**The Visual Recognition Memory Battery:
Test Development, Psychometric Evidence, and Normative Data**

by

Nancy J. McIntyre, M.A.

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

Department of Psychology

**Carleton University
Ottawa, Ontario
January, 1997**

copyright
1997, Nancy J. McIntyre

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

—

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

—

ISBN 0-612-19403-5

Name NANCY J. MCINTYRE

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

Developmental Psychology

SUBJECT TERM

5620

SUBJECT CODE

U·M·I

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS

Architecture 0729
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION

General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0688
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Financial 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematical 0280
Music 0522
Philosophy of 0998
Physical 0523

Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS

Language 0679
Ancient 0289
Linguistics 0290
Modern 0291
Literature 0401
Classical 0294
Comparative 0295
Medieval 0297
Modern 0298
African 0316
American 0591
Asian 0305
Canadian (English) 0332
Canadian (French) 0355
English 0593
Germanic 0311
Latin American 0312
Middle Eastern 0315
Romance 0313
Slavic and East European 0314

PHILOSOPHY, RELIGION AND THEOLOGY

Philosophy 0422
Religion 0318
General 0318
Biblical Studies 0321
Clergy 0319
History of 0320
Philosophy of 0322
Theology 0469

SOCIAL SCIENCES

American Studies 0323
Anthropology 0324
Archaeology 0326
Cultural 0327
Business Administration 0454
General 0310
Accounting 0272
Banking 0770
Marketing 0338
Canadian Studies 0385
Economics 0501
General 0503
Agricultural 0505
Commerce-Business 0508
Finance 0509
Labor 0510
Theory 0511
Folklore 0358
Geography 0366
Gerontology 0351
History 0578

Ancient 0579
Medieval 0581
Modern 0582
Black 0328
African 0331
Asia, Australia and Oceania 0332
Canadian 0334
European 0335
Latin American 0336
Middle Eastern 0333
United States 0337
History of Science 0585
Law 0398
Political Science 0615
General 0616
International Law and 0617
Rationalism 0617
Public Administration 0814
Recreation 0452
Social Work 0626
Sociology 0627
Criminology and Penology 0938
Demography 0631
Ethnic and Racial Studies 0628
Individual and Family 0629
Studies 0630
Industrial and Labor 0700
Reasons 0344
Public and Social Welfare 0709
Social Structure and 0999
Development 0453
Theory and Methods 0709
Transportation 0999
Urban and Regional Planning 0453
Women's Studies

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES

Agriculture 0473
General 0285
Agronomy 0475
Animal Culture and 0476
Nutrition 0359
Animal Pathology 0478
Food Science and 0479
Technology 0480
Forestry and Wildlife 0777
Plant Culture 0746
Plant Pathology 0306
Plant Physiology 0287
Range Management 0308
Wood Technology 0309
Biology 0379
General 0329
Anatomy 0353
Botany 0359
Cell 0379
Ecology 0329
Entomology 0353
Genetics 0359
Limnology 0793
Microbiology 0410
Molecular 0307
Neuroscience 0317
Oceanography 0416
Physiology 0433
Radiation 0821
Veterinary Science 0778
Zoology 0472
Biophysics 0786
General 0760
Medical

Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleontology 0445
Paleobotany 0426
Paleogeology 0418
Paleozoology 0985
Palynology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES

Environmental Sciences 0768
Health Sciences 0566
General 0566
Audiology 0300
Chemotherapy 0992
Dentistry 0567
Education 0350
Hospital Management 0769
Human Development 0758
Immunology 0982
Medicine and Surgery 0564
Mental Health 0347
Nursing 0569
Nutrition 0570
Osteotics and Gynecology 0380
Occupational Health and 0354
Therapy 0381
Ophthalmology 0571
Physiology 0419
Pharmacology 0572
Pharmacy 0382
Public Health 0573
Radiology 0574

Speech Pathology 0460
Toxicology 0383
Home Economics 0386

PHYSICAL SCIENCES

Pure Sciences 0485
Chemistry 0749
General 0485
Agricultural 0486
Analytical 0487
Biochemistry 0488
Inorganic 0738
Nuclear 0490
Organic 0491
Pharmaceutical 0494
Physical 0495
Polymer 0754
Radiation 0405
Mathematics 0605
Physics 0986
General 0605
Acoustics 0609
Astronomy and 0606
Astrophysics 0608
Atmospheric Science 0748
Atomic 0607
Electronics and Electronic 0798
Elementary Particles and 0759
Fluid and Plasma 0609
Molecular 0610
Nuclear 0752
Optics 0756
Radiation 0611
Solid State 0603
Statistics 0346
Applied Sciences 0346
Applied Mechanics

Engineering 0517
General 0539
Aerospace 0539
Agricultural 0540
Automotive 0541
Biomedical 0542
Chemical 0543
Civil 0544
Electronics and Electrical 0545
Heat and Thermodynamics 0546
Hydraulic 0547
Industrial 0548
Marine 0549
Materials Science 0549
Mechanical 0549
Metallurgy 0551
Mining 0552
Packaging 0549
Petroleum 0554
Sanitary and Municipal 0554
System Science 0790
Greater Technology 0426
Operations Research 0796
Physics Technology 0795
Textile Technology 0994
Psychology 0621
General 0621
Behavioral 0622
Clinical 0620
Developmental 0623
Experimental 0623
Industrial 0625
Personality 0989
Physiological 0349
Psychobiology 0632
Psychometrics 0632
Social

The undersigned hereby recommend to
the Faculty of Graduate Studies and Research
acceptance of the thesis,

**"The Visual Recognition Memory Battery:
Test Development, Psychometric Evidence, and Normative Data"**

submitted by

Nancy J. McIntyre, B.A., M.A.

in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

Chair, Department of Psychology

Thesis Supervisor

External Examiner

ABSTRACT

The need for a clinical memory test for evaluating memory abilities in individuals with known memory impairments is proposed. A review of the pertinent literature identified and described commonly used memory tests as well as evaluating their strengths and weaknesses. The current research was undertaken to address the difficulties observed. Pilot studies are described which led to the development of a battery of three visual recognition memory tests employing a 4-item multiple choice format using picture, word and abstract designs as stimuli. The tests involve two learning trials and a delay trial. A recognition paradigm was used because it is the most sensitive procedure available to ascertain the amount of material accessible in memory. Male and female volunteers were recruited from the community ranging from the ages of 16 to 85, and with education levels ranging from 8 years to post-graduate work. No gender effects were found, but education effects were present on some of the measures, although of a small nature. Ceiling effects were present for all tests. The most notable finding was the lack of age differences in performance on word and picture stimuli. Scores on the abstract designs test, however, reflected differences across age groups, and an interaction between age and trials. Other findings included learning from Trial 1 to Trial 2 and little or no difficulty with retrieval after a 10-20 minute delay. Appropriate normative tables were developed for each of the three tests to evaluate reliability and to examine the relationship of the test to other memory tests and neuropsychological measures. Results are discussed in terms of the theoretical literature on age effects, rates of forgetting, and stimuli differences. The

findings tend to underline the contribution of an automatic component to recognition memory. Finally, the clinical usefulness of the Recognition Memory Battery tests is discussed in light of present knowledge of memory functioning.

ACKNOWLEDGMENTS

This dissertation has been a long-time coming, and I would like my friends and colleagues, who stuck with me and encouraged me through the process, to know that I have appreciated their continuous support.

First and foremost, I would like to acknowledge the long hours of reading, repeated explanations, and continued confidence of my advisor, Dr. T.N. Tombaugh, who knew I would be successful in the long run. His second guesses throughout the years have always proved to be accurate.

Sincere thanks are in order for my colleagues at the Paul Menton Centre, who understood that a PhD. thesis is not written in a day! Without their time and effort, I would not have been able to complete this project.

My appreciation goes to my friends and peers, who never lost patience, knowing that whether or not I finished this degree, I would choose to do the right thing for me.

Finally, I would like all of my family to know that their love and sustenance through the years as I pursued my goal was essential in the successful culmination of this dissertation.

As long as I can remember, I will thrive on the knowledge that you all cared enough.

TABLE OF CONTENTS

Abstract	iii
Acknowledgments	v
Table of Contents	vi
List of Figures and Tables	x
 Introduction	 / 1
Visual Recognition Memory	3
Visual Memory	3
Recognition vs. Recall	8
Age Differences in Memory	12
Review of Clinical Recognition Memory Tests	14
Memory Test Considerations	15
Visual Recall Memory Tests	20
Wechsler Memory Scale: Visual Reproduction subtest	20
Wechsler Memory Scale-Revised: Memory Index	24
Benton Visual Retention Test	33
Complex Figures	35
Biber Figure Learning Test	38
Verbal Recognition Memory Procedures	42
California Verbal Learning Test	43
Visual Recognition Memory Tests	46
Randt Memory Test - Picture Reproduction	46
Rivermead Behavioural Memory Test	47
Recognition Memory Test	49
Recurring Figures Test	52
Continuous Recognition Memory Tests	55
Conclusions about Visual Memory Tests	59

Visual Recognition Memory . . .

Table of Contents (cont'd):

Development of a New Set of Visual Recognition Tests	60
Pilot Study #1	63
Pilot Study #2	68
Pilot Study #3	70
Pilot Study #4	75
Pilot Study #5	77
General Discussion	86
Methods / 92	
Participants	92
Materials	92
Procedure	96
Analyses	98
Results / 101	
Demographic Information on Original Ten Age Groups	101
Picture Recognition Memory	107
Normative Information	110
Reliability	113
Standard Error of Measurement	124
Correlated Measures	127
Other memory measures	127
Other neuropsychological factors	129
Depressive symptomatology	133
Word Recognition Memory	137
Normative Information	140
Reliability	144
Standard Error of Measurement	151
Correlated Measures	151
Other memory measures	151
Other neuropsychological factors	153
Depressive symptomatology	157

Visual Recognition Memory . . .

Table of Contents (cont'd):

Abstract Recognition Memory	157
Normative Information	162
Reliability	170
Standard Error of Measurement	178
Correlated Measures	180
Other memory measures	180
Other neuropsychological factors	182
Depressive symptomatology	189
Final Notes on the Three Recognition Memory Battery Tests	191
Discussion / 198	
Age Effects	200
Rates of Forgetting	210
Type of Stimulus Materials	213
Theoretical Implications	219
Clinical Usefulness	225
Implications for Future Research	230
End Notes / 232	
References / 233	
Appendices / 260	
Appendix A: Description of Recruiting Cells	260
Appendix B: Recognition Memory Battery Tests / 261	
B.01 Administration Directions	261
B.02 Recognition Memory for Pictures	262
B.03 Recognition Memory for Words	263
B.04 Recognition Memory for Designs	264
Appendix C: Supplementary Memory Tests Used in This Study / 265	
C.01 G-LAMB Word List	265
C.02 G-LAMB Paragraph	268
C.03 G-LAMB Rectangular Figure	271

Visual Recognition Memory

Table of Contents (cont'd):

Appendix D: Clinical Tests Used in This Study / 274	
D.01	Word Fluency 274
D.02	Trail Making Test A & B 275
D.03	Boston Naming Test 276
Appendix E: Demographic and Screening Scores / 278	
E.01	Recognition Memory for Pictures 278
E.02	Recognition Memory for Words 279
E.03	Recognition Memory for Designs 280
Appendix F: Correlations Among Demographic Variables / 281	
F.01	Recognition Memory for Pictures 281
F.02	Recognition Memory for Words 282
F.03	Recognition Memory for Designs 283
Appendix G: Proportion Remembered Scores 284	
Appendix H: Rank Order of Item Difficulty, Including Percent Wrong by Item Location / 285	
H.01	Recognition Memory for Pictures 285
H.02	Recognition Memory for Words 289
H.03	Recognition Memory for Designs 293
Appendix I: Data for 'High' and 'Low' Depression Subjects / 297	
I.01	Demographic Information for Paired Sample of 'High' and 'Low' Depression Subjects 297
I.02	Differences on G-LAMB scores between 'High' and 'Low' Depression Groups 298
I.03	Means and Standard Deviations on Neuropsychological Measures for Paired Sample of 'High' and 'Low' Depression Subjects 299

LIST OF FIGURES AND TABLES

List of Figures:

1. Sample from Wechsler Memory Scale Visual Reproduction subtest	22
2. Wechsler Memory Scale-Revised, Reproduction subtest	28
3. Sample of Wechsler Memory Scale-Revised, Visual Paired Associates	29
4. Sample from Wechsler Memory Scale Figural Memory subtest	30
5. Sample from Benton Visual Retention Test	34
6. Complex figure drawing tasks	36
7. Stimulus items from Biber Figure Learning Task	40
8. Sample stimulus and response sheets for facial stimuli from Recognition Memory Test	50
9. Samples from Recurring Figures Test	54
10. Sample stimuli from Continuous Visual Recognition Test	56
11. Sample of PRO Test response sheet for Pilot Study #1	65
12. Sample of PRO Test response sheet for Pilot Study #3	74
13. Sample of PRO Test answer page	79
14. Mean trial scores for age groups in Pilot Study #5	81
15. Mean scores for Intact and Impaired groups of matched sample for Pilot #5 . .	83
16. Mean PRO Test trial scores for young and for a matched sample of high functioning old and normal old, and a matched sample of normal old and impaired old	84
17. Sample items from Recognition Memory Battery	94
18. Overall trial scores for Recognition Memory Battery tests	193

List of Tables:

1. Order of Presentation of Tests	97
2. Demographic Information for Ten Age Groups for Entire Sample	102

List of Tables (cont'd)

3. Correlations Among Demographic Variables for Full RMB Sample	104
4. Education Levels for Original Ten Age Groups	106
5. Scores by Age Groups on Recognition Memory for Pictures Measures	109
6. Summary of Stepwise Regressions for Recognition Memory for Pictures Measures with Age Entered as the First Block	114
7. Cumulative Percentiles for Recognition Memory for Pictures Learning and Delay Trials for Normative Sample	115
8. Cumulative Percentiles for Recognition Memory for Pictures Acquisition Scores for Normative Sample	116
9. Rank Order of Item Difficulty for Recognition Memory for Pictures Test	119
10. Analysis of Item Responses by Location on Response Trial Page for Recognition Memory for Pictures (n = 342)	125
11. Variability (SEM) With Age Groups for Each Recognition Memory for Pictures Trial and for Total Acquisition	126
12. Pearson Correlations Between Comparable Recognition Memory for Pictures and G-LAMB Paragraph Measures (n = 300)	130
13. Pearson Correlations Between Comparable Recognition Memory for Pictures and G-LAMB Word List Measures (n = 300)	131
14. Pearson Correlations Between Comparable Recognition Memory for Pictures and G-LAMB Rectangular Figure Measures (n = 300)	132
15. Pearson Correlations Between Recognition Memory for Pictures Measures and Scores on Neuropsychological Factors (n = 229)	134
16. Mean Scores Comparing Recognition Memory for Pictures Measures for 'High' and 'Low' Depression Groups	136
17. Scores by Age Groups on Recognition Memory for Words Measures	139
18. Summary of Stepwise Regressions for Recognition Memory for Words Measures	141

List of Tables (cont'd)

