Self-Reports on Memory Functioning in a Longitudinal Study of the Oldest Old: Relation to Current, Prospective, and Retrospective Performance

Boo Johansson, Rebecca Allen-Burge, and Steven H. Zarit

1Institute of Gerontology, University College of Health Sciences, Jönköping, Sweden.
2Gerontology Center, The Pennsylvania State University.

Self-evaluation of memory performance, one aspect of metamemory, may be an important indicator of concurrent, retrospective, or future decline in memory functioning. The relationships among self-evaluations, cognition, and outcome were investigated in the OCTO study, a longitudinal, population-based panel of the oldest old. Using concurrent data, results indicated that overall cognitive ability, depression, gender, and education were associated with self-reports of memory for the entire sample. The relation of perception of decline to actual decline was also examined. Self-reported decline over a 2-year period was associated with actual decline in performance on three tests of memory. Finally, self-reported memory function was investigated as an indicator of future cognitive decline and diagnosis of dementia. These self-evaluations predicted decline on specific tests of memory over 2 years and subsequent diagnosis of dementia after 2 and 4 years. The amount of variance accounted for by self-evaluations, however, was relatively small, suggesting that complaints reflect different processes, only one of which is the pathological decline involved in dementia.

Most investigations of metamemory in adulthood and aging have relied on one of three operational definitions (Devolder & Pressley, 1989). These include (a) "recall-readiness," or ability to accurately know when new information has been acquired, (b) "performance prediction," or ability to accurately predict what information one knows prior to attempted recall or recognition, and (c) self-evaluation of memory ability, memory complaints, and ability to cope with declining memory skills. The present study focuses on self-evaluations of memory performance, and their relationship to concurrent, retrospective, and prospective memory functioning.

Whereas investigation of cognition and memory performance has a long history in the geropsychology literature (Schaie, 1990, 1994), the study of self-evaluations in old age is less well investigated despite the fact that self-reports are often used in clinical settings as an indicator of memory performance and decline. The clinical interest in using such judgments as possible early indicators of dementia has spawned research into the concurrent relationship between individuals' memory complaints and memory performance, judged either by the individual or by a collateral source (Grut et al., 1993; Scogin, Storandt, & Lott, 1985; Smith, Petersen, Ivnik, Malec, & Tangalos, 1996). For example, Grut et al. investigated the relationship between memory complaints, memory performance, and diagnosis of dementia for a representative sample of a district in Stockholm, Sweden. Memory performance ratings were made both by participants and a collateral source, usually the primary caregiver when the participant had dementia. Grut and collaborators found that there was little agreement between collateral sources' and participants' ratings of the participants' memory performance. Additionally, the ratings made by a collateral source were more effective at discriminating demented from nondemented participants and were associated with objective cognitive performance and dementia stage. Among nondemented participants, memory complaints were related to greater cognitive impairment, whereas among demented patients, memory complaints were related to lesser cognitive impairment. The results of this study argue for a distinction between the metamemorial abilities of nondemented individuals and demented individuals in various stages of decline (Kaszniaik, 1996).

Complaints not only vary with level of dementia but also with depressive symptoms at the time of assessment (Burt, Zembar, & Niederehe, 1995; Kahn, Zarit, Hilbert, & Niederehe, 1975). Complaints are therefore not a specific indicator of concurrent cognitive decline (Bolla, Lindgren, Bonaccorsy, & Bleeker, 1991; Scogin et al., 1985). For someone who is depressed, complaints are part of a general cognitive style in which negative experiences are emphasized. For other people, complaints may reflect a general pessimism. In fact, not only complaints but observed memory impairment may vary with person variables such as pessimism. For example, Burt et al. concluded, based on a meta-analysis of the memory and depression literature, that memory impairment may even be more strongly associated with general psychopathology than with depressive symptoms.

Several multidimensional self-report inventories have been developed, but only two have been recommended for general clinical use: (a) the Metamemory in Adulthood Questionnaire (MIA; Dixon, Hultsch, & Hertzog, 1988; Hertzog, Dixon, & Hultsch, 1990; Hultsch, Hertzog, & Dixon, 1987) and (b) the Memory Functioning Questionnaire.
naire (MFQ; Gilewski & Zelinski, 1988; Gilewski, Zelinski, & Schaie, 1990; Zelinski & Gilewski, 1988; Zelinski, Gilewski, & Anthony-Bergstone, 1990). These are relatively lengthy instruments (the MIA has 120 items; the MFQ has 92) that contain several different dimensions of metamemory. In clinical practice, self-evaluations of memory are typically addressed in a brief manner with a few questions about perceived decline and current difficulties (Kaszniaik, 1996).

