The analyses of competing models in Study IV, suggest that they do. The single-factor model did not provide a good fit to the data.

Finally, it should be acknowledged that the common variance of objective memory tests and the CDQ is small. There is likely systematic construct irrelevant variance in the scores. As such it is a fallible measure among others.

The approach taken for the development of the CDQ was that the same construct should be adopted for subjective and objective measures if validity

is to be evaluated by correlations between them. However, it may be that “subjective memory/cognition” should have the status of a subjective construct in its’ own right. This notion could be supported by the likelihood that the best correlate of any subjective memory/cognitive measure is any other subjective memory/cognitive measure. In that case, the theoretical framework surrounding the construct should dictate validation studies.

Future research

Future research on subjective memory/cognition in general should address the role of stress and cognitive load, particularly in relation to age. As Study II indicated, people of different ages perceive different causes for “memory problems”. Furthermore, the levels of stress and cognitive load are likely related to age in a similar manner to memory ability, that is, decreasing with increasing age. High levels of stress and cognitive load will likely impair cognitive performance in a given situation. There will however be individual differences in exposure to such high stress and cognitive load situations in everyday life. Inferring cognitive ability or dysfunction as defined by objective cognitive tests will be difficult without control for these extraneous variables. A study that includes measures of subjective memory/cognition and perceived stress and cognitive load, will enable addressing these issues.

Another issue that should be explored is the possibility of common method bias, that is, variance that is due to the measurement method rather than to the constructs the measures represent (e.g., Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Self-report measures may be particularly vulnerable to response biases, for example, social desirability, acquiescence, leniency effects, or tendencies to yes- or no-saying. The consequence is that relations between different constructs measured with the same method may be interpreted as “real”, when they are in fact caused by the method. These potential common method biases are likely present in the measurement of both subjective and objective constructs with subjective methods. However, it may be particularly challenging to infer them for subjective constructs when no objective response indicators exist.

A topic for future research that has a bearing on many of the issues in this thesis is an alternative strategy for developing potential screening instruments such as the CDQ. This strategy could be even more oriented at prediction, and less concerned with theoretical constructs and content. Such a strategy would perhaps be referred to as “blind empiricism” in a derogatory manner by some. However, I think that in the case of such devastating illnesses as dementia, the end should justify the means. The strategy could involve assembly of an item pool from various established self-report measures of memory and cognition, and perhaps items from other relevant domains (e.g., depression and stress). The selection of items should then be based on maximal predictive power in regard to a criterion, such as a future

dementia diagnosis. There are test development strategies focused on prediction, and predictive validation. However, if one adopts the unitary view of validity and the focus on construct validity, such a strategy appears dubious. Once again, the validity concept has become both too broad and too narrow. In my view, revision of the definition of the validity concept, introduction of an alternative term, such as utility, and additional theoretical work is needed.

Future research regarding the CDQ will be mainly concerned with validation and studies on clinical utility. Further evidence of construct validity is needed. A particularly important future study will adopt structural equation modeling to investigate the relation between scores from the CDQ and objective memory/cognitive functioning (convergent evidence), other subjective cognitive measures (method effects), depressive symptoms (discriminant evidence) et cetera.

Furthermore, the question of whether the CDQ can be used as a screening instrument for early signs of AD needs to be addressed. This could be considered part of providing concurrent and predictive evidence for construct validity. However, there is an additional issue as to what accuracy and precision can be obtained. It may well be that concurrent and predictive evidence in support of validity is obtained, but that the accuracy and precision is not sufficient for clinical purposes.

Moreover, studies should investigate whether the CDQ has utility in various clinical populations where cognitive dysfunction is a common denominator. Finally, if the CDQ will be used for clinical assessment purposes as a basis for diagnosis or decision-making, additional issues such as consequences of its use may be addressed.

Concluding remarks

The subjective–objective distinction and subjective memory in particular, are certainly challenging topics. The subjective–objective distinction in a way defies what has been, and still should be a fundament of science; objectivity. Many issues remain to be addressed, and maybe we have never actually known what the question is.