19. Cumulative Percentiles for Recognition Memory for Words Learning and Delay Trials for Normative Age Groups	142
20. Cumulative Percentiles for Recognition Memory for Words Total Acquisition Scores for Normative Age Groups	143
21. Rank Order of Item Difficulty for Recognition Memory for Words Test	145
22. Analysis of Item Responses by Location on Response Trial Page for Recognition Memory for Words (n = 340)	150
23. Variability (SEM) Within Age Groups for Each Recognition Memory for Words Trial and for Total Acquisition	152
24. Pearson Correlations Between Comparable Recognition Memory for Words Measures and G-LAMB Paragraph Measures (n = 298)	154
25. Pearson Correlations Between Comparable Recognition Memory for Words Measures and G-LAMB Word List Measures (n = 298)	155
26. Pearson Correlations Between Comparable Recognition Memory for Words Measures and G-LAMB Word List Measures (n = 298)	156
27. Pearson Correlations Between Recognition Memory for Words Measures and Neuropsychology Measures	158
28. Mean Scores Comparing Recognition Memory for Words Measures for 'High' and 'Low' Depression Groups	159
29. Scores by Age Groups on Recognition Memory for Designs Measures	161
30. Summary of Stepwise Regression for Recognition Memory for Designs Measures	164
31. Demographic Information for Final Three Age Groups for Recognition Memory for Designs Test	167
32. Performance on Recognition Memory for Designs for Three Age Groups	169
33. Cumulative Percentiles for Recognition Memory for Designs Learning and Delay Trials for Three Age Groups	171

List of Tables (cont'd)

34. Cumulative Percentiles for Recognition Memory for Designs Total Acquisition Scores for Three Age Groups	172
35. Rank Order of Item Difficulty for Recognition Memory for Designs Test	174
36. Error Rate for Recognition Memory for Designs Item Responses by Location on Response Trial Page	179
37. Variability (SEM) Within Age Groups for Each Recognition Memory for Designs Trial and for Total Sample	181
38. Pearson Correlations Between Comparable Recognition Memory for Designs and G-LAMB Paragraph Measures (n = 328)	183
39. Pearson Correlations Between Comparable Recognition Memory for Designs and G-LAMB Word List Measures (n = 328)	184
40. Pearson Correlations Between Comparable Recognition Memory for Designs and G-LAMB Rectangular Figure Measures (n = 328)	185
41. Pearson Correlations Between Recognition Memory for Designs and Scores on Neuropsychological Tests (n = 292)	187
42. Percent of Variance Explained Using Recognition Memory for Designs Measures and Scores on Neuropsychological Tests (n = 292)	188
43. Percent of Variance Explained Using Recognition Memory for Designs Measures and Scores on Neuropsychological Tests with Age and Vocabulary Levels Controlled (n = 292)	190
44. Comparison of 'High' and 'Low' Depression Groups on Recognition Memory for Designs Measures	192
45. Percent Correct on Trials of Recognition Memory for Designs Items for Three Age Groups	195
46. Percent Correct on Trials of Recognition Memory for Pictures and Recognition Memory for Words by Trials	197

INTRODUCTION

When undertaking an exploration of memory ability in an applied setting, the majority of tests available to the clinician concentrate on the assessment of free recall, both of auditory-verbal and visual information. Three problems arise from this unitary emphasis. First, research has shown that free recall is only one method of assessing memory ability. In order to sort out the contribution to a memory deficit of encoding and retrieval, one also must assess cued recall and recognition memory (Branconnier, Cole, Spera & DeVitt, 1982; Brandt, Corwin & Krafft, 1992; Buschke, 1973, 1984; Craig & Simon, 1980; Delis, Kramer, Kaplan & Ober, 1987; Grober, Buschke, Crystal, Bang & Dresner, 1988; Storandt, Botwinick & Danzinger, 1986; Tombaugh & Schmidt, 1992; Wilson & Kaszniak, 1986). In recent years, some memory batteries have included cued memory tasks and recognition memory procedures when assessing auditory-verbal retrieval (e.g., California Verbal Learning Test: Delis, Kramer, Kaplan & Ober, 1987; Learning and Memory Battery, or LAMB, Tombaugh & Schmidt, 1992; Wechsler Memory Scale-Revised: Wechsler, 1987). However, few visual memory tests extend beyond the measurement of free recall and, in addition, tend to be constructional in nature. When a visual recognition memory test is included, it can be more accurately described as a visual recognition procedure appended after recall memory for that same material has already been explored. The second problem with emphasizing free recall is that it does not allow for the assessment of different codes for memory. De Renzi (1968) concluded that, whenever possible, individuals attempt to make use of verbal mediation to encode

information. This makes the assessment of a visual code difficult to separate from auditory-verbal encoding. Although experimental visual memory tests often employ complex abstract stimuli to reduce the involvement of previously encoded semantic materials, clinical tests tend to use more meaningful visualizations, such as recognizable geometric shapes, in order to provide a level of 'face validity' and encourage ongoing involvement of the individual being assessed. A third difficulty is that because the individual must reproduce visual material by drawing it, free recall of visual information involves visuoconstructional and motor components, further confounding the measurement of "memory" (Glosser, Goodglass & Biber, 1989; Larrabee, Kane & Schuck, 1983; Ryan, Morrow, Bromet & Parkinson, 1987; Silverstein, 1962). In addition, it is difficult to evaluate the drawing of abstract nonsense designs.

One way of extending the usefulness of a memory battery is to include or supplement with tests that explore recognition memory. Recognition memory can be used to assess various components (e.g., verbal, non-verbal) while reducing the confounding effects of visuospatial, visuoconstructional, aphasia, and motor deficits. Like tests measuring recall, a recognition test can be used as a learning test for assessing acquisition rates as well as for measuring immediate and delayed memory. This thesis will look at the reliability and validity of a set of newly developed recognition tests of learning and memory using three different stimuli but employing comparable formats: the Recognition Memory Battery, made up of the Recognition Memory for Pictures (RM Pictures) Test, Recognition Memory for Words (RM Words), and Recognition Memory for Designs (RM Designs). Although the items in all three of these tests are presented visually, each test

was designed to assess recognition memory for a variety of materials, verbal as well as non-verbal, each of which may be encoded differently.

First, a short review of research on visual recognition memory will be presented. Following this, a review of the most commonly used clinical visual memory tests will be undertaken. Then the development of the present set of visual recognition memory tests will be presented, beginning with the initial format, the Picture Recognition of Objects (PRO) Test.

Visual Recognition Memory

Visual Memory

When memory problems are suspected, clinicians have learned that they can point to the underlying difficulty much more accurately when they examine the different codes which are used to store memories. Although neuropsychological studies have had difficulty discovering the exact location or pathways that these memories inhabit, it is clear that separate codes can be used for verbal and non-verbal information (e.g., Milner, 1968; Warrington, 1984; Warrington & James, 1967; Trahan, Larrabee, and Quintana, 1990). Researchers have tended to concentrate on various verbal and visual stimuli to examine these codes. This has resulted in confused findings surrounding localization studies, primarily because of terminology; it is quite a jump from a verbal/non-verbal encoding paradigm to a verbal/visual retrieval conceptualization.

“Visual memory” is an eclectic term that can be used to refer to memory for any information received through the visual system. During testing, this information may be

presented in a variety of forms, such as linguistic material (e.g., letters, words, or numbers), visuospatial information (e.g., a map, location of objects in a room), three-dimensional objects, photographs or movies of faces, pictures (e.g., line drawings or photographs of recognizable objects, or designs (e.g., drawings of geometric forms or abstract shapes). It is expected that linguistic information and easily-verbalized pictures are more readily assigned verbal codes, while more complex visual information, which is less easily encoded and processed verbally, accesses the non-verbal (or 'visual') system.

In general, most visual memories are felt to be more accessible than most verbal memories, whether due to deeper encoding (Lockhart & Craik, 1990), more efficient retrieval (Tulving & Thompson, 1973; 1984), or to the ease with which an automatic *dual encoding* process is used (Craik & Lockhart, 1972). However, both experimental and clinical literatures have supported the contention that a fundamental factor underlies both verbal and visual episodic memory (memory for newly learned material, the format most commonly used for clinical assessment of memory) (Borinstein & Chelune, 1988; Skilbeck & Woods, 1980; Wilson, Cockburn & Baddeley, 1985). This is particularly true when immediate recall is assessed (Delaney, Rosen, Mattson & Novelly, 1980). For the most part, episodic memory is tested with a free recall paradigm in clinical settings. In the past, this method reflected the standard form of reporting, which was to present memory as a unitary measure. In fact, when factor analyses have been conducted on test batteries purporting to measure various types of memory for a variety of memory impaired patients, verbal and visual memory loaded on the same factor (Loring, 1989; Tombaugh & Schmidt, 1992; Trahan et al., 1990). Clinical assessment has begun to reflect experimental findings and theoretical approaches as well as acknowledging memory as having several

codes (visual, verbal, and semantic) it includes various processes (encoding, storage, and retrieval). And within these processes, additional information about the individual's memory ability can be gleaned from making use of more than one procedure for assessing memory (free recall, cued recall, and recognition).

If neural damage is pervasive enough, a general memory deficit may be exposed regardless of the type of material presented for (learning and) memory. But if a localized deficit is suspected, whether severe or mild, the type of test chosen to assess that possibility is quite important to consider. There seems to be little question that a dissociation between performance on verbal and visual memory tasks occurs frequently in clinical practice (Delaney et al., 1980; De Renzi, 1968; Kimura, 1963; Mayes, 1987; Russell, 1975; Trahan et al., 1990; Trahan, Larrabee, Quintana, Goethe, & Willingham, 1989), even if the exact anatomical location of memory specific functions is still debated. Some studies have suggested that visual memory is associated more closely with functioning of the right cerebral hemisphere (e.g., Delaney et al., 1980; De Renzi, 1968; Kimura, 1963; Milner, 1968, 1971), while others have found that patients with left-hemisphere brain lesions perform more poorly on visual memory tests than do patients with right-hemisphere lesions (Boller & De Renzi, 1967; Boller & Spinnler, 1967; De Renzi, 1968; De Renzi & Spinnler, 1966). Still other researchers have found that although patients with (right or left) hemisphere lesions did more poorly than control subjects on tests of visual memory, these scores provided little support for lateralized damage (Brown, Sawyer, Nathan, & Shatz, 1987; Delaney et al., 1980; Trahan, Quintana, Goethe & Willingham, 1988).

In addition, although information from the various forms of stimuli mentioned at the beginning of this section shares a common code, each appears to involve slightly different neural networks (Lezak, 1983). It has been suggested that much of this localization discrepancy may be due to the methodology used when assessing visual memory (Glosser et al., 1989; Trahan et al., 1988). Two major factors are important in sorting out these methodological difficulties: the contribution of verbal codes to memory, and the confounding of visuospatial and visuoconstructional factors with memory.

The search for a truly "visual" memory test has been an important area of neuropsychological research (e.g., see Heilbrunner, 1992). Zangwill (1976) has pointed out that people process information through various means; some are verbalizers and others are visualizers. Many of the more widely used measures of visual memory such as the tests of the WMS-R Visual Memory Index, the Visual Retention Test (Benton, 1974), and the recall administrations of the various Complex Figure tasks (Milner, 1975; Osterrieth, 1944; Rey, 1941; Taylor, 1969, 1979), involve the use of fairly simple designs and geometric shapes (circles, lines, triangles) in order to facilitate drawing for recall. As a result, a large portion of the retrieval process may be undertaken in the verbal memory system. The most common way to avoid this verbal encoding is to use 'nonsense' figures as stimuli. However, if complex abstract shapes are used, the stimulus material may be too difficult for brain-damaged patients to recall (Cermak & Tarlow, 1978).

The experimental literature has presented a variety of other strategies to reduce the effects of verbal encoding for visual memory tasks. One is to reduce the exposure time allowed for viewing the stimulus. The less time available, the more the viewer is apt to rely

solely on visual memory. Trahan et al. (1990) recommend that a three-second exposure time be used to discourage verbal encoding. But De Renzi (1968) suggested that patients may be able to memorize even nonsense figures by encouraging them to attribute a verbal meaning to this type of visual information. In fact, Vanderplas & Garvin (1959) demonstrated that even when using complex random polygons, many subjects can readily attach verbal labels in periods as brief as three seconds. However, irrefutable results of research shows that for older individuals, memory ability can be confounded with slower processing times, depending on the research design (e.g., Ferris, Clark, Sathananthan, & Gershon, 1976; Flicker, Ferris, Crook, & Bartus, 1987; Welford, 1977).

A second element affecting the degree to which a visual factor is verbalized is the class to which distractor items belong. If the stimulus item was a rowboat or a shaded pentagon, another item from the same class (e.g., a canoe or a shaded octagon) is apt to be more difficult to discriminate than if the distractor item is from a different class (a car or an open figure). In addition, exposure times interact with the class of the item; it takes more time to encode a same-class item in a way that allows discrimination from other response items than if a separate-class item were used.

A third factor contributing to confused localization findings relates to the format generally used to assess visual memory. In order to equate free recall methods in visual memory tests with verbal formats, the use of visuoconstructional and visuomotor responses is required. This greatly complicates interpretation regarding the location of the deficit, and indeed whether the deficit scores are due to memory or to constructional ability, visuospatial memory, visual attention, primary visual perception, or an interaction

of any or all of these distinct factors. As mentioned previously, when large numbers of neuropsychological tests are submitted to factor analysis, a "visual" and a "verbal" factor may emerge, but the visuospatial and constructional components called upon in free-recall testing of visual memory are largely responsible for the "visual" memory factor.

One way to reduce the effects of visuospatial/visuoconstructional and verbal encoding deficits when testing for visual memory is to incorporate a delayed memory procedure (Daniel & Ellis, 1972; De Renzi & Spinnler, 1966; Ellis & Daniel, 1971; Heilbrunner, Buck & Adams, 1989; Larrabee, Kane, Schuck & Francis, 1985; Lee et al., 1989; Skilbeck & Woods, 1980; Trahan et al., 1990). A second way is to assess visual memory using a recognition format, which can also include an immediate and a delay procedure.

Recognition vs. Recall

Knowledge of the organization of human memory has benefited from research examining performance measures, such as recognition and recall, and their functional differences. But there is still debate about the relationship between recognition and recall. Are they part of the same process or do they involve separate processes?

Initial theories concentrated on the similarities of recognition and recall. One early model proposed that recall requires a greater strength (that is, more connections in storage to that information) than does recognition memory (e.g., Postman, 1963). The generate-recognize model (Anderson & Bower, 1973; Bahrick, 1970; Rabinowitz, Mandler, & Barsalou, 1979) presented recall and recognition memory as part of a single two-stage process which differed only during the recognition phase (Kintsch, 1978). According to

this theory, when an individual is exposed to a stimulus and later asked to recall it from memory, information must first be generated and provided to the working memory system, and then a decision must be made as to whether or not the information produced is the same as that previously presented. Recognition memory, however, merely accesses the decision portion of this formula, as the 'generation' portion of this dual process is replaced with the external presentation of the stimuli at the time of retrieval. Watkins and Gardiner (1979) pointed out that this theory fails to account for the weight of empirical findings (see Mandler, 1980, for an early review of this literature).