Reliance on self-evaluations of memory functioning for insight into concurrent memory performance is problematic, however, due to other general factors that affect self-reported memory and to conflicting evidence regarding the modest relationship between objective memory performance and metamemory reports. One reason for the modest association between memory and self-evaluations is that metamemorial abilities do not keep pace with changes in actual memory skills. For example, Bruce, Coyne and Botwinick (1982) found that increased age of participants was associated with overestimation of number of words learned in a list learning task, using the recall-readiness paradigm for metamemory. They proposed that participants likely made predictions based on past performance, but that their subjective evaluations of ability had not been adjusted for actual decline over time. These findings suggest that the relation between performance and self-evaluations involves a time lag, with reports of current functioning reflecting, in part, past levels of performance.

Other researchers conclude that metamemorial skills (i.e., feeling of knowing judgments in the performance prediction paradigm) remain intact with age (Bäckman & Karlsson, 1985; Butterfield, Nelson, & Peck, 1988; Rabinowitz, Ackerman, Craik, & Hinchley, 1982). Bäckman and Karlsson and Butterfield and associates found in their investigations of recall failures that there were no age differences in the accuracy with which individuals could predict what information they would later be able to recognize. This lack of an age effect could be due to individuals making predictions after having failed to recall information. Thus, participants made predictions with the benefit of knowledge that they failed to recall the information initially. While this seems a more ecologically valid measure of metamemory, it is unknown to what extent individuals’ monitoring of their own memory performances and memory failures gives rise to subjective memory complaints. There has also been a paucity of research addressing the very question in which clinicians are most interested: Do self-evaluations and memory complaints predict subsequent memory decline?

**Longitudinal Investigations of Metamemory**

Previous findings suggest that between 38% and 80% of older adults report subjective memory deficits (Bolla et al., 1991; Flicker, Ferris, & Reisberg, 1993; Grut et al., 1993; Zelinski & Gilewski, 1988). Some argue that memory complaints are an indicator of decline even in the absence of evidence of concurrent cognitive impairment. Reisberg, Ferris, de Leon, and Crook (1985) suggest that the relationship between complaints and level of dementia follows an inverted U function, peaking when dementia symptoms are incipient or mild. It has been argued that some individuals have an early awareness of incipient dementia and may begin to complain about memory losses even prior to the detection of such problems on neuropsychological tests (Reisberg et al., 1985).

Research has recently focused on the ability of memory complaints to predict subsequent memory decline (Flicker et al., 1993; Scogin & Bienias, 1988; Smith et al., 1996). For instance, Flicker and associates conducted a longitudinal investigation of memory complaints and memory performance in 59 healthy individuals (i.e., nondemented) with subjective memory complaints. At the 3-year follow-up, they found little evidence of progressive cognitive deterioration among the participants. What is remarkable about these findings is that all participants were selected due to subjective complaints of memory loss. There was evidence of decline in performance on only 2 of 12 tests in the assessment battery. Similarly, Smith and associates report that memory complaints had little sensitivity in predicting onset of dementia. Clearly, the relationship between complaint and subsequent performance warrants closer scrutiny.

**The Current Investigation**

To address the need for longitudinal research examining the predictive validity of memory self-evaluations, the current investigation examined whether such self-reports are restricted to modest concurrent relationships with memory performance, or whether they are more robust predictors of later decline. We hypothesized that (1) concurrent self-evaluations of memory performance would be associated with multiple independent variables, including cognitive ability, depressive symptoms, and participant characteristics such as gender and years of education; (2) self-reports of decline over a 2- or 4-year period would be associated with actual decline on cognitive tests; and (3) self-evaluations of memory functioning would predict subsequent decline in overall functioning. Specifically, we predicted that self-evaluation of memory would account for variance in decline in cognitive functioning and subsequent diagnosis of dementia.

**MATERIAL AND METHODS**

The sample was drawn from the OCTO [genarian] study, which is a population-based longitudinal panel of the oldest old conducted in the municipality of Jönköping located in the South Central region of Sweden. The total population of the municipality of Jönköping, including surrounding towns and rural areas, is approximately 113,000.

Using census data, four groups of 100 individuals each were identified with the following birth years: 1897, 1899, 1901, and 1903. All residents with those birth years were potential candidates for the sample, including people living in sheltered housing, nursing homes, or other institutional settings. The subjects were initially examined in 1987/1988 when they were 84, 86, 88, and 90 years of age. Following a cohort sequential design, they were reexamined after 2, 4, and 6 years. Only data from the first three measurements were used in this study.

From the initial 400 subjects identified from census data, 324 agreed to participate in the study. Twenty-one subjects were deceased before the investigation was scheduled, 1 potential subject had moved out of the area, and 54 de-
follow-up (Time 2) it was possible to re-examine 161 of the
participants, using the principal component method. The four
items all loaded on a single factor, with factor loadings
ranging between .70 and .78, and accounting for 57% of the
variance. Alpha for the scale was .69 (standard item alpha = .77). This single factor corresponds to the capacity (knowledge of one's own memory capacities) and change dimensions of the MIA questionnaire (Dixon et al., 1988) and to the general rating of memory in the MFQ battery (Gilewski & Zelinski, 1988). Higher scores indicate more positive evaluations of memory.