References

- Adams, D. (1997). *The Hitchhiker's Guide to the Galaxy*: Random House Publishing Group.
- Allen, P. A., Hall, R. J., Druley, J. A., Smith, A. F., Sanders, R. E., & Murphy, M. D. (2001). How shared are age-related influences on cognitive and noncognitive variables? *Psychology and Aging, 16*(3), 532-549. doi: 10.1037/0882-7974.16.3.532
- American Educational Research Association [AERA]; American Psychological Association [APA]; & National Council on Measurement in Education [NCME]. (1999). *Standards for educational and psychological testing*. Washington DC: American Educational Research Association.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th., text revision ed.). Washington, DC: Author.
- Anastasi, A. (1950). The concept of validity in the interpretation of test scores. *Educational and Psychological Measurement, 10*, 67-78. doi: 10.1177/001316445001000105
- Anastasi, A., & Urbina, S. (1997). *Psychological Testing* (7th International ed.). New Jersey: Prentice-Hall International.
- Annett, J. (2002). Subjective rating scales: Science or art? *Ergonomics, 45*(14), 966-987. doi: 10.1080/00140130210166951
- Armitage, R., Trivedi, M., Hoffmann, R., & Rush, A. J. (1997). Relationship between objective and subjective sleep measures in depressed patients and healthy controls. *Depression and Anxiety, 5*(2), 97-102. doi: 10.1002/(SICI)1520-6394(1997)5:2<97::AID-DA6>3.0.CO;2-2
- Baltes, P. B., & Lindenberger, U. (1997). Emergence of a powerful connection between sensory and cognitive functions across the adult life span: A new window to the study of cognitive aging? *Psychology and Aging, 12*(1), 12-21. doi: 10.1037/0882-7974.12.1.12
- Bassett, S. S., & Folstein, M. F. (1993). Memory complaint, memory performance, and psychiatric diagnosis: A community study. *Journal of Geriatric Psychiatry and Neurology, 6*(2), 105-111. doi: 10.1177/089198879300600207
- Bechtoldt, H. P. (1959). Construct validity: A critique. *American Psychologist, 14*(10), 619-629. doi: 10.1037/h0040359
- Bolla, K. I., Lindgren, K. N., Bonaccorsy, C., & Bleecker, M. L. (1991). Memory complaints in older adults: Fact or fiction? *Archives of Neurology, 48*(1), 61-64. doi: 10.1001/archneur.1991.00530130069022
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111*(4), 1061-1071. doi: 2004-19012-010 [pii], 10.1037/0033-295X.111.4.1061
- Bower, G. H. (2000). A brief history of memory research. In E. Tulvin & F. I. M. Craik (Eds.), *The oxford handbook of memory* (pp. 3-32). New York: Oxford University Press, Inc.