Other accounts from the literature view recognition and recall as a dual-process model: that is, recall and recognition represent another level of processing within the retrieval operation (Atkinson & Juola, 1974; Johnston, Dark & Jacoby, 1985; Mandler, 1980; Haist et al., 1992). This theory represents the two forms of retrieval as working in a parallel, but independent manner, each with separate, but related, functions. One example of a dual encoding theory is Tulving's (1983, 1985) *encoding specificity* principle stating, in simple terms, that successful retrieval is dependent on the match between the encoding of a stimulus at the time of learning and the information that is available at the time of retrieval. The more similar the cues at retrieval to those present at the time of encoding, the more successful is that retrieval. According to this theory, recall is usually more difficult than recognition: in order to access the correct information at the time of retrieval, recall requires more specific cues.

It has also been proposed that recognition memory may be a type of implicit, unconscious, or automatic memory (see Fagan, 1984; Hasher & Zacks, 1979; Haist et al.,

1992). This model proposes that while recall memory relies on declarative, or conscious processes, recognition memory depends on perceptual priming, which is an unconscious perceptual process improving the chance that a stimulus will be recognized in the future (Shimamura, 1986; Tulving & Schacter, 1990). However, using a different methodology than that employed during previous research, Haist et al. reported that recognition memory was not supported by automatic processing.

While retrieval appears to be easier when using recognition tasks than for comparable free recall, or even cued recall tasks, there are decrements seen in recognition performance in normal aging (see the following section for a short synopsis of memory and aging as it relates to this thesis) as well as in clinical samples (Shimamura & Squire, 1988; Squire & Shimamura, 1986; Trahan et al., 1990). In addition, there is support for the concept that the ease of retrieval in either type of situation depends a great deal on the task demands (Jacoby & Dallas, 1981; Lee et al., 1989). These variables include the type of material presented (Meudell & Mayes, 1981; Warrington, 1984; Warrington & James, 1967), the pace of the task (Delaney et al., 1980; Flicker et al., 1987), and the number of exposures to the information before it is accessed from long-term memory (Buschke, 1973; Delis et al., 1987; Trahan et al., 1990). In fact, given the right conditions, recognition memory can be more difficult than recall (Muter, 1984).

One obvious advantage of a recognition format is to reduce the effects of other cognitive and motor skills and abilities (Glosser et al., 1989; Larrabee et al., 1983; Pheley & Klesges, 1986; Trahan & Larrabee, 1984). This is particularly important in two situations: first, when specific right-hemisphere damage is suspected and non-verbal

memory is being assessed, and second, when traditional measures of memory assessment (i.e., free recall) may be confounded by motor difficulties, expressive language deficits, or visuospatial interference.

The traditional neuropsychological method of identifying specific hemisphere structural damage is to administer tests that theoretically measure performance supported by that hemisphere. For example, language ability is known to reside for most individuals in the left hemisphere. Therefore, language-based memory tests should result in deficit scores for those with unilateral damage in specific areas of the left hemisphere, but not for patients with unilateral right-hemisphere damage. Results such as this support a double dissociation hypothesis (e.g., see Delaney, Rosen, Mattson, & Novelly, 1980) in that unilateral right temporal lobe damage results in impaired visual memory but preserved verbal memory, while unilateral left temporal lobe damage results in impaired verbal memory but preserved visual memory. Of course, both right and left hemisphere damage may exist that could result in deficits on both types of memory tests. Numerous results from testing this hypothesis, however, reveal that while low scores on verbal memory tests for the most part reflect unilateral left-hemisphere damage, most visual memory tests yield reduced scores for patients with both left and right hemisphere damage compared to healthy controls (Benton, 1985; De Renzi, 1968; Glosser et al., 1989; Lee et al., 1989; Warrington, James & Kinsbourne, 1966). This may mean that the focus for all memory management is in the left hemisphere, but this interpretation does not explain results that do support the double-dissociation hypothesis. A more likely possibility is that tests used to assess visual memory are confounded by other factors, such as constructional

difficulties (left-hemisphere damage often results in semi-paresis, which could reduce the score of right-handed patients when measured by drawing tasks) or verbal encoding of pictures and designs. Recognition procedures help in avoiding this type of confound.

In addition, standard auditory-verbal memory testing are apt to be difficult to administer to an individual with expressive language deficits. Memory may be intact, but the individual is unable to repeat the information that had been presented. A recognition procedure using verbal information presented visually that involves pointing to a correct answer could avoid the need to test orally for verbal memory.

Finally, since tests presenting visual information have been shown to reflect deficits in both hemispheres, visually presented recognition memory tests could identify whether or not underlying memory deficits also exist when an individual has motor difficulties.

Age Differences in Memory

Regardless of age, large individual differences exist in performance on all types of memory testing, but several findings are consistent about aging and memory. Once the age of 60 has been reached, older people, as a whole, tend to perform less well on tests of recall than do their younger counterparts (Berry & Carpenter, 1992; Craik, 1977; Craik & Simon, 1980; Erber, 1974; Glosser et al., 1989; Hart & O'Shanick, 1993; Perlmuter, 1979; Rybarczyk, Hart & Harkins, 1987). This is particularly true when a time component is involved during presentation and/or retrieval of the material (Korsnes & Gilinsky, 1993; Craik & Simon, 1980; Perlmuter, 1979). Part of this may be due to a general slowing down of processing speed normally found with aging, but part may be due to the fact that older individuals on the whole do not automatically use learning and memory strategies.

When strategies are taught or suggested, they are successfully employed, but in normal functioning, it appears that older people generally rely more on environmental cues, such as calendars, lists, visual cues, and so on.

Although this same age phenomenon is present in recognition memory (Bartlett, Till & Levy, 1980; Cullum, Butters, Troster & Salmon, 1990; Erber, 1974; Ferris, Crook, Clark, McCarthy, & Rae, 1980; Flicker et al., 1987; Gordon & Clark, 1975; Hannay, Levin & Grossman, 1979; Harkins & Chapman, 1976; Perlmutter, 1979; Trahan et al., 1990; Winograd, Smith & Simon, 1982), it seems to exist to a lesser degree than when a free recall format is used (Haist et al., 1992; Measso, Romani, Martini & Zappalà, 1990; Menich & Baron, 1990; Riege & Inman, 1980; Trahan, Larrabee & Levin, 1986). In addition, scores on memory 'span' tests are relatively equivalent across age (Craik & Simon, 1980; Wechsler, 1981), and semantic information appears to be more easily accessed as age increases (Craik & Simon, 1980; Gordon & Clark, 1975). These findings combine to indicate that for the most part, lower levels of performance are present for older individuals on episodic tasks involving newly learned material that is transferred to long-term memory and subsequently re-assembled in working memory.¹ For these reasons, it is important to include normative information for each age group when standardizing a clinical memory test used for a variety of populations.

The reduction in scores on these tasks found with increasing age also appears when memory is assessed because of disease and conditions involving memory deficits. Because recognition memory appears to remain relatively similar across ages, it is

¹ Age differences also tend to disappear when episodic information is overlearned, but clinical tests seldom have the luxury of time to ensure that overlearning takes place.

therefore more suitable for exploring the degree of learning and memory ability present. If recall memory is not appropriate (because of floor effects, verbal or motor deficits, or language differences), recognition memory allows an exploration of at least a general memory ability, and more specifically, retained ability to attend to and learn from environmental cues.

Before presenting the development of a new visual recognition memory test, it is important to understand the choices a clinician has at the present time. The following section will review a representative sample of available clinical tests.

Review Of Clinical Recognition Memory Tests

While a wide variety of visual memory tests has been developed for experimental purposes, relatively few fit the category of visual recognition memory tests and fewer still are used in the clinical testing of memory. Some are included as part of a larger battery to screen for various cognitive deficits due to disease, open/closed head injury, or other focal damage causing a variety of cognitive disabilities, such as dementia. These include subtests in the CAMCOG Dementia Examination (Roth, Tym, Mountjoy, Huppert, Hendrie, Verma, & Goddard, 1986), the Mattis Dementia Rating Scale (Mattis, 1976), the Mini-Mental Status Examination (Folstein, Folstein & McHugh, 1975), or the Halstead-Reitan Neuropsychological Assessment Battery (originally introduced by Halstead in the late 1940's and last revised in 1974). Visual memory subtests or specific items from these batteries are often used in epidemiological research to screen for memory problems, even though validity and reliability measures unique to subtests, let alone individual items, are seldom available.

Larger memory assessment batteries often include visual memory tests using either a free recall or recognition format, but seldom both. Individual visual memory tests also concentrate either on one or the other of these processes.

Memory Test Considerations

Before evaluating the various recognition memory tests used by clinicians, it is important to understand what constitutes a good memory battery. In clinical practice, when a memory deficit is suspected, a battery of memory tests will be administered to support the absence or presence of a possible memory deficit. This battery is also used to determine the areas where specific deficits originate and/or processes that they affect. The clinician does not often administer a specialized memory test without first administering the battery. An ideal memory battery (1) would contain a variety of tests involving verbal and non-verbal materials which often are inaccurately termed verbal and visual, (2) would include measures to evaluate the contributions of learning rates as well as of free, cued, and recognition memory, (3) would incorporate subtests evaluating memory span, immediate and short-term memory, and delayed recall, and (4) would be used in conjunction with tests of attention/concentration and IQ. Of course, it is important that each subtest be a valid measure of those skills or deficits.

Ritchie (1986) presented five concerns when developing or evaluating the usefulness of any cognitive test. They include (1) the necessity of basing the test on theoretical concerns, (2) the importance of test format, (3) a careful choice of statistical analyses, (4) consideration of the measurements chosen to validate the new test, and (5) the reporting of reliability measures.

The first point Ritchie makes is that the test should be developed on the basis of theoretical considerations, not just clinical observations. In the case of visual memory, for example, do recognition and recall reflect different retrieval processes or are they points on the same continuum? Do learning, immediate recall, and delayed recall measure the same construct or are they parts of different constructs? And where do the concepts of motivation and attention fall in the importance of visual memory testing?

Verbal tests often are comprised of two or more learning trials, and occasionally have normative information on delayed (long-term) memory. However, with the exception of the LAMB (Tombaugh & Schmidt, 1992), visual tests in most 'comprehensive' memory batteries do not examine learning across various types of stimuli as well as various time periods; they assume the information presented was learned during a single exposure. 'Memory' for that information is tested immediately after a single presentation, and occasionally after a delay period. Pathological diagnoses, on the other hand, often depend on learning styles and patterns of learning as well as on differences in memory ability and rates of forgetting (Moss, Albert, Butters & Payne, 1986; Squire, 1987). The following examples may serve to highlight these differences. In a normal population and given the right conditions, such as self-pacing, performance for older adults on learning tasks can be brought up nearly to the level of younger adults, although perhaps not as rapidly (Craik & Simon, 1980). Both immediate and delayed free recall performance by those in older adults is significantly reduced on the average, compared to their younger counterparts (Craik & Simon, 1980; Cullum et al., 1990; Measso et al., 1990; Perlmutter,

1979). The addition of a recognition format, on the other hand, often raises memory performance of older adults closer to levels of younger adults.

Pathology leading to memory problems presents different patterns of performance on memory tests. In the case of Alcoholic Korsakoff's Syndrome, immediate memory is normal, but patients with this syndrome have a great deal of difficulty retaining new information over a series of trials, and then cannot recall it after a delayed period of time, even with recognition cues (Kopelman, 1985; Moss et al., 1986). In situations of more localized brain damage resulting from open or closed head injury, such as various agnosias, specific neuropsychological systems are affected while others remain intact. One example is face agnosia, which involves an inability to recognize faces while recognition of objects or other identifying features such as posture or birthmarks may be unaffected (Damasio, Tranel, & Damasio, 1989; McNeil & Warrington, 1993). On the other hand, both episodic learning and recall of semantic memories are impaired in moderate to more severe forms of memory loss such as Alzheimer's dementia (Gainotti & Marra, 1994; Morris & Kopelman, 1986). These differences serve to strengthen the need for exploration of the various memory processes to provide a differential diagnosis, and to develop treatment plans.

A second important consideration concerns the format of the test. Are the items too easy so that ceiling effects are produced, or too hard so that floor effects are present? Flicker et al. (1987) point to the need for a broad range of responses. Various researchers have stressed the importance of supplemental information surrounding the correct and incorrect answers in determining a differential diagnosis (e.g., see Delis, Massman,

Butters, Salmon, Cermak & Kramer, 1991; Osborne, Brown & Randt, 1982; Ritchie, 1988; Trahan et al., 1990). As a result, the format should leave room for providing information on how the items are answered, such as response time, false negatives and false positives, refusal to answer. Is this factor taken into consideration in the development of the test? Are the items chosen so that answers are not confounded with physical disability and/or demographic factors such as socioeconomic status and education? Do normal memory patterns correlated with aging interfere with an accurate interpretation of the results? For example, clinical tests of auditory-verbal memory usually involve everyday meaningful stimuli, but visual memory tests generally employ designs or nonsense figures in order to reduce verbal mediation and the effects of individual differences in drawing ability. This may bring into question the face validity of the test, with the possibility of reduced effortful performance in a clinical situation (Flicker et al., 1987). Another format considered in visual memory testing involves the use of drawing tasks to assess visual memory. With this type of assessment, more testing time is required, scoring is relatively complicated, and interrater reliability is generally reduced (Lezak, 1983; but see Tombaugh & Schmidt, 1992).

A third consideration presented by Ritchie relates to the data analysis of the results of the test. Do the norms take into consideration the age and IQ of the patient? If one fact about aging and memory is predominant, it is the certainty that performance decreases with increasing age on most episodic memory tasks (Perlmuter, 1979).

Fourth, how was the test validated? A memory test, particularly one that will likely be used with a relatively older population, should include normative data from a

large number of volunteer subjects from a wide variety of age groups. The *face validity* of this sort of test is especially important, as this intact normative population will tend to scrutinize the test and be unwilling to participate if it appears to be assessing something other than what is stated. *External validity* depends on other clinically recognized procedures that define memory ability. The validating procedure provides valuable information about the sensitivity and specificity of a new screening test. Sometimes a test is validated against similar existing tests. If so, that validating test should be carefully chosen. It does not make sense to justify the development of a new superior instrument on the basis of the inadequacy of available tests, and then use those same established tests as criterion measures for concurrent validity. A better measure of external validation is provided by outcome data based on clinical cases. This *predictive validity* is only available when conducting longitudinal studies, and for reasons of limited resources, is often bypassed.

Finally, Ritchie stresses that the reliability of a test should be carefully examined. Inter-rater, retest, and internal reliability are important especially for screening tests. The more reliability information available to the clinician or researcher, the more appropriate will be his or her choice of tests for a given population.

In addition, Erickson and Scott (1977) stress that when testing for memory, it is important to include measures of delayed memory as well as learning and immediate memory. They allege that most memory tests measure perception and learning rather than retention, and they emphasize the importance of including a delayed memory period long enough to allow encoded information to be integrated with semantic information in long-

term memory (see also Delaney et al., 1986; De Renzi, 1968; Tombaugh & Schmidt, 1992; Trehan & Quintana, 1990).

It does not take a great deal of perusal to discover that no single test addresses all of these criteria, but the clinician should be aware of the various strengths and weaknesses of each of these criteria and take them into consideration when interpreting whatever test has been chosen for use.