Memory and cognitive performance. — Cognitive functioning was assessed with a battery of five cognitive tests. The tests were selected with two purposes in mind: (1) to represent different cognitive dimensions known to be affected by dementia, and (2) to use an ecological approach employing familiar stimuli. This approach was taken to maximize subjects' interest and motivation for the tests. The tests and their validity are described in more detail in Johansson and Zarit, 1991 (see Note 2). They included:

   Participants are asked to recall 10 common objects and where they placed these objects in a three-dimensional model of an apartment. The test consists of a free recall test, followed by a recognition task for items that were not recalled. Finally, individuals are asked to place the objects in the same locations as they did previously. The score reported here is the sum of free recall plus recognition.

2. Prose recall. In this verbal memory test, participants are asked for free recall of a brief (100 words) story that has a humorous point. This test has a similar format and similar scoring procedures as the story recall (i.e., logical memory) in the Wechsler Memory Scale. The Guttman split-half reliability coefficient across the 16 sentences was .88.

3. Clock test. In this visuospatial test, individuals are asked to (a) draw a clock and set the hands to a certain time, (b) set the hands on a wooden clock with no numbers on the clock face to certain standard times, and (c) with the wooden clock set to certain

clined participation. Excluding those who were deceased and the one who had moved, the overall participation rate was 86%. Because the sampling frame consisted of everyone in the population, including residents of nursing homes and other specialized housing for disabled elderly, many participants were quite frail and could not complete parts of the interview. Missing data were almost entirely due to severe sensory loss or severe dementia. For the current investigation, 267 people (172 women and 95 men) from the pool of 324 subjects completed the self-report measures on memory. In some analyses, missing data on cognitive tests and other scales used in these analyses reduced the sample size further. This high rate of missing data underscores the frailty of this population. The sample used in analyses thus has somewhat better functioning than the whole population, as people with significant cognitive and/or sensory deficits were unable to complete relevant portions of the interview.

Of the 267 participants with memory self-report data, 215 were nondemented at Time 1, 51 were diagnosed as demented at Time 1, and 1 participant did not have sufficient data to complete a dementia diagnosis. At the first 2-year follow-up (Time 2) it was possible to re-examine 161 of the 215 individuals with memory self-evaluation data and without dementia at Time 1. One hundred and five individuals survived to the second follow-up (Time 3); of these participants, 85 have complete data at Time 3. The attrition in the sample mainly reflects mortality. The overall nonparticipation rate in the study for reasons other than mortality was 3.7% between Time 1 and Time 2 and 6.7% between Time 2 and Time 3.

The mean age of the participants was 86.85 (SD = 2.28) at baseline (Time 1). As generally found in this age group, there were significantly more women (64%) than men (36%). Average education of the sample was 6.5 years (SD = 1.80), which is typical for older Swedish cohorts (see Note 1). Eight percent of the subjects had less than the basic 6 years of elementary schooling and 27% had more. Most participants were widowed (60%), 24% were married, 15% were never married, and 1% were divorced. Sixty-eight percent lived in ordinary housing while 28% were in service apartments (in which some housekeeping, meal, social and health care services are available) and 4% resided in institutional settings. Seventy-seven percent were living in urban areas of the municipality, which corresponds to the figure for the entire population. A comparison between participants and those who declined participation at Wave 1 showed no significant differences for age, gender, marital status, or housing arrangements.

Procedures and Measures
A comprehensive battery of structured behavioral, social, and biomedical assessments was administered in the participants' places of residence, including measures of functional abilities, cognitive performance, mood, demographics, and socioeconomic information. Medical status was assessed by determining current symptoms, diagnosed illnesses, and medications, using a combination of self-reports and medical records. The entire in-person examination took about 3.5 to 4 hours, including breaks for coffee and social talk. Table 1 shows baseline performance of the entire sample on each of the instruments used in this study.

Self-evaluations on memory performance. — Four questions drawn from typical clinical assessments were used to assess self-evaluations of memory. The items were:

1. On the whole, do you think that your memory is good or poor?
2. Do you think you have problems with your memory that make your life more difficult?
3. Do you think that your memory has gotten worse during the past 2 years?
4. On the whole, do you think that you have problems remembering things that you want to do or say?

The items had a Likert-type response format, indicating the degree of severity of a problem. The third item about experiences of declining memory was used separately for some of the analyses, while all items were used to form a self-evaluation scale. To determine if the self-report items could be formed into a scale, a factor analysis was performed, using the principal component method. The four items all loaded on a single factor, with factor loadings ranging between .70 and .78, and accounting for 57% of the variance. Alpha for the scale was .69 (standard item alpha = .77). This single factor corresponds to the capacity (knowledge of one's own memory capacities) and change dimensions of the MIA questionnaire (Dixon et al., 1988) and to the general rating of memory in the MFQ battery (Gilewski & Zelinski, 1988). Higher scores indicate more positive evaluations of memory.