- Broadbent, D. E., Cooper, P. F., FitzGerald, P., & Parkes, K. R. (1982). The Cognitive Failures Questionnaire (CFQ) and its correlates. *The British Journal of Clinical Psychology* 21 (Pt 1), 1-16.
- Burgess, P. W., & Shallice, T. (1997). The relationship between prospective and retrospective memory: Neuropsychological evidence. In M. Conway (Ed.), *Cognitive models of memory* (1st MIT Press ed., pp. 247-272). Guildford and King's Lynn: MIT Press.
- Byrne, M. D. (1998). Taking a computational approach to aging: The SPAN theory of working memory. *Psychology and Aging*, 13(2), 309-322. doi: 10.1037/0882-7974.13.2.309
- Bäckman, L., Jones, S., Berger, A. K., Laukka, E. J., & Small, B. J. (2005). Cognitive impairment in preclinical Alzheimer's disease: A meta-analysis. *Neuropsychology*, 19(4), 520-531. doi: 2005-08223-013 [pii] 10.1037/0894-4105.19.4.520
- Bäckman, L., Small, B. J., & Fratiglioni, L. (2001). Stability of the preclinical episodic memory deficit in Alzheimer's disease. *Brain*, 124(1), 96-102. doi: 10.1093/brain/124.1.96
- Bäckman, L., Small, B. J., & Wahlin, Å. (2001). Aging and memory: Cognitive and biological perspectives. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (5th ed., pp. 349-377). San Diego, CA: Academic Press.
- Carroll, J. B. (1993). *Human Cognitive Abilities*. Cambridge: Cambridge Univ. Press.
- Chakravartty, A. (2011). "Scientific Realism". *The Stanford Encyclopedia of Philosophy* (Summer 2011 Edition), Edward N. Zalta (ed.). from <http://plato.stanford.edu/archives/sum2011/entries/scientific-realism/>
- Choppin, B. H. (1997). Objective tests. In J. Keeves (Ed.), *Educational research, methodology and measurement: an international handbook* (2nd ed.): Elsevier Science Ltd.
- Cizek, G. J., Bowen, D., & Church, K. (2010). Sources of validity evidence for educational and psychological tests: A follow-up study. *Educational and Psychological Measurement*, 70(5), 732-743. doi: 10.1177/0013164410379323
- Cizek, G. J., Rosenberg, S. L., & Koons, H. H. (2008). Sources of validity evidence for educational and psychological tests. *Educational and Psychological Measurement*, 68(3), 397-412. doi: 10.1177/0013164407310130
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7(3), 309-319. doi: 10.1037/1040-3590.7.3.309
- Clarnette, R. M., Almeida, O. P., Forstl, H., Paton, A., & Martins, R. N. (2001). Clinical characteristics of individuals with subjective memory loss in Western Australia: Results from a cross-sectional survey. *International Journal of Geriatric Psychiatry*, 16(2), 168-174. doi: 10.1002/1099-1166(200102)16:2<168::AID-GPS291>3.0.CO;2-D

- Cleary, P. D. (1997). Subjective and objective measures of health: Which is better when? *Journal of Health Services Research & Policy*, 2(1), 3.
- Cole, G., Tucker, L., & Friedman, G. (1986). Measures of objective and subjective stress by level of income. *Psychological Reports*, 59(1), 139-142.
- Cook, S., & Marsiske, M. (2006). Subjective memory beliefs and cognitive performance in normal and mildly impaired older adults. *Aging & Mental Health*, 10(4), 413-423. doi: V614N6477K182272 [pii], 10.1080/13607860600638487
- Craik, F. I. M. (2000). Age-related changes in human memory. In D. C. Park & N. Schwarz (Eds.), *Cognitive Aging: A Primer*. (pp. 75-92). New York: Psychology Press.
- Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. New York: Holt, Rinehart & Winston.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin*, 52(4), 281-302. doi: 10.1037/h0040957
- Danna, K., & Griffin, R. W. (1999). Health and well-being in the workplace: A review and synthesis of the literature. *Journal of Management*, 25(3), 357-384. doi: 10.1177/014920639902500305
- Díez, J. A. (1997a). A hundred years of numbers. An historical introduction to measurement theory 1887-1990. Part I: The formation period. Two lines of research: Axiomatics and real morphisms, scales and invariance. *Studies in History and Philosophy of Science*, 28(1), 167-185. doi: 10.1016/S0039-3681(96)00014-3
- Díez, J. A. (1997b). A hundred years of numbers. An historical introduction to measurement theory 1887-1990. Part II: Suppes and the mature theory. Representation and uniqueness. *Studies in History and Philosophy of Science*, 28(2), 237-265. doi: 10.1016/S0039-3681(96)00015-5
- Dixon, R. A., Hultsch, D. F., & Hertzog, C. (1988). The Metamemory in Adulthood (MIA) questionnaire. *Psychopharmacology Bulletin*, 24(4), 671-688.
- Dobbs, A. R., & Rule, B. G. (1989). Adult age differences in working memory. *Psychology and Aging*, 4(4), 500-503. doi: 10.1037/0882-7974.4.4.500
- Eysenck, M. W., & Keane, M. T. (2005). Learning and memory. *Cognitive psychology: A student's handbook* (5th ed.). New York: Psychology Press.
- Gauthier, S., Reisberg, B., Zaudig, M., Petersen, R. C., Ritchie, K., Broich, K., et al. (2006). Mild cognitive impairment. *The Lancet*, 367(9518), 1262-1270. doi: 10.1016/S0140-6736(06)68542-5
- Geerlings, M. I., Jonker, C., Bouter, L. M., Ader, H. J., & Schmand, B. (1999). Association between memory complaints and incident Alzheimer's disease in elderly people with normal baseline cognition. *American Journal of Psychiatry*, 156(4), 531-537.
- Gilewski, M. J., & Zelinski, E. M. (1986). Questionnaire assessment of memory complaints. In L. W. Poon (Ed.), *Handbook for clinical memory*