Although the emphasis of this thesis is on visual recognition memory, it is important to understand how clinical memory tests currently assess visual memory. To this end, the most popular memory battery will be reviewed, followed by the most commonly used clinical tests of visual memory (usually free-recall tests with a structural component). Recognition memory for verbal information usually consists of administering a recognition test after free-recall testing is completed. Therefore, a representative of this type of test is reviewed. Finally, tests used by clinicians that specifically assess recognition memory will be presented.

Visual Recall Memory Tests

The most commonly used visual reproduction or free recall memory tasks are the Visual Reproduction (VP) test from the Wechsler Memory Scale (WMS; Wechsler, 1945), tests from the Visual Memory Index included in the WMS-R, the Benton Visual Retention Test (BVRT; Benton, 1974), and the Taylor and the Rey-Osterrieth Complex Figure tests (Milner, 1975; Osterrieth, 1944; Rey, 1941; Taylor, 1969, 1979).

Wechsler Memory Scale: Visual Reproduction subtest. The Wechsler Memory Scale (WMS) and its revised form is probably the most widely employed memory test

battery in use today (Brinkman et al., 1985; Herman, 1988; Roid, Prifitera & Ledbetter, 1988). In its original form, the WMS included the Visual Reproduction (VR) task, a subtest measuring immediate recall of four designs (see Figure 1). In this subtest, four shapes drawn in ink are presented on a card, each for 10 seconds. Immediately after viewing each card, the individual is asked to draw the figure from memory on a blank piece of paper.

 Insert Figure 1 about here

The 14-point score makes up approximately 15% of the total WMS score. An age-correction score is added, and then a table is used to determine a Memory Quotient (MQ), which is comparable to the WAIS Intelligence Quotient in that the mean score is 100 with a standard deviation of 15 points. Thus, although the WMS includes 5 other subtests, the battery presents 'memory ability' as a unidimensional score that, besides 'visual memory ability', equally includes orientation to time, place and person, verbal span, immediate verbal recall, drawing competency, and mental tracking. Factor studies of the WMS consistently have indicated that, in fact, the WMS score consists of three main factors. These include a general learning and recall factor (incorporating both verbal and visual subtests), and weaker factors that include attention/concentration and 'mental control' (see Skilbeck & Woods [1980], for a review of these factor studies). A general conclusion has been drawn that the Visual Reproduction task of the WMS is not a test of

Figure 1. Sample from Wechsler Memory Scale Visual Reproduction Subtest.

visual memory (also see Larrabee et al., 1985). And, in fact, norms are only available for the overall MQ and not for the various subtests that make up this memory battery.

Finally, no delayed memory tasks are included in this battery. Thus, it appears that although the WMS is useful for calling attention to memory dysfunction (Wechsler, 1945), its use, for the most part, is restricted to an assessment of immediate verbal memory (Black, 1973; Milner, 1975; Shapiro, Sadowsky, Henderson & Van Buren, 1973). It is insensitive to the detection of lateralized damage or for identifying various types of memory impairment (Prigatano, 1978; Skilbeck & Woods, 1980; Squire, 1987).

Yet, it appears that the VR subtest has been granted a degree of face validity. When Russell (1975) attempted to correct some of the above problems, he included only Logical Memory and Visual Reproduction in his revision of the WMS. He identified these two subtests as measures of immediate recall for assessing 'semantic' and 'figural' memory. He expanded the scoring scheme to include half-points for partially correct responses. In addition, this revised test allows the participant to be prompted if he/she is unable to remember any of the figures. Russell also included measures of 30-minute delayed retention and of percent retained. When looking at the usefulness of this version of the WMS, test-retest reliability was found to be quite high (in the .83 to .96 range. Subsequent studies using Russell's version included norms for older age groups (Bak & Green, 1980; Cauthen, 1977; McCarty, Logue, Power, Ziesat & Rosenstiel, 1980) and various studies were conducted to assess the usefulness of this revision with elderly demented patients (e.g., Brinkman et al., 1983; Logue & Wyrick, 1979). Studies of the criterion validity of Russell's version of the WMS indicated that although the delayed

reproduction of the visual memory task is a more valid measure of memory deficits than immediate recall, it still reflects visual/perceptual/motor difficulties more than visual memory deficits (Brinkman et al., 1985). Despite the addition of a delayed recall procedure, the same three-factor structure was maintained for Russell's revised version of this test (Skilbeck & Woods, 1980).

In summary, the WMS provides a unitary score to measure both general memory ability and general awareness/orientation for a limited age range of individuals. Russell's revision may be more useful in identifying types of memory problems relating to localized brain damage. Normative data for this revision is more comprehensive than for the original WMS. Although the Visual Reproduction subtest is employed regularly as a test of visual memory in experimental studies, its usefulness is limited for clinical memory testing, since it appears to be related more to visuospatial and constructional abilities than to memory, *per se* (Lezak, 1983).

Wechsler Memory Scale-Revised: Visual Memory Index. In 1987, the Wechsler Memory Scale-Revised (WMS-R) was normed and presented for clinical use. Modified and weighted scoring, the inclusion of older age groups in its normative sample, increasing the number of visual items, and adding tests of delayed memory were changes meant to answer to some of the more critical reviews of this frequently used clinical memory battery (Herman, 1988).

To reflect research revealing the multidimensionality of memory, the WMS-R was designed to encompass five scales, including the General Memory Index, which consists of the weighted sum of scores from two other indexes, the Verbal Memory Index and the

Visual Memory Index. The test manual reports Index score equivalents of sums of weighted raw scores for both Visual Index and Verbal Index tests. However, the validity of neither of these WMS-R indexes was supported in factor analytic studies (Bornstein & Chelune, 1988; Herman, 1988; Loring, 1989; Loring, Lee, Martin & Meador, 1989; Roid et al., 1988; Weschler, 1987). Rather, as with the Wechsler Memory Scale, a General Memory factor including both visual and verbal subtests was found. A second factor was termed the attention/concentration factor, which makes up another Index in the WMS-R. Bornstein and Chelune (1988) presented a slightly different factorial structure for a psychogeriatric sample, but all of the tests of immediate retrieval, including the visual recognition test (Figural Memory) still loaded on a General Memory factor. In a review of the WMS-R literature, Elwood (1991a, 1991b) attributed this lack of subscale score reliability to high standard error, low reliabilities, and large standard error of the means. He stated that only the General Memory and Attention/Concentration scales meet even the most liberal standards for reliability, and that any other subscale score can only provide a general approximation of memory ability.

Although Bornstein and Chelune intimated that the WMS-R was only able to assess global memory impairment, Loring et al. (1989) felt the inclusion of visual-verbal discrepancy scores in the test manual represents an implicit suggestion that the Visual and Verbal Memory Indexes are designed to test lateralized memory dysfunction. In their validity study of WMS-R using patients with temporal lobectomies, material-specific deficits on tasks involving verbal (the Selective Reminding Task: Buschke & Fuld, 1974)

and visual (Complex Figure²; Lezak, 1983) memory. They found that the WMS-R was not sensitive to these deficits. Loring et al. found that left hemisphere damage could be assessed through either WMS-R auditory or visual codes, but that 25% of the right hemisphere lobectomy patients were not identified correctly through the visual memory tests of the WMS-R. In fact, using a series of three verbal/visual index discrepancy scores, they found that two of their right hemisphere patients had scores nearly reaching significance levels, but in the wrong direction. Loring et al. concluded that because the WMS-R shows relatively high correlation with intelligence (Bornstein & Chelune, 1989) and is sensitive to a general memory factor, the WMS-R is a good measure of cognitive decline, and because it includes delayed memory measures, it can reflect the differentially affected impaired episodic memory deficits of amnesia. However, it is not sensitive to material-specific memory losses, particularly when those specific losses involve visual memory, and warned that caution should be used in interpreting Verbal/Visual Memory Index discrepancies. Similar findings were reported by Naugle, Chelune, Cheek, and Lunders (1993) in a study of right temporal lobe epileptics.

Perhaps the reason for this relative insensitivity to material-specific memory losses lies in the type of visual subtests used in this scale. The Visual Memory Index adds together the scores of three subtests: Visual Reproduction, Visual Paired Associates, and Figural Memory. The WMS Visual Reproduction subtest was expanded in the WMS-R to include five figures instead of four, and scoring was structured and expanded to encompass a total of 41 points instead of 12 (see Figure 2). The figures still consist of

² Lezak reviews two separate Complex Figures in her text, the Rey Complex Figure and the Taylor Complex Figure. It is not clear from Loring et al.'s papers which of the two figures was used in their study, but it is likely to have been the Rey Figure, since that is the one most commonly used.

abstract designs made up of geometric shapes or lines. Inter-rater reliability for this improved scoring scheme was reported to be 0.97.

 Insert Figure 2 about here

The second test in this index, Visual Paired Associates, requires that the individual learn a set of six figure-colour pairs so that when the figure is subsequently shown, the participant can correctly point to the colour with which it was first presented (see Figure 3). Both of these subtests have a delayed memory trial, with free recall for Visual Reproduction, and cued recall for Visual Paired Associates.

 Insert Figure 3 about here

The third subtest, Figure Memory, was designed as a test of recognition memory. The test taker is shown a set of abstract designs, each a square with different portions filled in with white, black, or various shades of grey (see Figure 4). After viewing a subset of (3) designs, the participant is asked to identify them within a larger set of (9).

 Insert Figure 4 about here

Figure 2. Wechsler Memory Scale-Revised, Visual Reproduction subtest

Figure 3. Sample of Wechsler Memory Scale-Revised, Visual Paired Associates

Figure 4. Sample from Wechsler Memory Scale Figural Memory subtest.

The Visual Memory Index score, therefore, is a combination of memory processes along with various other cognitive abilities: constructional ability, motor coordination, visuospatial memory, and perceptual ability, as well as free and cued recall, and recognition. In addition, the authors expected that "verbal mediation" is apt to be used to memorize and respond to the figure-colour pairs. This also is likely to occur during the visual reproduction task, which uses easily-recognizable geometric shapes such as triangles, squares, and circles. It is not surprising that despite the addition of several more visual tests, principle component analyses have found the WMS-R to consist of a factor structure similar to that of the WMS, with all immediate recall visual memory tests loading on the same general memory factor as the immediate recall verbal tests.

Another change made in the WMS-R was the inclusion of delayed memory tests, and a separate Delayed Memory Index comprised of weighted scores of the delayed measures for Visual Reproduction, Visual Paired Associates, and the verbal memory subtests. Interestingly, in the normative sample, all delayed recall tests loaded on the same factors as their immediate memory counterparts rather than on a separate delayed factor (Bornstein & Chelune, 1988). It may be that verbal factors in these tests are so intertwined with visual factors that, combined, they outweigh the effects of an immediate memory task versus a delayed memory task. Heilbronner (1992) pointed out that when different trials of the same subtest are included in factor analysis, there will be a strong tendency for them to be included on the same factor. Likewise, Loring (1989) suggested that when a factor analysis includes immediate and delayed tests of the same information, clusters are encouraged to form around the set of identical tests. Bornstein and Chelune (1989)

proposed that, when given sufficient time, older individuals rely more on verbal support for visual encoding, and thus the successful delayed memory scores will be associated with individuals who are better able to encode visual information verbally. Heilbrunner (1992) indicated that perhaps verbal and visual memory storage (measured by long-term retrieval) are extremely similar, if not the same, but that the process of encoding and consolidating this material-specific information (measured by learning processes) is important to investigate.

Finally, reliability measures for some subtests are relatively low. The average test-retest reliability for the Visual Paired Associates immediate and delayed subtests was 0.58 to 0.59; for the two administrations of the Visual Reproduction subtest, α levels ranged from 0.46 to 0.49; and the Figural Memory subtest yielded an α level of 0.44. Initial validity studies included only one study comparing WMS-R results to an independent memory measure, the California Verbal Learning Test (Delis, Cullum, Butters, Cairns, and Prifitera, 1988). Most of the validity studies depended on correlations with WMS scores. The WMS, however, only has an unweighted unitary memory scale, and is inappropriate for validating individual index measures (except, perhaps, the General Memory Index).

In conclusion, it appears that although the surface structure of the WMS-R has been unquestionably improved, the test's deeper structure remains essentially unchanged. The WMS-R is still mainly a test of verbal learning, but many of the measures may be sensitive to generalized, nonlateralized memory dysfunction. Only further validation studies comparing these subtests and Indexes with other memory measures known to be well-founded will indicate whether the Verbal and Visual Index scores have any significant

clinical usefulness. A complete critical review of the WMS-R may be found in Loring (1989).

Benton Visual Retention Test. The BVRT is widely used in clinical settings. It is a constructional task involving 10 cards, each with a set of three simple designs and geometric shapes that the patient is to draw immediately after a 5-10 s exposure (see Figure 5). Three "roughly equivalent" forms of this test are available (Benton, 1974; Riddell, 1962), with norms and exposure times varying among them. Norms include age groups (16 to 64) and IQ levels. One of the forms includes a short delay period (15 s) before the drawing commences, but normative information is not available for this form. Lezak (1983) indicates that the three-figure format allows some measure of immediate span and is particularly good for determining memory defects, since the individual must hold two of the figures in memory while drawing the third. A 'copy' form, which does not have normative information, is available. In addition, an unnormed multiple choice version is in use at various clinical settings.

- - - - -
 Insert Figure 5 about here
 - - - - -

As age rises, the expected number correct on the BVRT decreases and the expected error score increases. Unfortunately, normative information is not available beyond the age of 64, and when Klonoff and Kennedy (1965, 1966) tried to extrapolate correction scores for the original set of the BVRT tests to a sample of 80-92 year olds,

Figure 5. Sample from Benton Visual Retention Test.

estimated scores were much too high, indicating (incorrectly) what should have been “very defective” IQ levels.

While this test provides a wealth of information about various types of perceptual and drawing defects, it is insensitive to differential damage in the left versus right hemisphere, probably because its designs and shapes are easily encoded verbally (Zubrick & Smith, in Lezak, 1983). In addition, it has been found to correlate more highly with copying tasks than with memory (Silverstein, 1962). Thus, it appears that the perceptual and constructional components of this test far outweigh memory ability or dysfunction.

Complex Figures. Another widely used set of neuropsychological tests for assessing visual memory is the Rey-Osterrieth and the Taylor Complex Figures (see Figure 6). These tests were primarily designed to measure constructional praxis, but are also used to assess visual memory functioning. After the patient has copied the figure directly onto a sheet of paper, the original and copy drawings are removed, and the figure is drawn from memory. This free recall procedure is again conducted from 15 min to 1 hr after the original recall trial. These immediate and delayed memory tests measure incidental memory since the person is not informed that memory will be tested prior to its administration. Within the limits of an hour, Ebert found the exact time of the delay period does not appear to make a significant difference in scoring the figure for memory purposes (cited in Wood, Ebert & Kinsbourne, 1982). These findings were supported by Berry & Carpenter (1992) who used four different delay periods ranging from 15 to 60 min.