Memory and cognitive performance. — Cognitive functioning was assessed with a battery of five cognitive tests. The tests were selected with two purposes in mind: (1) to represent different cognitive dimensions known to be affected by dementia, and (2) to use an ecological approach employing familiar stimuli. This approach was taken to maximize subjects' interest and motivation for the tests. The tests and their validity are described in more detail in Johansson and Zarit, 1991 (see Note 2). They included:

   Participants are asked to recall 10 common objects and where they placed these objects in a three-dimensional model of an apartment. The test consists of a free recall test, followed by a recognition task for items that were not recalled. Finally, individuals are asked to place the objects in the same locations as they did previously. The score reported here is the sum of free recall plus recognition.

2. Prose recall. In this verbal memory test, participants are asked for free recall of a brief (100 words) story that has a humorous point. This test has a similar format and similar scoring procedures as the story recall (i.e., logical memory) in the Wechsler Memory Scale. The Guttman split-half reliability coefficient across the 16 sentences was .88.

3. Clock test. In this visuospatial test, individuals are asked to (a) draw a clock and set the hands to a certain time, (b) set the hands on a wooden clock with no numbers on the clock face to certain standard times, and (c) with the wooden clock set to certain
times, respond to questions about: “What’s the time?” The Guttman split-half reliability coefficient across the 13 items was .94.

(4) Coin test. In this sorting task, concept formation and basic arithmetic abilities are tested. People are presented with 10 each of 4 Swedish coins (10 ore, 50 ore, 1 krona, 5 kronor) and asked to choose the least number of coins that add up to four preselected totals. The Guttman split-half reliability coefficient was .96.

(5) The Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975). This brief measure assesses global cognitive ability and has been used extensively in epidemiological studies to identify dementia-related problems. Test-retest reliabilities are typically found as high as .90, and interrater reliabilities are generally high (Farber, Schmitt, & Logue, 1988). The Guttman split-half reliability coefficient in the present sample was .71.

Diagnosis of dementia. — DSM-III-R criteria (American Psychiatric Association, 1987) were used for the diagnosis of dementia. Diagnosis was based on cognitive test performance and on the ability to perform instrumental activities of daily living (IADL), especially more complex functions such as managing finances and shopping. Sensory and motor impairment were considered when evaluating cognitive and IADL performance. The protocols in the assessment battery and information from medical records were reviewed for consistency with the diagnosis, including information on depression, delirium, and other diseases known to affect memory and cognition. Additional medical evaluations were not undertaken to determine type of dementia. Two independent ratings were made, using all the above information, to examine interrater agreement. The raters, who were blind to prior or subsequent status of the cases, agreed on 93.8% of the cases, with a corresponding kappa coefficient of 0.88. For more detailed information on the dementia diagnostic procedure and interrater agreement, see Johansson and Zarit (1995).

Depression. — For the present study depression was measured with a short form of the Center for Epidemiological Studies Depression Scale (CES-D; Kohout, Berkman, Evans, & Cornoni-Huntley, 1993; Radloff, 1977; Radloff & Teri, 1986). The short form consists of 11 items designed to measure an individual’s current level of depressive symptomatology. Respondents are asked about the frequency with which they experienced each of 11 symptoms and problems during the past week. The items are rated on a 4-point scale, ranging from “rarely or none of the time” to “most or all of the time.” A total depression score is calculated by summing the items. In the present investigation the CES-D was reverse-scored, so that higher scores indicate lower depression.

RESULTS

The analyses were conducted in the following steps, as shown in Figure 1. First, the five cognitive tests, depression, and demographic variables including gender and education were evaluated as indicators of concurrent assessments of memory functioning (Paths A in Figure 1). Second, the ability of individuals to identify previous decline in memory (i.e., provide retrospective assessments) was examined by comparing reported decline to actual decline on the five cognitive tests (Paths B in Figure 1). Third, logistic regressions were used to examine three questions about the prospective relationships between self-evaluation and memory performance (Paths C in Figure 2): (1) Do self-evaluations among individuals who were nondemented at baseline predict decline in memory performance (i.e., decline versus no decline) from Time 1 to Time 2 and from Time 1 to Time 3? (2) Do self-evaluations at Time 1 among individuals who remained nondemented at Time 2 predict decline in memory performance from Time 1 to Time 3? (3) Do self-evaluations among individuals who were nondemented at baseline predict risk of diagnosis of dementia at either Time 2 or Time 3? It should be noted that due to the frailty of this age group and the whole population sampling frame, there was a great deal of missing data for certain items or tests. For each analysis we included participants without missing data. Therefore, some analyses use a subset of the maximum sample available at each wave.