- assessment of older adults*. Washington: American Psychological Association.
- Goldiamond, I. (1958). Indicators of perception. I. Subliminal perception, subception, unconscious perception: An analysis in terms of psychophysical indicator methodology. *Psychological Bulletin*, *55*(6), 373-411. doi: 10.1037/h0046992
- Guarch, J., Marcos, T., Salamero, M., & Blesa, R. (2004). Neuropsychological markers of dementia in patients with memory complaints. *International Journal of Geriatric Psychiatry*, *19*(4), 352-358. doi: 10.1002/gps.1074
- Herrmann, D. J. (1982). Know thy memory: The use of questionnaires to assess and study memory. *Psychological Bulletin*, *92*(2), 434-452. doi: 10.1037/0033-2909.92.2.434
- Herrmann, D. J., Sheets, V., Gruneberg, M., & Torres, R. (2005). Are self reports of memory failure accurate? *Applied Cognitive Psychology*, *19*(7), 821-841. doi: 10.1002/acp.1125
- Hertzog, C., & Dixon, R. A. (1994). Metacognitive development in adulthood and old age. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: knowing about knowing* (pp. 227-251). Cambridge, MA: MIT Press.
- Hertzog, C., Park, D. C., Morrell, R. W., & Martin, M. (2000). Ask and ye shall receive: Behavioural specificity in the accuracy of subjective memory complaints. *Applied Cognitive Psychology*, *14*(3), 257-275. doi: 10.1002/(SICI)1099-0720(200005/06)14:3<257::AID-ACP651>3.0.CO;2-O
- Hindmarch, I., & Parrott, A. (1977). Repeated dose comparison of nomifensine, imipramine and placebo on subjective assessments of sleep and objective measures of psychomotor performance. *British Journal of Clinical Pharmacology*, *4*(Suppl 2), 167S-173S.
- Howard, G. S. (1994). Why do people say nasty things about self-reports? *Journal of Organizational Behavior*, *15*(5), 399-404. doi: 10.1002/job.4030150505
- Howsepian, A. A. (2008). The DSM-IV-TR 'Glossary of Technical Terms': A reappraisal. *Psychopathology*, *41*(1), 28-34. doi: 10.1159/000109952
- Hughes, C., Berg, L., Danziger, W., Coben, L., & Martin, R. (1982). A new clinical scale for the staging of dementia. *The British Journal of Psychiatry*, *140*(6), 566-572. doi: 10.1192/bjp.140.6.566
- Hänninen, T., Reinikainen, K. J., Helkala, E. L., Koivisto, K., Mykkänen, L., Laakso, M., et al. (1994). Subjective memory complaints and personality traits in normal elderly subjects. *Journal of the American Geriatrics Society*, *42*(1), 1-4.
- Johnston, D. W., Propper, C., & Shields, M. A. (2009). Comparing subjective and objective measures of health: Evidence from hypertension for the income/health gradient. *Journal of Health Economics*, *28*(3), 540-552. doi: 10.1016/j.jhealeco.2009.02.010
- Jonker, C., Geerlings, M. I., & Schmaud, B. (2000). Are memory complaints predictive for dementia? A review of clinical and population-based