 Insert Figure 6 about here

Figure 6. Complex figure drawing tasks

In addition to quantitative scores, qualitative information from the recall trials provides sensitive information for determining location of neural damage from lesions (Milner, 1975; Osterreith, 1944; Taylor, 1969) as well as of traumatic brain injured patients (e.g., Brooks, 1974). As could be expected, performance on this test of memory is highly correlated with constructional ability, visuospatial ability, and visual perception. While Ebert (in Wood et al., 1982) reported a small one to two point difference between scores on immediate and delay recall trials, and Lee et al. (1989) reported that a delay trial did not significantly discriminate between individuals' scores with left and right temporal lobe lesions, factor studies have shown that a delay trial eliminates some of the verbal and visuospatial confounds of this test (Bigler, Rosa, Schultz, Hall & Harris, 1989; Brouwers, Cox, Martin, Chase & Fedio, 1984; De Renzi & Spinnler, 1966; Ellis & Daniel, 1971; Larrabee et al., 1985; Trahan & Larrabee, 1984; Trahan et al., 1990).

Several methodological problems are involved when using the standard Complex Figure test to assess for memory. First, because it is a constructional task, it has poor face validity for practical memory ability, and it is susceptible to impaired motor function as well as other localized damage unrelated to memory. Second, a potentially important confound is introduced by allowing the patient unrestricted time to copy the figure before remembering it. The length of time the figure is exposed varies individually, depending on visual processing speed, motor speed, and personality characteristics. As a result, the amount of encoding that can take place varies, even when memory is incidental. Third, the Rey-Osterreith and Taylor figures have been used interchangeably for follow-up testing, but while copy scores are equivalent, significant differences have been found across a

variety of retention conditions, various scoring schemes, and both incidental and intentional learning paradigms (Tombaugh & Hubley, 1991). And finally, inter-rater reliability measures are not readily available for these two tests, but Taylor's (1969) scoring scheme (presented in Lezak) leaves considerable latitude for subjective decisions, leading to what is suspected to be poor interrater reliability. Recent articles have presented normative data using more structured scoring schemes (Duley, Wilkins, Hamby, Hopkins, Burwell & Barry, 1993; Schmidt & Tombaugh, 1992). In addition, Tombaugh & Hubley (1991) have further standardized the administration of the Taylor figure when using it for memory assessment by developing norms for the administration of four learning trials and a delay memory trial followed by a copy trial for evaluating visuospatial, visuoperceptual, motor, and construction difficulties.

Biber Figure Learning Test (BFLT). The BFLT (Glosser et al., 1989) is unique as a non-verbal visuospatial analogue to auditory-verbal recognition procedures. It was designed to examine different aspects of visuospatial memory that have neuropsychological significance, such as the analysis of the rate and pattern of new visuospatial acquisition, examination of short-term and long-term visual memory functions, measurement of the differences between learning and retention, discrimination between visuospatial memory impairments and primary visuoperceptual and visuospatial deficits, and a comparison of voluntary retrieval with passive recognition of recently learned information.

The BFLT contains ten designs, each a combination of two simple geometric forms (see Figure 7). Specific rules were followed in creating each distractor design in

order to explore possible patterns of recognition errors (e.g., visuospatial orientation, shape discrimination). Besides immediate and delayed free recall, immediate and delayed recognition are assessed.

 Insert Figure 7 about here

This test is administered over a series of three conditions: (1) Five consecutive free recall learning trials, each consisting of the presentation of the items exposed at the rate of 3 s each. Immediately after the 10th design is removed, the subjects draw the designs in any order they wish; 2) immediate recognition trial after the 5th free recall trial, with each test design presented along with three similar distractor items in a forced-choice recognition procedure; 3) delayed free recall.

The only study found in the literature reporting the use of the BFLT was its original (1989) introduction, with 92 subjects, each from one of four groups: left hemisphere CVA, right hemisphere CVA, amnesia (from various causes), and normal controls. Performance within various groups followed predicted patterns (i.e., normal acquisition but reduced recall for patients with left and right cortical lesions; normal short-term visual memory and intact primary visuo-perceptual/visuo-constructional abilities along with severely impaired acquisition retention found in amnesics; and minimal age effects within the healthy control group). Scores in the immediate reproduction and copy trial were found to correlate with scores on the WAIS-R Picture Completion and Block Design

Figure 7. Stimulus items from Biber Figure Learning Task.

subtests. In addition, the principle of memory (free recall, delayed recall, and recognition scores) were found to be unrelated to visuo-perceptual and visuo-spatial impairments.

A significant test-retest correlation for sessions administered approximately 18 months apart was found for scores on learning, immediate and delayed recall, recognition, immediate reproduction, and copy scores, with no significant differences on any of the BFLT measures.

The recognition trial was incorporated to discriminate impairments in retrieval from deficits due to encoding and consolidation. Although both right hemisphere and left hemisphere patient groups had reduced recall scores, left hemisphere subjects had recognition scores equivalent to normal control subjects, whereas right hemisphere subjects were impaired compared to both of these groups. The copy trial is then used to separate visuo-constructional or visuo-spatial deficits from retrieval deficits. These findings appear to support the concept that left hemisphere lesions are related to constructional "apraxia" (Benton, 1985; Warrington, James & Kinsbourne, 1966), while right hemisphere lesions contribute in some way to a disturbance of visual perceptual and visual scanning that impacts on both free recall and recognition (Diller & Weinberg, 1977; Lezak, 1973). It appears, therefore, that the BFLT can separate the effects of primary visuo-spatial and constructional ability often found in other free recall visual memory tests (Loring & Papanicolaou, 1987).

The authors of this test reported that no significant correlation was found between BFLT measures and 'naming abilities' (the Boston Naming Test), even within a group of aphasic patients. They interpret this as evidence that the BFLT successfully excluded

verbal encoding from contributing to visual memory performance on this test. In addition, effects of educational grade level were not found, and only minimal effects of normal aging were present (the latter found only in free-recall scores). Age and educational effects, however, may emerge with a larger sample size, or when a longer list of test stimuli are used (e.g., Drachman, Glosser, Fleming, & Longnecker, 1982).

Normative information was not published for any of the scores, but this test appears to be promising as a comprehensive visual learning and memory test. However, it is not appropriate for individuals with motor impairments affecting the writing hand. If the recognition procedure was normed as a separate recognition memory test, it is suspected it might not be useful, as the task would not provide a challenge to any but those with the most serious memory impairments. It has a small range (10 items only) and, despite the findings that scores on the test are not correlated with word finding ability, it is likely its geometric shapes (circles, squares, triangles, etc.) could be easily identified with verbal labels.

Verbal Recognition Memory Procedures

Except for the Recognition Memory Test (described in the following section), very few tests exist of verbal recognition memory ability. However, many auditory-verbal memory tests include recognition procedures after free recall has been measured. The WMS-R, LAMB, and California Verbal Learning Test (Delis et al., 1987) are prime examples of this type of test. Research on the California Verbal Learning Test (CVLT) has examined performance and reported specific results on the recognition procedures for a

variety of patient groups. Therefore, the CVLT will be presented as an example of this type of recognition test.

California Verbal Learning Test (CVLT). The CVLT was designed to increase discrimination among various causes of memory dysfunction. For this purpose, it includes methods for examining multiple strategies and processes involved in learning and remembering verbal material. It consists of a list of 16 words (4 words from each of four semantic categories) presented orally over five learning trials. After the fifth trial, an interference list is presented for one learning trial. Immediately after the interference trial, immediate and category-cued recall of the first list is administered. Finally, after a 20-min delay, free recall, cued recall and recognition, again of the first 16 words, is assessed.

The recognition procedure consists of a yes/no response to each of 44 words sequentially presented in an auditory format. This list includes the 16 words from the target list along with 28 distractor words. Each distractor also names an object that could be on a shopping list. Each answer is marked as a hit, a miss, a false positive, or a correct rejection.

Normative information is presented for a variety of patient groups, including Alzheimer's disease, Parkinson's disease, Huntington's disease, and Korsakoff's syndrome (Delis et al., 1987). Because the distribution of age and sex was uneven in the reference groups, smoothed age curves were used to estimate the mean score and standard deviations within each age for each gender. Norms have been published for seven age groups, from 18 through 85 yrs.

In the studies described above, split-half internal consistency for the normative sample ranged from 0.74 for total test reliability to 0.92 for total trial scores. Internal consistency for a mixed clinical and non-clinical sample (including brain damage, normal adults, and detoxified chronic alcoholics) was 0.85 (Delis et al., 1991). A one-year test-retest measure for the original sample revealed significant correlations for immediate recall across all five trials, short- and long-delayed recall, and recognition hits and false positives. The score for recognition memory increased 1 by word, on the average, for the one-year test-retest subjects.

Studies of factor structure yielded 6 factors for the 19 CVLT indices (Delis, Freeland, Kramer & Kaplan, 1988). Recognition "hits" were included in the general memory factor while false positive scores on the recognition trial formed a second factor along with free recall and cued recall intrusions. Correlations with the WMS memory quotient were reported to be 0.66 for CVLT total immediate recall scores, and 0.56 and -0.23, respectively, for recognition hits and false positives for a varied group of clinical patients (Delis et al., 1987).

Performance across specific patient groups (chronic alcoholism, Parkinson's disease, multiple sclerosis, Huntington's disease, and Alzheimer's disease) reflected previous findings and theoretical differences for these patient groups. For example, chronic alcoholics had difficulty with this episodic memory task, even on recognition testing. Huntington's patients did better on the recognition task relative to free recall in earlier stages of the disease than those in later stages, and Alzheimer's patients'

recognition performance was worse than long-delay free recall performance, because of a high “yes” response bias and endorsing of a large number of false positives.

Subsequent research using the CVLT found identical patterns of performance for Alzheimer’s disease and Korsakoff’s syndrome patients. Huntington’s patients reflected a better overall performance, including better performance on recall measures. They also made more perseverative responses than those in the other two groups (Delis et al., 1991).

Researchers examining this test in various clinical settings conclude that the CVLT is useful in characterizing memory deficits associated with a variety of neurological and psychiatric disorders, including head injury (Crosson, Novack, Trenerry & Craig, 1988), temporal lobe epilepsy (Herman, Wyler, Steenman & Richey, 1988), human immunodeficiency virus (Saykim, Jansseu, Sprehn, Kaplan, Spira & Weller, 1991), Parkinson’s and Huntington’s diseases (Massman, Delis, Butters, Levin & Salmon, 1990), and Alzheimer’s disease (Kramer, Levin, Brandt & Delis, 1989).

Although this test assesses verbal recognition memory, this format is not a pure measure of recognition memory. The material in this test is limited to auditory presentation and requires verbal expression of memory. In addition, as it was pointed out earlier, there are strong indications that recall and recognition are separate concepts. Moreover, when an individual knows that a recognition format is going to be used, he/she will probably encode the differently than when information must be available for free recall. Finally, no localization studies have been reported for this auditory-verbal memory test to validate neurological deficits.

Visual Recognition Memory Tests

Visual recognition tests more commonly assess long-term as well as immediate visual memory. Included in this set of tests are the the Visual Paired Associates subtests from the WMS-R (reviewed with the other subtests from the WMS-R Visual Memory Index), the Visual Recognition subtest of the Randt Memory Test (Randt & Brown, 1983), the Recognition Memory Test (Warrington, 1984), and a subset of visual recognition tests involving repeated exposure to stimulus items. Representing this type of test are reviews of the Recurring Figures Test (Kimura, 1984) and the Continuous Visual Recognition Test (Trahan & Larrabee, 1988).

Randt Memory Test - Picture Reproduction. The Randt Memory Test, or the New York Test (Randt & Brown, 1983; Randt, Brown & Osborne, 1980) was created with the purpose of identifying memory loss and measuring deficits in storage and retrieval due to diffuse neural dysfunction, such as that existing in mild stages of dementia. This battery contains seven subtests, one of which is a Picture Recognition task. In this subtest, seven pictures (line drawings of common objects) are shown, with a 2 s exposure allowed for each. Immediately afterwards, the subject is given 3 s to indicate by a yes/no answer whether each of fifteen pictures was one of the original seven drawings. Then he or she is asked to name the seven objects first shown to assess free recall of the pictures. After 3 min of intervening tasks, the recognition memory trial is repeated. It is administered yet again the following day. This yields measures of immediate recall, short delayed recall, and long-term delay recall between sessions. In addition, after the delay recognition portion, subjects undertake a verbal free recall of the original items.

The original normative study included 350 control subjects, but no clinical sample. The Memory Index consists of the sum of scaled scores from the seven subtests. Standard scores are available for each of seven decades, but no subtest norms are included. A later study by Osborne et al. (1982) analyzed scores from an additional 300 subjects divided into three groups (community dwelling older, younger, and a clinical group of patients with at least a one-year history of progressive memory loss without known physical or neurological involvement). On the whole, scores on the 24-hr delay tests tended to discriminate older from younger subjects, while short-term memory tasks were more useful in differentiating community-based older individuals from clinical patients. Canonical correlation correctly predicted group membership for 91% of the younger and 84% of the older individuals.

This test has the advantage of collecting a wide range of information in a short time period. Again, however, more validity information on the Picture Recognition subtest needs to be gleaned from research before it is likely to be used by the clinical community as a test of visual memory.

Rivermead Behavioural Memory Test (RBMT). The RBMT (Wilson et al., 1985) was developed to complement existing memory assessment procedures by measuring and monitoring everyday verbal and visual memory in patients with acquired brain damage (Cockburn, Wilson, Baddeley & Hiorns, 1990; Wilson, Cockburn, Baddeley & Hiorns, 1989). It is a brief test designed to bridge assessments obtained by questionnaires and observation with laboratory-based measurements of memory. Therefore, while its administration and scoring are standardized, it makes use of ordinary memory situations

which are often troublesome for some patients with brain injury. It was intended that performance for most non-memory impaired people would reach ceiling levels.

There are twelve components to the RBMT, four of which are specific visual tasks. These include: (a) Free and cued recall for paired-associate memory for an unknown face and name; (b) free and cued recall for the location of a hidden object when an additional cue is given to the subject ("We have now finished this test"); (c) picture recognition (yes/no format) for twenty pictures of common objects (the original 10 pictures and 10 distractors); and (d) yes/no recognition of 10 photographs of faces (5 original and 5 distractors). The latter two tests involve testing recall a few moments after the original visual stimuli have been presented. No long-term delay measures are taken for any of these visual tests.

Normative information is provided for subjects between the ages of 16 and 69 years of age (Wilson et al., 1989). This study reported high reliability (inter-rater = 100%; composite scores for the four parallel forms ranged from 0.83 to 0.88). Validity measures were also reported to be satisfactory. The Profile Score correlated significantly with both subtests from the Recognition Memory Test (described in a later section). In addition, it correlated significantly with the Paired Associate Learning subtest from the Randt Memory Test. Performance on the RBMT was found to have a low correlation with intelligence for control subjects, with higher correlations for patient groups. Performance on visual memory subtests, however, was not related to the site of right vs. left lesions from stroke. Therefore, although this test is able to validly detect general memory problems, it cannot identify specific memory deficits. A further criticism is that perceptual

deficits interfere with performance on this test (Cockburn et al., 1990). These researchers developed a subscale to reduce this difficulty. Using this subscale, Goldstein, McCue and Rogers (1992) found that performance on the RBMT was related to functional capacity in 60-87 year olds who were cognitively intact or suffering from major depression. However, no significant correlation existed between scores on the RBMT and functional ability for those with dementia. In some cases, the correlation was even in the wrong direction (i.e., high RBMT performance but low functional performance).