Concurrent Relations Between Self-Evaluation and Performance

As a first step, regressions were performed to examine factors associated with memory self-evaluations at Time 1. Simultaneous regressions were conducted using the whole sample (N = 220) and then separately for people with (N = 33) and without (N = 187) a diagnosis of dementia. The independent variables were (1) the five cognitive tests, (2) depression, and (3) demographic variables such as gender and education (see Paths A in Figure 1). The simultaneous regression for the whole sample was significant, F(8,210) = 10.14, p < .0001 accounting for 28% of the variance in the self-evaluation scale. Among the cognitive variables only the MMSE was significantly related to self-evaluations, r(210) = 3.02, p < .01, with better performance associated with fewer complaints. The other cognitive tests had significant univariate correlations with the scale (see Table 2), but because of the shared variance among the tests only the MMSE remained significant in the regression. Along with performance on the MMSE, the self-evaluation scale was significantly as-

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Range*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory self-evaluation</td>
<td>9.89</td>
<td>1.91</td>
<td>4-12</td>
</tr>
<tr>
<td>MMSE</td>
<td>24.11</td>
<td>5.48</td>
<td>0-30</td>
</tr>
<tr>
<td>MIR</td>
<td>8.08</td>
<td>3.43</td>
<td>0-10</td>
</tr>
<tr>
<td>Prose recall</td>
<td>11.73</td>
<td>4.65</td>
<td>0-15</td>
</tr>
<tr>
<td>Clock test</td>
<td>7.03</td>
<td>2.17</td>
<td>0-08</td>
</tr>
<tr>
<td>Depression</td>
<td>35.66</td>
<td>5.45</td>
<td>16-44</td>
</tr>
</tbody>
</table>

Note. Where MMSE = Mini-Mental State Examination, MIR = Memory-in-Reality Test, and Depression is a reverse-scored short form of the Center for Epidemiological Studies Depression Scale.

*Observed score range.
associated with depression, \( t(210) = 3.26, p < .001 \). People who were less depressed had a more positive evaluation of their memory. Finally, both demographic variables were significant, demonstrating that women, \( t(210) = 2.16, p < .05 \), and those with less education, \( t(210) = 3.13, p < .01 \), rated their own memory more positively. Table 3 shows the results of these analyses.

Separate regressions for nondemented and demented subjects revealed similar findings. Each accounted for a similar amount of variance (\( R^2 = .27 \) and .33, respectively). For the nondemented group, the MMSE, depression, education, and gender were again significantly associated with self-evaluations, reflecting the same relationships as in the total sample. Because of the smaller sample size, only the MIR was significantly related to concurrent self-evaluation in the dementia group, \( t(24) = 2.49, p < .05 \). A better performance on the MIR was associated with positive evaluations of memory. Comparing the 95% CI for the beta weights in the two samples for each test showed there was no overlap in performance for two of the eight measures, the MMSE (95% CI = .68, .12) and the MIR (95% CI = -.13, -.63). Again, a better performance on the MIR was associated with positive evaluations of memory. By contrast, the MMSE had a negative relation with the self-evaluation scale in the dementia sample (i.e., better performance was associated with more complaints of poor memory), and a positive relation in the nondemented (i.e., better performance was associated with better memory evaluations).

**Self-Evaluations as a Retrospective Indicator of Decline**

In examining the retrospective relationship among the cognitive variables and self-evaluations, we examined the extent to which reports of decline during a time interval corresponded to actual decline in performance. Participants with dementia at Time 1 would be expected to decline and thus were excluded from this analysis. Data were complete for a maximum of 157 of the possible 161 survivors at Time 2. Respondents who reported that their memory had declined \( (N = 75) \) had significantly lower self-evaluation scores at Time 1, \( t(155) = 4.15, p < .01 \), in comparison with individuals who thought that their memory had remained stable over the 2-year interval \( (N = 82) \).

We related changes in cognitive performance between Time 1 and Time 2 to whether subjects reported at Time 2 that their memory had declined in the recent past (see Paths B in Figure 1). Respondents’ reports of decline were examined in relation to decline in scores on the five cognitive tests and incidence of dementia. A reported decline was associated with actual decline on three of the five memory tests, as follows: MIR, \( t(91.43) = 2.35, p < .02 \), prose recall, \( t(142) = 3.50, p < .001 \), and the coin test \( t(113.35) = 2.61, p < .01 \). Diagnosis of dementia at Time 2 was highly associated with self-report of decline at Time 2, \( \chi^2 = 12.22, p < .01 \). Similar analyses using reports of decline at Time 3 for people nondemented at Time 2, however, failed to reveal significant relationships. This was likely due to the smaller sample size (stable = 33, decline = 38).