- studies. *International Journal of Geriatric Psychiatry*, 15(11), 983-991. doi: 10.1002/1099-1166(200011)15:11<983::AID-GPS238>3.0.CO;2-5
- Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, 38(4), 319-342. doi: 10.1111/j.1745-3984.2001.tb01130.x
- Kelley, T. L. (1927). *Interpretation of educational measurements*. New York: McMillan.
- Kendell, R., & Jablensky, A. (2003). Distinguishing between the validity and utility of psychiatric diagnoses. *American Journal of Psychiatry*, 160(1), 4-12. doi: 10.1176/appi.ajp.160.1.4
- Kerlinger, F. N. (1964). *Foundations of behavioral research - educational and psychological enquiry*. New York: Holt, Rinehart and Winston, Inc.
- Kliegel, M., & Zimprich, D. (2005). Predictors of cognitive complaints in older adults: a mixture regression approach. *European Journal of Ageing*, 2(1), 13-23. doi: 10.1007/s10433-005-0017-6
- Kline, P. (2000). *Handbook of psychological testing* (2nd ed.). London: Routledge.
- Lezak, M. D. (2004). *Neuropsychological assessment* (4th ed.): Oxford University Press.
- Lincoln, N. B., & Tinson, D. J. (1989). The relation between subjective and objective memory impairment after stroke. *The British Journal of Clinical Psychology*, 28(1), 61-65.
- Lissitz, R. W., & Samuelsen, K. (2007). A suggested change in terminology and emphasis regarding validity and education. *Educational Researcher*, 36(8), 437-448. doi: 10.3102/0013189x07311286
- Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, 3(Suppl 9), 635-694. doi: 10.2466/PRO.3.7.635-694
- MacCorquodale, K., & Meehl, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review*, 55(2), 95-107. doi: 10.1037/h0056029
- Mari, L. (2003). Epistemology of measurement. *Measurement*, 34(1), 17-30. doi: 10.1016/S0263-2241(03)00016-2
- McDonald-Miszczak, L., Gould, O. N., & Tychynski, D. (1999). Metamemory predictors of prospective and retrospective memory performance. *Journal of General Psychology*, 126(1), 37-52. doi: 10.1080/00221309909595350
- McDonald, R. P. (1999). *Test theory. A unified treatment*. Hillsdale: Lawrence Erlbaum Associates, Inc.
- Mehrens, W. A. (1997). The consequences of consequential validity. *Educational Measurement: Issues and Practice*, 16(2), 16-18. doi: 10.1111/j.1745-3992.1997.tb00588.x
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: American Council on Education/Macmillian Publishers.

- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741-749. doi: 10.1037/0003-066x.50.9.741
- Michell, J. (1999). *Measurement in psychology: Critical history of a methodological concept*: Cambridge University Press.
- Mitchell, A. J. (2008). Is it time to separate subjective cognitive complaints from the diagnosis of mild cognitive impairment? *Age and Ageing*, *37*(5), 497-499. doi: 10.1093/ageing/afn147
- Mol, M. E. M., van Boxtel, M. P. J., Willems, D., & Jolles, J. (2006). Do subjective memory complaints predict cognitive dysfunction over time? A 6-year follow-up of the Maastricht Aging Study. *International Journal of Geriatric Psychiatry*, *21*(5), 432-441. doi: 10.1002/gps.1487
- Moore, J. (1995). Radical behaviorism and the subjective-objective distinction. *The Behavior Analyst*, *18*(1), 33-49.
- Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures: Issues and applications*: Sage Publications, Inc.
- Nilsson, L.-G. (2003). Memory function in normal aging. *Acta Neurologica Scandinavica*, *107*, 7-13. doi: 10.1034/j.1600-0404.107.s179.5.x
- Nilsson, L.-G., Adolfsson, R., Bäckman, L., de Frias, C. M., Molander, B., & Nyberg, L. (2004). Betula: A prospective cohort study on memory, health and aging. *Aging, Neuropsychology, and Cognition*, *11*(2-3), 134-148. doi: 10.1080/13825580490511026
- Nilsson, L.-G., Adolfsson, R., Bäckman, L., Molander, B., Norin, E., Nyberg, L., et al. (2000). The Betula Study: Objectives, assessment and design. *International Journal of Psychology*, *35*(3-4), 234-234.
- Nilsson, L.-G., Bäckman, L., Erngrund, K., Nyberg, L., Adolfsson, R., Bucht, G., et al. (1997). The Betula prospective cohort study: Memory, health and aging. *Aging, Neuropsychology, and Cognition*, *4*(1), 1-32. doi: 10.1080/13825589708256633
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill, inc.
- Nyberg, L., Maitland, S. B., Rönnlund, M., Bäckman, L., Dixon, R. A., Wahlin, Å., et al. (2003). Selective adult age differences in an age-invariant multifactor model of declarative memory. *Psychology and Aging*, *18*(1), 149-160. doi: 10.1037/0882-7974.18.1.149
- Nyberg, L., Marklund, P., Persson, J., Cabeza, R., Forkstam, C., Petersson, K. M., et al. (2003). Common prefrontal activations during working memory, episodic memory, and semantic memory. *Neuropsychologia*, *41*(3), 371-377. doi: 10.1016/S0028-3932(02)00168-9
- paradigm. (2011). In *Merriam-Webster Online Dictionary*. Retrieved May 12, 2011, from <http://www.merriam-webster.com/dictionary/paradigm>
- Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging*, *17*(2), 299-320. doi: 10.1037/0882-7974.17.2.299

- Pearman, A., & Storandt, M. (2004). Predictors of subjective memory in older adults. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 59(1), 4-6. doi: 10.1093/geronb/59.1.P4
- Pearman, A., & Storandt, M. (2005). Self-discipline and self-consciousness predict subjective memory in older adults. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 60(3), 153-157. doi: 10.1093/geronb/60.3.P153
- Perez Garcia, M., Godoy Garcia, J. F., Vera Guerrero, N., Laserna Triguero, J. A., & Puente, A. E. (1998). Neuropsychological evaluation of everyday memory. *Neuropsychology Review*, 8(4), 203-227. doi: 10.1023/A:1021622319851
- Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, 256(3), 183-194. doi: 10.1111/j.1365-2796.2004.01388.x
- Podewils, L. J., McLay, R. N., Rebok, G. W., & Lyketsos, C. G. (2003). Relationship of self-perceptions of memory and worry to objective measures of memory and cognition in the general population. *Psychosomatics*, 44(6), 461-470.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879-903. doi: 10.1037/0021-9010.88.5.879
- Ponds, R. W. H. M., & Jolles, J. (1996). Memory complaints in elderly people: the role of memory abilities, metamemory, depression, and personality. *Educational Gerontology*, 22(4), 341 - 357. doi: 10.1080/0360127960220404
- Popham, W. (1997). Consequential validity: Right concern wrong concept. *Educational Measurement: Issues and Practice*, 16(2), 9-13. doi: 10.1111/j.1745-3992.1997.tb00586.x
- Reid, L. M., & MacLulich, A. M. J. (2006). Subjective memory complaints and cognitive impairment in older people. *Dementia and Geriatric Cognitive Disorders*, 22(5-6), 471-485. doi: 10.1159/000096295
- Reisberg, B., Ferris, S., de Leon, M., & Crook, T. (1982). The Global Deterioration Scale for assessment of primary degenerative dementia. *American Journal of Psychiatry*, 139(9), 1136-1139.
- Roediger, H. L., III, Marsh, E. J., & Lee, S. C. (2002). Kinds of memory. In H. L. Roediger, III, E. J. Marsh, S. C. Lee, H. Pashler & D. Medin (Eds.), *Steven's handbook of experimental psychology, 2: Memory and cognitive processes* (3 ed., Vol. 2, pp. 1-41). Hoboken, NJ: John Wiley & Sons Inc.
- Rothstein, J. M. (1989). On defining subjective and objective measurements. *Physical Therapy*, 69(7), 577-579.
- Rönnlund, M., Mäntylä, T., & Nilsson, L.-G. (2008). The Prospective and Retrospective Memory Questionnaire (PRMQ): Factorial structure, relations to global subjective memory ratings, and Swedish norms.