Recognition Memory Test (RMT). Developed as an attempt to localize areas of neural damage leading to memory deficits, the RMT (Warrington, 1984) includes two different types of stimuli, both presented visually. An early test of facial recognition (Milner, 1968) pointed to the usefulness of this format as a non-verbal memory test of localized damage. Poor performance on this type of task may represent deficits related to memory for complex visual stimuli, or they may represent the functional loss of a very specific location in the brain which has been identified as being essential to the processing and memory of faces. The RMT includes two subtests: Recognition Memory for Words (RMW) and Recognition Memory for Faces (RMF; see Figure 8 for examples of stimulus items). In order to reduce the effects of readily available semantic information, the words are common 4-6 letter words. The faces are frontal views of unknown males.

 Insert Figure 8 about here

For each subtest, 50 unique stimulus items are exposed for 3 s each during which the patient is asked to state whether the word (or face) is 'pleasant' or 'unpleasant.'

Figure 8. Sample stimulus and response sheet for facial stimuli from Recognition Memory Test.

Immediately following this initial presentation, 50 cards are shown, each containing the 'old' stimulus item along with a 'new' word (or face). Using a forced-choice format, the patient is allowed 3 s to point to the word (or face) previously shown.

The original standardization sample, consisting of 310 inpatients aged 18-70, screened out those with cerebral disorders. Both raw scores and discrepancy scores can be compared with this age-related normative information.

Two validity studies were reported in the RMT manual. In the first study, scores on the RMW and the RMF were lower for patients with lateralized cerebral lesions who were proven to be unilaterally right-handed than were scores of the control group subjects. Scores on the RMF correctly predicted damage in the right hemisphere, while lower scores on either of the subtests predicted the presence of lesions in the left hemisphere. The second study involved patients with diffuse brain disease, divided into "mild-atrophy" and "moderate-atrophy" groups. Sixty-one percent of the patients in the atrophy groups were identified as having a deficit score on at least one of the memory tests. The mild-atrophy group was significantly worse than the normal subjects on both subtests, but score differences between this group and the group of patients with more extensive atrophy were only detected by the RMF, not the RMW.

In addition, 36 patients with traumatic brain injury showed significant memory deficits on both the RMW and RMF subtests, again with the latter subtest the most clinically useful (Millis & Dijkers, 1993). Although scores on the RMW and RMF were not significantly different for left- and right- hemisphere lesion patients, this may be due to the small number of individuals included in these groups. Of the five subjects with left-

hemisphere lesions, four had defective performances on the RMW subtest, and all of the subjects with identified right-hemisphere lesions had impaired performances on the RMF subtest.

Unfortunately, reliability data could not be found for this test. In addition, no allowance is made in the norms for vocabulary level, even though initial studies indicated that correlations with vocabulary levels are higher than correlations with age on both subtests (Kapur, 1987). In addition, the RMF may be confounded with specific deficits due to prosopagnosia rather than due to general visual recognition memory. Using the RMT, McNeil and Warrington (1993) found severe facial recognition deficits in one patient, even though he could visually recognize most of the individual sheep in a flock that he acquired post-stroke.

Overall, this test appears to be promising in helping to identify specific hemispheric deficits as well as identifying a general memory dysfunction. In Loring's review of the WMS-R (1989), he suggests that the revised WMS should have made use of tests already proven to relate to memory deficits associated with right cerebral dysfunction, such as Warrington's RMT. However, more research needs to be published to prove the validity and reliability of the Recognition Memory Test before it will be commonly used as a clinical test.

Recurring Figures Test. This test of visual recognition memory (Kimura, 1984) was designed to assess right hemisphere damage. Twenty stimulus cards with irregular or geometric nonsense figures are presented one by one for 3 s each (see Figure 9). Immediately afterwards, 140 cards are shown successively. This second set of cards

contains 7 sets of 8 of the original designs along with 84 new designs, for a possible hit rate of 56. The subject has 3 s to tell the tester whether the picture was or was not one of the original designs. To correct for guessing, the false positive responses are subtracted from a correct answer score.

 Insert Figure 9 about here

Internal reliability for this test was reported to be 0.94 (Rixecker and Hartje, 1980). Age norms were not available when the test was first introduced, but a subsequent study providing norms for three age groups (20-29, 30-39, and 40-54) found that scores decreased by 25-30% with age (Nielsen, Knudsen and Daugbjerg, 1989). This normative study represents the age group most often affected by traumatic head injury such as motor vehicle and work-related accidents. However, it is not appropriate normative information for the large number of patients experiencing cerebrovascular illness (such as stroke) or general cognitive decline (various dementias), which tend to occur beyond the age of sixty.

In validity studies, it was found that stimulus material was learned by both traumatic head injury patients and by matched controls (Brooks, 1974). But in another study, although overall scores did not differ significantly when left hemisphere (as opposed to right hemisphere) damage occurred, right hemisphere patients had nearly two times as many false positive responses (Brooker and George, 1984). In addition, geometric shapes were identified correctly more often than nonsense figures, and for those with left

Figure 9. Samples from Recurring Figures Test.

hemisphere damage, more false positive errors were earned on nonsense figures than on geometric shapes (Newcombe, 1969).

Continuous Recognition Memory Test /Continuous Visual Recognition Test. The Continuous Recognition Memory Test (CRMT) was introduced in 1979 by Hannay et al., and was revised as the Continuous Visual Recognition Test (CVRT) in 1988 by Trahan et al. Like the Recurring Figures test, these recognition memory tests evaluate ongoing learning as well as memory for visually presented stimuli. They were designed to effect differential diagnosis of right hemisphere damage involving visual memory systems. (Samples of items from these tests may be seen in Figure 10.)

- - - - -
 Insert Figure 10 about here
 - - - - -

The CRMT uses line drawings of common objects, while the CVRT uses complex, ambiguous geometric shapes and nonsense figures. This change in stimuli was prompted by research suggesting that tests designed to identify visual memory deficits should present stimuli with low verbalizability and should be administered with a minimal exposure time for the stimuli, both for encoding and for retrieval (Boller and De Renzi, 1967; Delaney et al., 1980; Trahan et al., 1985; Trahan et al., 1986). The interpretation is that the more complex the visual material, the more discriminating it will be in evaluating right hemisphere losses (Lee et al., 1989; Trahan et al., 1988; Trahan et al., 1990).

Both the CRMT and the CVRT follow a procedure similar to that found in Kimura's Recurring Figures test. After presenting 5 practice items, 20 stimulus items are

Figure 10. Sample stimuli from Continuous Visual Recognition Test.

presented for 2 s each. This administration is immediately followed by the presentation of 120 items, each on a separate card. Eight of the original 20 items reappears once in each of 6 blocks of 20 pictures. The other 12 items are unique drawings, some from the original category of shapes, but others from new categories. There is a 30-minute delayed recognition trial, where the individual is asked to identify the recurring design on a sheet containing that design along with six other distractor items from within the same stimulus class. After the delayed trial, individuals are asked to match-to-sample to examine possible visual processing problems. Signal detection analysis is used to determine the final score, which examines within-class errors as well as outside-class errors.

Results from both of these tests showed a strong age-related decrease in performance (Trahan et al., 1986; Trahan et al., 1990). A subsequent validity study of the CRMT, however, revealed that although it was able to detect generalized cerebral dysfunction, there was little or no evidence to support the prediction of lateralized brain lesions (Trahan et al., 1988). Therefore, the rest of this section will concentrate on findings of 'laboratory' research relating to the CVRT, as opposed to its clinical use.

In a normative study (Trahan et al., 1990), four age groups were used for the CVRT normative sample of 310 healthy, non-hospitalized individuals ranging from 18 to 88 (18-29; 30-49; 50-69; and 70+). Mean education levels were from 13 to 15 years for the various groups. In addition, 60 patients were administered this test, 30 with right-hemisphere and 30 with left-hemisphere cerebrovascular accidents. All subjects were screened for a history of other medical/neurological conditions known to affect brain functioning. Those stroke patients who had moderate to severe receptive language deficits

were excluded from the study. In addition, patients with visual field neglect or orientation deficits were excluded. There were no significant differences among age groups for correct answers, but there was a significant increase in the false alarm rate (i.e., identification of a newly occurring item as one that is “old”) for each successive age group, resulting in a significant difference across age groups for the d' (the Z score for the probability of hits minus the probability of false alarms) and Total Scores. Suggested cutoff scores were derived so that fewer than 5% of the subjects in the age sample fell within the “impaired” range. For the oldest age group (over 70 years), however, the cutoff score for the delayed trial fell at chance level, so that 34.5% of the subjects fell below the criterion. Therefore, the user is cautioned about interpretation of delayed trial scores for that age group.

Results from this validation study indicated that right CVA patients performed significantly worse than left CVA patients, and almost half of the patients exhibited the predicted double-dissociation. The authors report that 25% of the clinical sample did not reflect memory deficits on “both verbal and visual memory measures” included in the study, but these measures were not identified, so it is difficult to interpret whether the CVMT had any construct validity. In addition, 20% of the patients exhibited generalized memory deficits, the majority of these being in the right CVA group. It was suggested that there may be a generalized attentional impairment with a right-hemisphere focus (or perhaps caused by diffuse generalized damage to neural pathways) responsible for both visual and verbal memory deficits.

In conclusion, this test of visual recognition memory appears to have face validity, but criterion validity has not yet been shown. Future research is required to answer that question. In addition, it is not particularly useful for identifying lateralized damage in those over 70 years of age.

Conclusions about Visual Memory Tests

As can be seen by the above reviews, a range of recognition memory tests are available. Clinical visual recognition memory tests fall into one of two categories. In the first instance, they are presented as a recognition module in a free recall memory test, in which case they contain a constructional component. In this type of test, information to be remembered tends to be easily verbalized, whether that information is in the form of actual concrete nouns printed on a page, or a 'complex' figure. The second category consists of recognition procedures for visual constructional tests, used to sort out the contribution of various components of memory (e.g., visual scanning from visual construction), particularly if a delayed memory task is involved. As such, these tests are very useful in identifying generalized memory deficits, and some of them hold promise in identifying the hemispheric localization of deficits.

However, recall and recognition memory have been found to represent two separate components of the memory system. As such, it stands to reason that information presented as a free recall memory task will be encoded, and perhaps retrieved, in a way different from information that must be recognized. It isn't just that recognition memory is 'easier' than free recall memory. As pointed out earlier, the degree of difficulty between these two types of memory tasks often depends on the material or format used. In

addition, there are cases where recall memory for visual information cannot be assessed accurately (e.g., language deficits, motor problems).

Therefore, a second type of clinical recognition memory test has arisen in which the participant is informed that the information is to be recognized, as opposed to recalled. This second group of tests is also able to assess general memory deficits, but often contains complex stimuli presented at a rapid rate. Because of possible floor effects, it is questionable whether this sort of test would be useful in assessing seriously deficit memory, such as that found in later stages of Alzheimer's disease. For this reason, it is felt that a less complex test of visual recognition memory is apt to be useful in finding the range of intact memory when memory deficits are already known to exist. The most successful tests are the ones which involve learning as well as memory assessment, which have a delayed memory portion, and which contain complex abstract stimuli which strongly discourage the incorporation of verbal cues. The next section involves the rationale and procedure used for developing such a clinical test.

Development Of A New Set Of Visual Recognition Tests

It was clear from a review of the literature that recognition tests, in general, are important in examining retrieval processes (e.g., Delis et al., 1989; Storandt et al., 1986; Flicker et al., 1987; Haist et al., 1992; Moss et al., 1986; Mitchell, 1989; Tombaugh & Schmidt, 1992). There is a strong indication from these and other studies that recognition memory and recall memory are distinct processes, albeit highly related. When recognition memory procedures are included in a memory test along with spontaneous, or free recall

information, additional information about possible memory deficits is available in making clinical decisions. That is, free recall will assess what is available without cues, but recognition will allow a further comparison of what was encoded in long term memory beyond the material that retrieved with a free recall format. On the other hand, separate recognition memory tests add to a knowledge of an individual's memory abilities, particularly when spontaneous recall does not accurately reflect an ability to learn and remember. However, while recognition memory tests identify underlying memory deficits, they are limited to one memory process (i.e., recognition).

The intention of this thesis to develop a test of visual recognition memory which would supplement other clinical memory tests. When standard auditory-verbal memory testing is not possible because of expressive language deficits such as aphasia, this test would be useful for examining underlying memory deficits. When depression interferes with effortful processing, it would allow an assessment of memory ability because of the relatively automatic nature of the test. And when visuoconstructional or motor problems interfere with standard free recall of visual stimuli, it could act as a visual learning and memory test. Furthermore, its focus would be centered on moderate to severe cognitive deficits, where, because of near-floor effects, few available clinical memory tests are practical. Finally, it would be created with an older population in mind.

The development of a purely 'visual memory' task is not an easy chore. Several authors appear to be headed in the right direction for exploring hemispheric differences in memory through eliminating, as much as possible, the verbal and constructional components of visual memory tests by creating complex, abstract visual stimuli presented

in a recognition format (Warrington, 1984; Trahan & Larrabee, 1984). However, this format appears to have several drawbacks when used with an older population. For example, due to the additional effort required to encode complex abstract stimuli, attempting to test recognition memory by means of purely abstract information may reduce cooperation. How often are we required to identify abstract figures in our environment? Additionally, the practice of very limited exposure time for visual stimuli in existing clinical visual recognition tests is detrimental for an older population that, in general, requires a longer time to process incoming stimuli. Thus, when initial testing results in the discovery of possible visual memory or recognition memory deficits, appropriate tests are not available for use with an older population to further document the nature or degree of those deficits.

With these factors in mind, the development of a new visual recognition memory test, the Picture Recognition of Objects (PRO) Test, was begun several years ago. Evolving from this was the development of a second-generation set of tests, the Recognition Memory Battery. This process is outlined below.

In order to examine both learning and memory, it was intended that the PRO Test include two learning trials and one delayed memory trial. The goal was that in a normal population, the majority of pictures would be 'learned' and 'remembered' during the first trial, while the second trial would be used to consolidate that learning. There should be a significant difference in scores on the delay trial compared to scores on the final learning trial when memory is impaired. The first step in this study involved the development of appropriate materials, format, and test difficulty.

Pilot Study #1

The purpose of this pilot study was twofold. The first was to determine whether the use of unique distracter items for each trial would result in better memory scores than when the same distracter items were used for all trials. Confusion may result from repeating distracter items across trials ('point to the picture I showed you before' would not be very instructive after the first trial, since the individual already will have seen all of the pictures, some as stimuli and some as distractors). Therefore, it was expected that when unique distracter items for each trial were interspersed with the repeating target items, scores would be higher than when the same distracter items were administered for all trials. In addition, it was expected that for both conditions, scores would increase from the first to second trial, and that little or no forgetting would occur in a normative sample during a delay period. The second purpose was to obtain feedback about the test format and the test items as viewed by a sample of senior citizens.

Participants. Twelve community-dwelling seniors from the ages of 65-79 ($M = 73$, $SD = 5.04$), gleaned through personal contact, volunteered as participants.

Materials. Line drawings of 20 everyday objects representing a wide range of familiarity and of name agreement (Snodgrass & Vanderwart, 1980) were drawn with black marker, each on a 3 x 5 inch card. Three distracter items were drawn for each target item, each on a separate card. This resulted in a pool of 60 distracter items. Each distracter was chosen to represent the same semantic category as its corresponding target item. This format, similar to that used in the CAMCOG visual memory subtest, was selected to prevent encoding of the target stimuli by using the category name. Lee et al.