**Table 2. Concurrent Associations Between Self-Evaluations, Performance on Cognitive Tests, and Sociodemographic Variables \( (N = 267, \text{maximum}) \)**

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Memory self-evaluations</td>
<td>1.00</td>
<td>.03</td>
<td>.09</td>
<td>.40**</td>
<td>.34**</td>
<td>.32**</td>
<td>.22**</td>
<td>.32**</td>
<td>.31**</td>
</tr>
<tr>
<td>2. Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Years of school</td>
<td>.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. MMSE</td>
<td>.09</td>
<td>.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Story recall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Clock test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Coin test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. MIR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Depression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. Where MMSE = Mini-Mental State Examination, MIR = Memory-in-Reality Test, and Depression is a reverse-scored short form of the Center for Epidemiological Studies Depression Scale.

*p < .05; **p < .01.
Self-Evaluations Predicting Subsequent Decline

The relationship between self-evaluation at Time 1 and decline in memory performance at Time 2 or Time 3 was examined using logistic regressions (see Paths C, Figure 1). Because people with dementia at Time 1 would be expected to decline, they were excluded from these analyses. A dichotomous variable indicating decline (i.e., a change in score of at least .5 SD) was computed for each of the five cognitive tests and used as the dependent variable. This was done because change between time periods was not normally distributed, with a sizable proportion of the sample having little or no change on a measure. The amount of change was set at .5 SD because of the concern of catching not only pronounced impairments, but also more subtle decline. A negative score indicates decline while a positive score indicates stability or improvement. For the periods from Time 1 to Time 2 and from Time 1 to Time 3, respectively, the model included self-evaluations as the only predictor variable. Data for these analyses were available for 155 of the 161 survivors at Time 2.

Self-evaluation at Time 1 predicted decline in performance on the following measures: coin test at Time 2, R^2 = .05, p < .01; and Time 3, R^2 = .03, p < .05; MIR at Time 2, R^2 = .05, p < .01, and Time 3, R^2 = .12, p < .01; and prose recall at Time 2, R^2 = .04, p < .01. Tables 4 and 5 summarize the results obtained from this series of analyses.

The relation of self-evaluations at Time 1 to subsequent incidence of dementia was examined next (see Figure 1). Only individuals who were nondemented at Time 1 and survived to Time 2 (N = 161) or Time 3 (N = 85) were used in these analyses. Two logistic regressions were performed using self-evaluations at Time 1 as the predictor variable. The dependent variable in the first regression was incidence of dementia at Time 2. In the second regression, the dependent variable was incidence of dementia at any point over the 4-year period examined. The results showed that self-evaluations at baseline were significantly associated with subsequent diagnosis of dementia. That is, people who evaluated their memory more negatively were more likely to develop dementia at Time 2, R^2 = .06, p < .001, or at any point during the period examined, R^2 = .03, p < .01. Table 6 summarizes these results.

**DISCUSSION**

This study examined the relation of self-evaluations of memory and cognitive functioning in the context of a longitudinal study of the oldest old. While most work on memory complaints has focused on the young-old and concurrent as-

---

**Table 3. Simultaneous Regression (Standardized) Accounting for Variance in Self-Evaluations at Time 1 for the Total Sample With Complete Interviews (N = 220)**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta</th>
<th>SE</th>
<th>95% CI</th>
<th>t</th>
<th>p-value</th>
<th>R^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>.50</td>
<td>.23</td>
<td>.04 / .96</td>
<td>2.16</td>
<td>.03*</td>
<td>.05</td>
</tr>
<tr>
<td>Years of education</td>
<td>-.20</td>
<td>.06</td>
<td>-.32 / .07</td>
<td>-3.13</td>
<td>.00**</td>
<td></td>
</tr>
<tr>
<td>MIR</td>
<td>.01</td>
<td>.05</td>
<td>-.09 / .10</td>
<td>.14</td>
<td>.89</td>
<td></td>
</tr>
<tr>
<td>Prose recall</td>
<td>.04</td>
<td>.04</td>
<td>-.05 / .12</td>
<td>1.20</td>
<td>.23</td>
<td></td>
</tr>
<tr>
<td>Clock test</td>
<td>.05</td>
<td>.04</td>
<td>-.02 / .12</td>
<td>1.29</td>
<td>.20</td>
<td></td>
</tr>
<tr>
<td>Coin test</td>
<td>-.02</td>
<td>.08</td>
<td>-.18 / .14</td>
<td>-.02</td>
<td>.84</td>
<td></td>
</tr>
<tr>
<td>MMSE</td>
<td>.13</td>
<td>.04</td>
<td>-.04 / .21</td>
<td>3.02</td>
<td>.00**</td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td>.07</td>
<td>.02</td>
<td>.03 / .11</td>
<td>3.26</td>
<td>.00**</td>
<td></td>
</tr>
</tbody>
</table>

*Note. Where CI = 95% confidence interval (lower/upper limits), and R^2 is the estimated proportion of variance explained.

* p < .05; ** p < .01.