- Scandinavian Journal of Psychology*, 49(1), 11-18. doi: 10.1111/j.1467-9450.2007.00600.x
- Rönnlund, M., Nyberg, L., Bäckman, L., & Nilsson, L.-G. (2005). Stability, growth, and decline in adult life span development of declarative memory: Cross-sectional and longitudinal data from a population-based study. *Psychology and Aging*, 20(1), 3-18.
- Salthouse, T. A. (1994). The aging of working memory. *Neuropsychology*, 8(4), 535-543. doi: 10.1037/0894-4105.8.4.535
- Salthouse, T. A., & Ferrer-Caja, E. (2003). What needs to be explained to account for age-related effects on multiple cognitive variables? *Psychology and Aging*, 18(1), 91-110. doi: 10.1037/0882-7974.18.1.91
- Schacter, D. L., & Tulving, E. (1994). What are the memory systems of 1994? In D. L. Schacter & E. Tulving (Eds.), *Memory systems 1994*. Cambridge, MA: The MIT Press.
- Schacter, D. L., Wagner, A. D., & Buckner, R. L. (2000). Memory systems of 1999. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 627-643). New York: Oxford University Press, Inc.
- Scott, W. A., & Johnson, R. C. (1972). Comparative validities of direct and indirect personality tests. *Journal of Consulting and Clinical Psychology*, 38(3), 301-318. doi: 10.1037/h0032897
- Sechrest, L. (2005). Validity of measures is no simple matter. *Health Services Research*, 40(5p2), 1584-1604. doi: 10.1111/j.1475-6773.2005.00443.x
- Sedlacek, W. E. (2004). *Beyond the big test: Noncognitive assessment in higher education* (1st ed.). San Francisco: Jossey-Bass Publishers.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York: Houghton Mifflin.
- Shuttleworth, M. (2008). What is a Paradigm? Retrieved February 15, 2011, from <http://www.experiment-resources.com/what-is-a-paradigm.html>
- Smith, G., Della Sala, S., Logie, R. H., & Maylor, E. A. (2000). Prospective and retrospective memory in normal ageing and dementia: A questionnaire study. *Memory*, 8(5), 311-321. doi: 10.1080/09658210050117735
- Smith, G. E., Petersen, R. C., Ivnik, R. J., Malec, J. F., & Tangalos, E. G. (1996). Subjective memory complaints, psychological distress, and longitudinal change in objective memory performance. *Psychology and Aging*, 11(2), 272-279.
- Solomon, Z., Mikulincer, M., & Hobfoll, S. E. (1987). Objective versus subjective measurement of stress and social support: Combat-related reactions. *Journal of Consulting and Clinical Psychology*, 55(4), 577-583. doi: 10.1037/0022-006x.55.4.577
- Spector, P. E. (1994). Using self-report questionnaires in OB research: A comment on the use of a controversial method. *Journal of Organizational Behavior*, 15(5), 385-392. doi: 10.1002/job.4030150503

- Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory*, 82(3), 171-177. doi: 10.1016/j.nlm.2004.06.005
- Squire, L. R., Knowlton, B., & Musen, G. (1993). The structure and organization of memory. *Annual Review of Psychology*, 44(1), 453-495. doi: 10.1146/annurev.ps.44.020193.002321
- Sunderland, A. (1983). Do laboratory tests predict everyday memory? A neuropsychological study. *Journal of Verbal Learning and Verbal Behavior*, 22(3), 341-357. doi: 10.1016/S0022-5371(83)90229-3
- Thompson, B., & Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement*, 56(2), 197-208. doi: 10.1177/0013164496056002001
- Tulving, E. (2000). Concepts of memory. In E. Tulving & F. I. M. Craik (Eds.), *The oxford handbook of memory* (pp. 33-43). New York: Oxford University Press, Inc.
- Urbina, S. (2004). *Essentials of psychological testing*. Hoboken, NJ: John Wiley & Sons, Inc
- Valentijn, S. A. M., van Hooren, S. A. H., Bosma, H., Touw, D. M., Jolles, J., van Boxtel, M. P. J., et al. (2005). The effect of two types of memory training on subjective and objective memory performance in healthy individuals aged 55 years and older: a randomized controlled trial. *Patient education and counseling*, 57(1), 106-114. doi: 10.1016/j.pec.2004.05.002
- Veenhoven, R. (2004). Subjective measures of well-being. *WIDER Discussion Paper*, 07.
- Verhaeghen, P., & Salthouse, T. A. (1997). Meta-analyses of age-cognition relations in adulthood: Estimates of linear and nonlinear age effects and structural models. *Psychological bulletin*, 122(3), 231-249. doi: 10.1037/0033-2909.122.3.231
- Wilson, M. (2005). *Constructing measures: an item response modeling approach*. Mahwah, N.J: Lawrence Erlbaum Associates.
- World Health Organization. (1992). *International Statistical Classification of Disease and Related Health Problems, Tenth Revision (ICD-10)*. Geneva: Author.
- Young, T., Rabago, D., Zgierska, A., Austin, D., & Laurel, F. (2003). Objective and subjective sleep quality in premenopausal, perimenopausal, and postmenopausal women in the Wisconsin Sleep Cohort Study. *Sleep*, 26(6), 667.
- Zandi, T. (2004). Relationship between subjective memory complaints, objective memory performance, and depression among older adults. *American Journal of Alzheimer's Disease and Other Dementias*, 19(6), 353-360. doi: 10.1177/153331750401900610

Appendix

Abbreviations and acronyms

AD	Alzheimer's disease
ANOVA	Analysis of variance
BDT	Block Design Test
CDQ	Cognitive Dysfunction Questionnaire
CES-D	Center for Epidemiological Studies Depression Scale
CFA	Confirmatory factor analysis
CFI	Comparative fit index
EFA	Exploratory factor analysis
IQ	Intelligence quotient
LISREL	Linear Structural Relations (statistics software)
MCI	Mild cognitive impairment
ML	Maximum likelihood (estimation)
MMSE	Mini-Mental State Examination
PASW	Predictive Analytics Software (statistics software)
PCA	Principal components analysis
PRMQ	Prospective and Retrospective Memory Questionnaire
RMLE	Robust maximum likelihood estimation
RMSEA	Root mean squared error of approximation
S(1-6)	Sample in the Betula study
SB	Satorra-Bentler
SPSS	Statistical Package for the Social Sciences (statistics software)
SRMR	Standardized root mean squared residuals
T(1-5)	Occasion for data collection in the Betula study
TCI	Temperament and Character Inventory
WAIS	Wechsler Adult Intelligence Scale
WAIS-R	Wechsler Adult Intelligence Scale Revised

Statistical notation

N	Number of participants in sample
n	Number of participants in subsample
F	F-ratio (in ANOVA)
MSE	Mean square error
p	Probability (statistical significance)
η^2	Effect size (in ANOVA)
SD	Standard deviation
α	1. Probability level 2. Cronbach's α (alpha); internal consistency reliability estimate
r	Pearson correlation coefficient

ns	Not statistically significant
CI	Confidence interval
R^2	Coefficient of determination (explained variance)
ΔR^2	Difference in R^2 between models
B	Unstandardized regression coefficient
β	Standardized regression coefficient
t	t statistic (in t-tests)
r_s	Spearman rank order correlation coefficient
χ^2	Chi-square
S-B χ^2	χ^2 adjusted with the Satorra-Bentler scaling procedure
Δ S-B χ^2	Difference in S-B χ^2 between nested models

Errata

Article III

Page 221, heading *General knowledge*: “ten questions” should be “twenty-seven questions”.