(1989) encouraged the inclusion of same-class items during the response trials in visual recognition tests. For example, for the target item 'broom', all distracter items were brooms of different shapes or details (see Figure 11).

 Insert Figure 11 about here

The Modified Mini-Mental State test (3MS; Teng & Chui, 1988) was administered to screen for cognitive impairment. This test was chosen because it yields both a 3MS score and a comparable MMSE score (Mini-Mental Status Exam; Folstein et al., 1975). The MMSE is still widely used to screen for cognitive deficits, but the 3MS provides a more comprehensive screening exam with higher validity and reliability (Teng, Chui & Gong, 1990; Tombaugh & McIntyre, 1992). In addition, the Vocabulary subtest from the Wechsler Adult Intelligence Scale, Revised (WAIS-R; Wechsler, 1981) was administered, using alternate items. The resulting score was extrapolated for this study as an estimate of intelligence. This Vocabulary subtest has been shown to correlate 0.84 with the Verbal Scale on the WAIS-R and 0.82 with the overall WAIS-R IQ score.

Procedures. All participants were tested individually. Two practice items were used to familiarize the participants with the test procedure of the PRO Test.

The 20 target figures were shown successively using a two second exposure for each figure. Immediately after viewing all targets, a second set of 40 pictures was shown, with the original target figures randomly interspersed with a set of 20 of the distracter

Figure 11. Sample of PRO Test response sheet for Pilot Study #1.

figures. Participants were asked to respond 'yes' or 'no' to indicate whether or not each item was one of the target items. During a second learning trial, all participants saw the 20 original target items in an order differing from this initial exposure. Then they again responded to the original targets and distracters, presented in a different order from the previous trial. Finally, after a 10 to 20 min. delay during which the participant completed the WAIS-R Vocabulary subtest, the 20 targets and 20 distracter pictures were used to test delayed recall, this time without initial exposure to the target pictures. There was no feedback about accuracy of responses during any of the trials.

The participants were randomly assigned to either the same-distracter or unique-distracter group. The five participants in the former group were shown identical distracter items during all three trials while the seven participants in the latter group saw a different set of 20 distracters for each of the three trials.

At the end of the study, participants were asked to provide information about item difficulty on the basis of similarity, discriminating the target from the distracters, attention to picture detail, and the methods of encoding that they used.

Results. The mean 3MS score was 96 (out of 100 possible points), with scores ranging from 89 to 100 ($SD = 2.87$). The mean vocabulary score was equivalent to a scale score of 13.25 ($SD = 4.12$), indicating that this sample was above average in verbal ability. Education levels ranged from some grade school through higher graduate degrees, with the 'average' education level being approximately a university undergraduate degree. There were no significant differences between the two PRO groups (described below) for any of the demographic variables.

An ANOVA was conducted on the PRO Test data which revealed no significant differences on PRO Test scores between the same-distracter and unique-distracter groups. No interaction was found among the type of distracter items and the trials factor. However, scores increased significantly for both groups from the first to the second trial ($F = 7.20 [2,11], p < .01$). A small reduction in scores from the second learning trial to the delay trial was not significant (32.67 out of 60 for Trial 2, 31.67 out of 60 for the Delay Trial).

Discussion. In this yes/no format, it did not seem to matter whether 'same' or 'different' distracter items were used. Although the results may be confounded by the relatively simple nature of the task, it is clear that the unique nature of the distracter items did not appear to help participants learn or retain target items.

As expected, scores increased on the PRO Test on the second trial, indicating that with more experience, items that had not previously been identified were remembered. In addition, little forgetting occurred when assessed on the delayed memory trial.

One of the goals in the development of this test was to allow a normal population to attain very high scores. An average of 20% of the answers were either false positives or false negatives for this group of high-functioning seniors. This leads to conjecture about the effectiveness of discriminability for this format when used with the general population.

Finally, information gained from subjective responses indicated that although the automatic response for participants was to encode the pictures verbally, this strategy did not help in identifying the correct answers on the response trial because of the similarity of

the response items to the target item. Items that were too difficult (answered incorrectly by all participants) and too easy (every participant got these items correct) were identified.

Pilot Study #2

The purpose of this study was to determine whether a reduced number of items (16 versus the original 20) would affect scores when using a yes/no format for the PRO Test. By reducing the number of items, the administration time would be decreased. It was desirable that the reduction would not affect the error rate.

Participants. Because of the difficulty in obtaining a large number of participants from an older age-range (over 65), and since the stimulus and distractor items had already been piloted with a group of older adults, it was decided that some of the pilot work on test format would be completed with individuals from a younger population. Ninety-three students at Carleton University volunteered to participate in this pilot study. The students were members of two different psychology courses at the second year level. Ages of the participants ranged from 20 to 45 years ($M = 22.9$, $SD = 5.1$).

Materials. Target and distracter pictures from Pilot Study #1 were used, but were developed into a slide format in order to facilitate group presentation. Since there were no significant differences found in the first study between the same-distracter and unique-distracter formats, this study used the same-distracter format for all trials.

Procedure. Participants were assigned to one of two groups. The 20-picture group was presented with the original 20 pictures followed by 40 items for the response trial (one distracter for each original item). The 16-picture group was shown 16 target items followed by 32 items for the response.

Similar to Pilot Study #1, two practice trials were conducted prior to the presentation of the test stimuli in order to familiarize the participants with the task. However, because the allotted time was limited and because a short delay period does not normally result in reduced scores for younger individuals (Craik, 1977; Hart & O'Shanick, 1993; Perlmutter, 1979), no delay trial was used. Each trial consisted of successive exposure to the complete set of target pictures with a different order of presentation for each trial. This was immediately followed by presentation of the set of target items interspersed with distracter pictures, again with the order varying for each of the two trials. Exposure time to the target items was increased from two to three seconds in order to accommodate the slide format. A five-second response time was allowed, during which participants marked their answer ('Yes' or 'No') on a score sheet.

Results. The total score for each of the two trials was converted to a 'percent recognized' score in order to allow direct comparison of the two forms used in this study. An ANOVA showed a significant increase in scores from Trial 1 to Trial 2 for both groups ($F = 138.47 [1,92], p < .001$) but no difference occurred between the scores for the 20-picture group compared to the 16-picture group.

Discussion. As expected, the number of target items did not significantly affect scores in either trial. However, the administration of a yes/no format was still felt to be less useful than a multiple choice format for the comparable number of distractor pictures; the latter would be shorter to administer (all distractor items on one page instead of having individual pages for each item). In addition, a multiple-choice format would eliminate the response bias often found in a single-item format. In determining subsequent development,

the test's intended use in a clinical setting was taken into consideration, as discussed in the next study. Format and length of administration time are important factors when choosing which tests to use in a clinical setting.

Pilot Study #3

As mentioned in the introduction to this thesis, many factors affect the difficulty of a recognition memory test, including altering the degree of similarity among items, varying the number of distracters, changing the number of items to be remembered, and choosing the type of response format. This study concerned the latter two of these factors. The purpose of the first part of this study was to determine whether a multiple choice format, as compared with a yes/no format, would affect the usefulness of the PRO test, while the second question related again to the number of items included.

Although only 40 items were exposed for each response trial in the pilot studies for this test, in an actual testing situation the target and all three distracter items would be shown on each learning trial for a yes/no response format. In addition, the larger testing booklets required would be bulkier, while clinicians are looking for short concise ways to get information with as little equipment as possible. It was thought that a multiple-choice version would provide similar data in less time and with fewer materials.

Most currently available clinical recognition memory tests use a single or a two-item response format. In Warrington's Recognition Memory Test, for example, each response item consists of a distracter picture shown along side a target picture. The participant knows that one of the pictures is correct and one is incorrect. Therefore, there is a 50% chance of guessing each item correctly. This multiple choice type of yes/no response

seems to be subjectively easier than tests such as the CVMT, Kimura's Recurring Figures test, and the Randt Visual Memory subtest, where response pictures are shown sequentially. For example, the Randt Visual Memory subtest has approximately the same 1:1 ratio of distracter pictures to target pictures. However, in using a sequential presentation, the test taker has to determine whether each and every picture has previously been seen. Although it allows for an exploration of response style through the examination of false positive and false negatives along with the recording of hits and misses, it still results in a 50/50 chance of getting each item correct. It also increases the number of items in the test. And when a successive response test increases the number of distracter pictures compared to target pictures (e.g., the Recurring Figures test and the CVMT), a 50/50, or chance level of getting each item correct or incorrect still exists.

For this study, a different strategy was employed. While maintaining a forced multiple choice format, the number of distracters was increased from one, as found in the RMT, to three, as found in the CVMT. For each item, the participant would always know that one of the item choices was correct, but with four pictures, the chance of getting a correct answer for each picture would be 25%.

The second purpose of this pilot study was to determine whether the number of distracters used in response trials of the PRO Test would affect scores on each trial. If the number of response choices was increased from four to six, the probability of getting a correct answer by chance on any one item would be reduced to approximately 17%. Therefore, two versions of the multiple choice format were prepared: the first version used

four response pictures (target and three distracters), while the second included six response pictures (target and five distracters).

Two predictions were made. First, it was expected that there would be an increase of scores from the first to the second trial on both formats of the multiple choice test. Second, it was predicted that scores on the six-item multiple choice test would be lower than scores on the four-item test. In addition, it was hoped that the percentage of correct responses on the multiple choice tests would be equivalent to percentages collected during the previous pilot study for the more lengthy yes/no format.

Participants. From two separate undergraduate psychology classes, 73 students volunteered to participate in this study. Ages of the participants ranged from 20 to 53 years ($M = 27$ years, $SD = 8.63$). Participants were randomly assigned to one of two groups: a four-picture multiple choice group and a six-picture multiple choice group.

Materials. Stimuli from the previous pilot study were used for the four-picture group, while for the six-picture group, an additional two distracter pictures were developed for each target item. Following the pattern of another experimental visual memory test (Biber Figure Learning Test [BFLT]: Glosser et al., 1989), each distracter picture adhered to a specific rule (orientation distracter, shape distracter, and unique distracter). However, the BFLT used only three distracters, so two additional rules were developed for the 6-picture multiple choice format. Each distracter picture on the PRO followed one of the following rules:

- a. *Orientation distracter:* Target item in a different orientation (from 30 to 180° rotation) retaining the same detail and perspective shape (this sometimes involved a mirror image)

- b. *Shape distracter*: Target item with different shape and/or detail, but always with the same orientation; if detail difference, involved an obvious difference
- c. *Unique distracter*: but in the same semantic category and with similar degree of detail.

(Additional rules for 6-picture test)

- d. Matched with either the target or (b) above, with a different orientation, detail, and/or shape
- e. Either matched with (c) above, or similar in shape and/or orientation to the target or (b) above

Thus, each response sheet contained four (or six) drawings positioned in such a way to make sure that no more than two items in a row contained the target item in the same position on the sheet (see Figure 12) Each multiple choice alternative was accompanied by a letter (a, b, c, d, e, or f). All materials were prepared on slides.

Insert Figure 12 about here

Procedure. The procedure from Pilot Study #2 was followed, with the exception that the participants marked the letter of their answer of choice on the score sheet, as opposed to writing 'Yes' or 'No.'

Results. An ANOVA was conducted and, as predicted, scores for both the 16-item and 20 item tests increased significantly from the first to the second trial ($F = 41.96$ [1, $p < .001$]). The second prediction was not supported; there were no significant differences in

Figure 12. Sample of PRO Test response sheet for Pilot Study #3.

scores between the six- and four-picture tests. Finally, for both the 16-item and 20-item tests, the percentages correct on the multiple choice format were significantly higher than those earned on the true/false format ($F = 4.8$ [1,72], $p < .05$ for the 20-picture test; $F = 7.47$ [1,72], $p < .01$ for the 16-picture test).

Discussion. Since the multiple choice format is considerably less time-consuming to administer and uses less materials, it was decided to continue with this format. As mentioned in the Introduction, it has been demonstrated repeatedly in the literature that older individuals perform less well than younger ones on many types of learning and memory tests. However, it remained to be determined whether the desired ceiling effects for the PRO Test would show up when testing intact older individuals.

Pilot Study #4

The purpose of this study was to determine how a sample of community-dwelling seniors would perform on a multiple choice format of the PRO Test. It was expected that relatively high scores would be earned, since the sample chosen was relatively high functioning.

Participants. Participants in this study included 72 senior citizens from Canada and the United States who were attending an Elderhostel program in Ottawa. They ranged from 55 to 84 years of age ($M = 68$, $SD = 6.0$). Although no measure of education or IQ was taken, it is common knowledge in the area of life-span development that older adults participating in educational activities such as the Elderhostel program are highly educated compared to a random community sample of adults in the equivalent age range.

Materials. The 16-picture, 6-choice slide format prepared for Pilot Study #3 was used for this study.

Procedure. As in the previous study, two learning trials were given, but no delayed memory portion was included.

Results. Four age groups were formed on the basis of reported age: 69 years and under, 70 - 74, 75 - 79, and 80 years and older. ANOVAs were performed revealing that scores increased significantly from Trial 1 to Trial 2 for all age groups ($F = 39.39$ [3,71], $p < .001$), but that there were no significant differences in scores among the four age groups.

Results from this study were compared with results from the multiple choice data collected from the younger group for the 16-picture test administered in Pilot Study #3, and no significant difference in scores was observed.

Discussion. This study indicated that, regardless of age, learning and memory did occur in this test of recognition memory. The degree of mastery, even on a recognition task, varies according to the materials used and the format chosen. Age differences that usually exist in memory testing did not appear in this study using a multiple choice format. Various factors may be responsible for these findings. First, they may be due to the relatively persistent nature of recognition memory when compared to free recall. Second, it may be due to the nature of the test items, that is, recognizable items that can be easily encoded verbally, despite using distracter pictures from the same semantic category. Third, the small number of items may have produced ceiling effects. A fourth possibility is that the high functioning of the group of seniors participating in Pilot Study #4 may be

responsible. It remained to be seen what performance in a normal older population would be.

Pilot Study #5

The purpose of this study was to collect normative information from a random sample of seniors (65 and older) on the multiple choice version of the PRO Test. The Pilot Study #4 version was used, with the exposure time increased and a delay trial included. It was predicted that scores would increase from Trial 1 to Trial 2, and that cognitively intact participants would have similar scores on Trial 2 and the Delay Trial. Furthermore, it was expected that participants who were identified as cognitively deficit would not show this increase from Trial 1 to Trial 2, and in addition would show reductions in Delay Trial scores.

Previous research on visual memory has found significant correlations between IQ levels and scores on these tests (e.g., Fagan, 1984). It is not certain whether this represents an interaction of IQ with visual memory or with visuospatial ability. It was suspected that the Elderhostel group tested in Pilot Study #4 would represent higher levels of cognitive functioning than this random sample of community-based seniors. Therefore, if visual memory is a significant factor in the correlation between IQ and visual memory tests, it was expected that for this sample, average scores on the PRO Test would be significantly lower than those earned by participants from the Elderhostel group.