---

**Table 4. Logistic Regression Models (Standardized) With Self-Evaluations at Time 1 Predicting Decline in Performance on Cognitive Tests at Time 2 (N = 155, maximum)**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta</th>
<th>SE</th>
<th>p-value</th>
<th>95% CI</th>
<th>R^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIR</td>
<td>.38</td>
<td>.13</td>
<td>.00**</td>
<td>.12 / .63</td>
<td>.05</td>
</tr>
<tr>
<td>Prose recall</td>
<td>.34</td>
<td>.11</td>
<td>.00**</td>
<td>.12 / .56</td>
<td>.04</td>
</tr>
<tr>
<td>Clock test</td>
<td>.12</td>
<td>.12</td>
<td>.01**</td>
<td>-.11 / .35</td>
<td>.00</td>
</tr>
<tr>
<td>Coin test</td>
<td>.42</td>
<td>.16</td>
<td>.01**</td>
<td>.11 / .73</td>
<td>.05</td>
</tr>
<tr>
<td>MMSE</td>
<td>.08</td>
<td>.10</td>
<td>.42</td>
<td>-.12 / .28</td>
<td>.00</td>
</tr>
</tbody>
</table>

*Notes. Where CI = 95% confidence interval (lower/upper limits), and R^2 is the estimated proportion of variance accounted for by metamemory.  
*Model χ^2 = 9.50, p = .00.  
*Model χ^2 = 10.43, p = .001.  
*Model χ^2 = 9.096, p = .34.  
*Model χ^2 = 6.69, p = .01.  
*Model χ^2 = 0.65, p = .42.  

**p < .01.

---

**Table 5. Logistic Regression Models (Standardized) With Self-Evaluations at Time 1 Predicting Decline in Performance on Cognitive Tests at Time 3 (N = 98, maximum)**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta</th>
<th>SE</th>
<th>p-value</th>
<th>95% CI</th>
<th>R^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIR</td>
<td>.66</td>
<td>.21</td>
<td>.00**</td>
<td>.25 / 1.07</td>
<td>.12</td>
</tr>
<tr>
<td>Prose recall</td>
<td>.11</td>
<td>.14</td>
<td>.41</td>
<td>-.16 / .38</td>
<td>.00</td>
</tr>
<tr>
<td>Clock test</td>
<td>.02</td>
<td>.14</td>
<td>.87</td>
<td>-.25 / .29</td>
<td>.00</td>
</tr>
<tr>
<td>Coin test</td>
<td>.33</td>
<td>.15</td>
<td>.03*</td>
<td>.04 / .62</td>
<td>.03</td>
</tr>
<tr>
<td>MMSE</td>
<td>.15</td>
<td>.12</td>
<td>.24</td>
<td>-.08 / .38</td>
<td>.00</td>
</tr>
</tbody>
</table>

*Notes. Where CI = 95% confidence interval (lower/upper limits), and R^2 is the estimated proportion of variance accounted for by metamemory.  
Model χ^2 = 12.27, p = .0005.  
Model χ^2 = 6.09, p = .41.  
Model χ^2 = .025, p = .87.  
Model χ^2 = 4.72, p = .03.  
Model χ^2 = 1.39, p = .24.  

**p < .05; **p < .01.

---

**Table 6. Logistic Regression Models (Standardized) With Self-Evaluations at Time 1 Predicting Diagnosis of Dementia Prospectively.**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta</th>
<th>SE</th>
<th>p-value</th>
<th>95% CI</th>
<th>R^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dementia at 2*</td>
<td>-.41</td>
<td>.12</td>
<td>.60**</td>
<td>-.64 / -.17</td>
<td>.06</td>
</tr>
<tr>
<td>Dementia at 2 or 3*</td>
<td>-.28</td>
<td>.10</td>
<td>.06**</td>
<td>-.48 / -.08</td>
<td>.03</td>
</tr>
</tbody>
</table>

*Notes. Where CI = 95% confidence interval (lower/upper limits), and R^2 is the estimated proportion of variance accounted for by metamemory.  
Model χ^2 = 12.91, p = .0005.  
Model χ^2 = 7.42, p = .006.

**p < .01.
associations, that is, assessment of how people evaluate their memory and actual performance measured at the same point in time, only a few studies have examined the predictive power of self-evaluations of memory over time (Flicker et al., 1993; Scogin & Bienias, 1988; Smith et al., 1996). Self-reported memory decline may be an early indicator of incipient cognitive problems associated with dementia prevalent among the oldest-old, or with more subtle age-related changes. Therefore, the present study examined concurrent relations between memory self-evaluations and cognitive performance, the validity of retrospective self-reported decline, and the predictive power of self-evaluations to detect subsequent changes in performance and diagnosis of dementia.

In the concurrent analyses our results were similar to previously reported findings in younger samples on the relation of memory self-evaluations, cognition, and depression. Our results provided no support for the idea that self-evaluations in the oldest old would be more influenced by expectations of inevitable decline, than by actual performance (Erber, Scuchman, & Rothberg, 1990). People who rated their memory more poorly had lower performance and higher depression scores. In addition, women and people with low education rated their memory more positively. For the whole sample, cognitive performance and depression were significantly associated with self-evaluations. When the sample was divided between people with a diagnosis of dementia and those without, somewhat different patterns emerged for the MIR test and MMSE. The MIR had a straightforward relation with self-evaluations in the dementia sample; that is, better performance was related to better evaluation of memory, but had little influence on self-reports in the non-demented. The MMSE had a more complex relation to self-evaluations. In the nondemented sample better performance was associated with better evaluations of memory. By contrast, people with dementia who performed better on the MMSE evaluated their memory more accurately and are more likely to be aware of their cognitive problems. With more pronounced symptoms of dementia, however, awareness of deficits declines (Kaszniai, 1996; Reisberg et al., 1985). Concurrent assessments of memory self-evaluations and memory performance, then, have a small, significant relation to performance for people who are cognitively intact and for those with mild symptoms of dementia.

We hypothesized that actual decline in cognitive performance over a 2-year or 4-year interval would be associated with retrospective self-reports of decline. Roughly half of the individuals who were cognitively intact at the initial assessment reported their memory had declined between Time 1 and Time 2. Self-reported decline at Time 2 was associated with actual decline on three tests, indicating that individuals are able to a certain extent to evaluate change in their memory functioning retrospectively over a 2-year time frame. This was not found for the 4-year sampling interval, probably because there was not enough power to detect an effect.

Most importantly, this investigation yields evidence of the prospective relation of memory self-evaluations to subsequent memory decline as measured by objective tests and by diagnosis of dementia. Memory complaints are often conceptualized in clinical settings as a possible early indicator of a dementing illness (Reisberg et al., 1985), but there has been a glaring lack of longitudinal research to substantiate this intuition, especially in the oldest old. In this sample, self-evaluations at Time 1 were predictive of decline in performance on three of five cognitive measures and of risk of diagnosis of dementia at both Time 2 and Time 3. Though significant, the magnitude of these associations was nonetheless low, suggesting that self-evaluations are not a powerful clinical sign of impending decline.

Although this study addresses the need for longitudinal evaluation of the predictive power of memory self-evaluations, several limitations to the interpretation of the current findings must be noted. First, the brief clinical instrument assessed self-evaluations in a global way and did not include other dimensions of metamemory that might be important for specific outcomes. It may be, for example, that memory self-efficacy is a better predictor of subsequent decline on objective memory tests, but change is a better predictor of subsequent diagnosis of dementia or mortality (factors taken from the MIA, Hertzog, Dixon, & Hultsch, 1990). Another limitation of the current research is the lack of racial/ethnic or educational diversity in the Swedish sample. The generalizability of these results to other countries, especially with greater diversity in racial and ethnic compositions, is therefore unknown. Of particular note is the small amount of variance in education in this cohort of older Swedes. It may be, however, that the relation between memory self-evaluations and performance can emerge more clearly in a sample like this one, in which education is not a confounding factor.

In summary, these results show that self-evaluations of memory functioning are predictive of later cognitive and functional decline although the magnitude of the relation is small. Thus, self-reports cannot substitute for objective cognitive assessment, but it may be possible to use self-reports of decline to identify at-risk individuals who may then be targeted for intervention, whether training in cognitive abilities or pharmacological trials. Further studies are needed to resolve these issues using a broader battery of assessments of both metamemory and memory performance. The early detection of cognitive problems remains an important goal for research aimed at identifying treatments that can retard the dementing process or improve the quality of life for older individuals.

ACKNOWLEDGMENTS

The OCTO study ("Functional capacity and its determinants in the oldest old") was conducted at the Institute of Gerontology in Jönköping. The study was supported by a grant from the Swedish Social Research Council, SFR. Partial preparation of the manuscript was supported by Training Grant T32-18904, Research Training in Mental Health and Aging, from the National Institute of Mental Health.

Address correspondence to Boo Johansson, Institute of Gerontology, University College of Health Sciences, Box 1038, S-551 11, Jönköping, Sweden. E-mail: Boo.Johansson@hhj.hj.se

REFERENCES


Received January 22, 1996
Accepted December 16, 1996

Appendix

Notes

1. Basic school when these cohorts were children consisted of 6 years (Statistics Sweden, 1996). More advanced education was rare and generally restricted to the upper class.

2. Reliabilities for the prose recall test, MIR recall test, clock test, and MMSE were also evaluated by comparing the scores obtained by interviewers at Time 1 of this investigation. Interviewers were assigned randomly to prospective participants. One-way analysis of variance (test by interviewer) and Scheffé’s post-hoc comparisons found no significant differences between interviewers on any measure. This finding suggests that tests were administered and scored in a reliable way.