Participants. A random sample of 335 community-dwelling seniors in the Ottawa-Cornwall area was accessed through an "add-on" study to the Canadian Health and

Welfare Study of Aging and Health. Participants were selected from the general population by their age, which ranged from 65 to 85 years ($M = 73.7$, $SD = 6.6$).

Materials. A test booklet was prepared using the 16 target items and 2 practice items from the slide format used in Pilot Study #4. For ease of administration, the same order of presentation was used for all three trials. A set of standard instructions for administering the test was prepared and a score sheet was designed to allow rapid scoring during the test (see Figure 13).

- - - - -
 Insert Figure 13 about here
 - - - - -

Procedure. The PRO Test was administered by trained testers following the interview portion of the formal Health and Welfare study. A questionnaire for a second “add-on” study was administered between the two learning trials and the delay trial of the PRO Test. The delay interval, recorded by the tester, varied from 4 to 22 minutes, depending on the time it took to complete that questionnaire.

Exposure time for the target items was increased to five seconds. On the test trial, if the participant did not answer within ten seconds, he or she was encouraged to respond. If no response was forthcoming, the item was marked wrong. Answers for the PRO Test were recorded by the tester, with no feedback given regarding the accuracy of the response.

Other information from this study that was used in analyzing the data from the PRO Test included the participants’ age, gender, education level, and scores on the 3MS. Since

Visual Recognition Memory Test . . . 79

PRO TEST SCORING SHEET

Name: Date of Testing: June 11/91

Age: 24 Gender: F M (Circle one) Time of Delay: 4-10 min.

SUBJECT

6	5	4
3	2	1

TESTER

TRIAL 1

SUBJECT

6	5	4
3	2	1

TESTER

TRIAL 2

SUBJECT

6	5	4
3	2	1

TESTER

DELAY

***** DO NOT SIGNIFY AGGRESSIVE BEHAVIOR *****

Item	Key	Answer	Score	Key	Answer	Score	Key	Answer	Score
1. Knife	(2)			(2)			(2)		
2. Cup	(1)			(1)			(1)		
3. Fruit	(5)			(5)			(5)		
4. Box	(4)			(4)			(4)		
5. Key	(6)			(6)			(6)		
6. Watch	(4)			(4)			(4)		
7. Bed	(2)			(2)			(2)		
8. Cap	(6)			(6)			(6)		
9. Umbrella	(3)			(3)			(3)		
10. Wallet	(1)			(1)			(1)		
11. Mailbox	(4)			(4)			(4)		
12. Lamp	(6)			(6)			(6)		
13. Pan	(5)			(5)			(5)		
14. Flower	(3)			(3)			(3)		
15. Boat	(6)			(6)			(6)		
16. Broca	(2)			(2)			(2)		
	Sum:			Sum:			Sum:		

Figure 13. Sample of PRO Test answer page.

the participants had the choice whether the interview was to be conducted in French or English, language preference was also recorded.

Results. There was no significant correlation between PRO Test scores and the durations of the delayed retention interval ($M = 9.5$ min, $SD = 4.5$), or between PRO scores and language preference, gender, or education levels. As a result, these four variables were not examined further for this pilot study.

For the purpose of statistical analyses, participants were divided into six age groups, each containing a five-year age range, with the oldest group containing individuals from 80-85. A MANOVA was conducted to examine differences in scores between the various age groups. There was a statistically significant, but small difference in scores between age groups ($F = 3.19$ [4,334], $p < .05$), with scores increasing significantly for all age groups from Trial 1 to Trial 2 ($F = 129.81$ [2,334], $p < .001$). This indicated that the pattern of learning and memory was equivalent for all age groups (see Figure 14).

 Insert Figure 14 about here

Participants were then separated into two groups on the basis of a 3MS cut-off score of 78/77: 'cognitively intact' ($n = 297$) and 'cognitively impaired' ($n = 38$). Thirty-one of the latter group had completed all three PRO Test trials, and were matched by age with cognitively intact participants (see Figure 15). An ANOVA was performed using trials and cognitive status, with age groups collapsed. Once again, there was a significant trials effect ($F = 26.87$ [2,334], $p < .001$). Individual t -tests indicated that this effect was due to

Figure 14. Mean trial scores for age groups in Pilot Study #5.

an increase in scores from Trial 1 to Trial 2, and that the small reduction in scores from Trial 2 to the Delay Trial was not significant for either group. The scores on each trial were significantly higher for the cognitively intact group compared to the cognitively impaired group ($F = 16.06 [1,334]$, $p < .001$). No significant interaction occurred among age and trial variables.

 Insert Figure 15 about here

Finally, 52 participants from the Elderhostel group in the previous study were matched by age with participants from this pilot study who had earned intact scores on the 3MS. The participants from Pilot Study #4 had a significantly higher score on the two learning trials than the Health and Welfare intact participants by approximately one and a half items ($F = 39.81 [1,51]$, $p < .001$). Since there was a significant increase in scores from Trial 1 to Trial 2 ($F = 84.39 [1,51]$, $p < .001$), and there was no interaction between trials and the cognitive ability of the group, it was concluded that the difference in scores was consistent for both trials (see Figure 16). Comparisons on delayed memory ability could not be evaluated in this analysis since no delayed memory trial was included in Pilot Study #4.

 Insert Figure 16 about here

Figure 15. Mean scores for Intact and Impaired groups of matched sample for Pilot Study #5.

* Matched for age

Figure 16. Mean PRO Test trial scores for young and for a matched sample of high functioning old and normal old, and a matched sample of normal old and impaired old.

Discussion. This test represented the predicted pattern of a typical learning function of a memory test in that scores were higher on the second learning trial than on the first learning trial. Additionally, the seniors in the Elderhostel group performed better on this test than did seniors in the general population, and this general population performed better than cognitively impaired seniors. Varying degrees of learning were found from Trial 1 to Trial 2, even among most of those who were cognitively impaired.

Although it was no surprise that delayed memory scores for the intact group followed the typical pattern of little or no forgetting, it was surprising that scores for the cognitively impaired group, which were expected to be comparatively lower on the Delay Trial than those of the intact group, repeated the same pattern. Based on previous findings in the literature, these findings would indicate that on the PRO Test, the delay time was too short, the test length too short, or items too easy. In spite of the fact that there was a statistically significant difference between PRO Test scores of the cognitively intact and cognitively impaired groups, the clinical utility of this test should be questioned due to a clear lack of overall variability in group scores. The mean increase in scores for all participants from Trial 1 to Trial 2 consisted of 1.7 items, but the standard deviation was approximately 2.35. At the same time, there was a large within group variability. Although participants with the lowest 3MS scores earned the lowest scores on learning (and delayed memory) trials on the PRO Test, there were individual participants with 3MS scores above the cut-off point of 78/77 with low PRO Test scores, and vice versa. Keeping these patterns in mind, it is difficult to decipher the meaning of an individual's test scores.

Results from this pilot study are difficult to interpret since there was very little in the way of validity measures available for perusal. Although the final clinical identification of “cognitively impaired” or “cognitively intact” was available, individual neuropsychological tests were limited in number, and were administered only to the final clinical sample. In addition, those conducted were not particularly applicable to this “add-on” study. For example, no measure of IQ was taken, and no other visual memory tests were administered. While the available results may be interpreted to mean that recognition memory is not meaningfully affected by mild to moderate levels of cognitive impairment, they may also represent the relatively easy nature of the test, or imply that the test was too difficult since those with the lowest 3MS scores seldom were administered the delay trial for a variety of reasons and thus were not included in the final analyses. Or it may be that cognitive impairment is not synonymous with visual recognition memory. It is most likely, however, that the test is not difficult enough to challenge even those with mild to moderate cognitive deficits. A controlled study would need to be undertaken with known diagnoses of memory deficits and other validity measures in order for the clinical usefulness of this test to be evaluated.

General Discussion

The aim of these pilot studies was to develop a test that would differentiate levels of memory ability, and to develop a test that ensured high performance among intact individuals of any age while allowing those with lowered memory ability to reflect learning ability and to earn a range of immediate and delayed memory scores.

It was clear from the studies that the multiple choice format of a short visual recognition test acted as a test of learning and memory ability for pictures of objects in an intact sample. Furthermore, this test was sensitive enough to reflect different degrees of intact memory ability in a sample of individuals with lowered cognitive ability. When an individual has serious memory problems, standard learning and memory tests that involve free recall of information may involve floor effects. With the PRO Test, however, these floor effects did not exist in a mild to moderately impaired group when visual recognition learning and memory were assessed. It is not certain what scores would be like for a severely impaired group, as the testers conducting the assessment did not administer this test to those individuals.

The literature reveals that pictures are on the easy end of an 'easy-hard' continuum, but the PRO Test was developed to examine retained memory ability in seniors when memory ability was known to be impaired. It was for that reason that a 'picture' format was chosen, rather than more complex visual information (abstract shapes and designs). But because pictures are relatively easy to encode verbally, every attempt was made within that format to reduce any obvious verbal encoding. This was done by using a larger number of items on the response trials (6), and by making use of same-category response items. Differences in the response and stimulus items concentrated on orientation and rotation, rather than within-picture details and shading. For example, on the picture of a watch, the time setting was kept identical for all exemplars; for the same reason, shading and minor detail differences were removed from the first generation of drawings. It is

likely that these precautions were not successful in decreasing the natural tendency to encode visual items verbally, however.

As the results of the community sample (Pilot Study #5) were examined, several serious problems emerged. The first problem was its lack of sensitivity and specificity of the PRO Test. Although the results from these pilot studies indicated a pattern comparable to that found in most memory tests, it became obvious that a 16-item test would limit its clinically utility because the range of responses was relatively small. Although the PRO Test detected differences in memory ability across a large group of seniors, performance could not be predicted for individuals.

A second obvious difficulty with the format of the PRO Test was its reliance on identical distracter items for each trial, all presented in the same order and placed in the same location on the page for each trial. In fact, the same target and distracter sheets were used for each of the three trials. The opportunity to become involved in the Health and Welfare study was quite unexpected. There was only time to develop the Pilot Study #4 version into a test booklet format, whereas designing a larger test with unique response items could not have been completed in time. It was felt to be more important to get some community-based information on how a random sample of seniors would perform on the available items than it was to pass up this chance. It was known that interpretation of the results could be affected by an individual's misunderstanding the task demands. That is, when asked to "point to a picture you have been shown", all of the target and distracter items were 'old' items after the first trial. However, it was decided that basic information about the usefulness of this 16-item test could be gathered while further development of

alternate response items was being undertaken. If the results supported further development, the change could then be made before undertaking the gathering of norms and examination of the test with a clinical sample.

In the meantime, confirmatory evidence was collected in a Boston hospital from neuropsychological testing being conducted with participants representing a wide range of neurological impairments including stroke, seizure disorders, Alzheimer's and other dementias, and head wounds. The results indicated that this short version of a recognition test was not predictive of memory disorders as they were being assessed with free recall tests. That is, while scores on spontaneous recall tests indicated fairly serious episodic memory deficits, recognition scores on the PRO Test were unreasonably high - higher than the literature generally shows for this type of format. Using an intentional recognition memory format provides a measurement of memory that is obviously much more robust than was expected.

Consequently, changes were made to the original PRO Test presentation. Renamed the Recognition Memory for Pictures Test (RM Pictures), this alternate format uses a much larger number of items. It, too, consists of two learning trials and a delayed memory trial, and is presented in a forced multiple choice format. However, similar to the CVRT, the distractor items for each trial are unique. That is, for each target item, there are 9 different distractors (3 distractors x 3 trials). This allows less opportunity for confusion in determining whether or not an item is 'old' or 'new' for each response trial.

At the same time, two companion tests were developed which accessed a variety of visual codes and reflected a continuum of difficulty. Pictures appear to be the least difficult

items to remember in visual memory tests: words seem to represent more of a challenge to recognition, and abstract designs provide the greatest challenge. Rather than combining these types of stimuli into one test, we developed three separate, but comparable tests in order to provide the most useful format. We learned that a much larger number of items than 16 must be used in order to overload what appears to be a very sturdy part of the memory system. But to incorporate large numbers of each type of item in a single test would result in a very long test, and would make the administration impractical in a clinical setting. Two additional tests were developed, the Recognition Memory for Words (RM Words), with 48 items and, providing more of a challenge, a 24-item test using complex, abstract designs, the Recognition Memory for Designs (RM Designs).

This set of three tests attempts to provide a visual, as opposed to auditory, method of assessing memory. The three tests make up the Recognition Memory Battery. This battery could be used as a whole to examine memory when spontaneous recall is impractical, such as in cases of chronic or acute depression, or when neural damage reduces a patient's ability to respond to free recall tests.

On the other hand, these tests may be used individually. Even though the format of the three tests is directly comparable, the different types of stimuli (pictures, words, and designs), emphasize different systems in visual memory (verbal and non-verbal).

Because picture stimuli are easily encoded in both the visual and spatial systems (i.e., roughly relating to left and right hemisphere functions), the RM Pictures test could be used to assess memory losses due to damage in both hemispheres. For example, when global deficits affecting non-localized damage to the memory system are suspected, as in

various dementias, or other disorders exist which make it difficult to assess memory ability with more traditionally auditory-verbal free recall tests, this test using picture stimuli may be useful.

Recognition memory for words has been shown to relate more to left- than to right-hemisphere functioning. The RM Words test could thus be used to document damage to the verbal memory system, particularly when expressive language function has been disrupted.

Damage confined to right-hemisphere memory functions has been more difficult to localize using neuropsychological tests. Following Trahan et al.'s (1990) suggestions for a test that emphasizes right hemisphere differences in the degree of involvement of verbal versus visual recognition memory, the RM Designs test administers a set of abstract designs that are difficult to verbalize, it uses recognition memory to avoid confounds such as visuospatial and constructional deficits, distracter items are similar to the original stimulus, limited exposure time is used (3 sec exposure time per item), a learning component is included, and a delayed recognition task is included. The present study makes use of these three tests to (a) gather normative information in a community-based sample, and (b) to explore the usefulness of these tests in a clinical setting.

METHODS

Participants

A total of 200 female and 170 male volunteers from the community (ages 16 through 85) were employed in this study. For recruitment purposes, the participants were recruited to fill 40 cells in a manner that should result in comparable numbers of males and females, as well as individuals with or without a high school diploma in each of ten age groups. (For a complete description of these cells, see Appendix A).

Materials

The Recognition Memory Battery consists of the Recognition Memory for Words (RM Words) test, the Recognition Memory for Pictures (RM Pictures) test, and the Recognition Memory for Abstract Designs (RM Designs) test. The RM Words and RM Pictures tests each consist of three separate test booklets. The first booklet for each of these tests contains two practice items. It then includes 48 stimulus items, each on a separate page. This is followed by 48 response trial sheets, each containing one of the target items and three distractor items. This comprises the first learning trial. The second booklet of each subtest contains 48 stimulus items and 48 response sheets (the second learning trial), but no practice items. The third booklet contains the 48-response-sheet retention trial, without stimulus sheets.

If the error response rate varies, it is important to identify whether systematic errors are inherent in a test in the form of order effects or effects due to the position of an item on the page. To reduce the possibility of order effects, the sequence of item presentation

was randomly varied for stimulus trials and response trials for each of the Recognition Memory Battery tests. As a result, it would be extremely difficult to learn items by sequences. As a further control, target items were counterbalanced for location on the page. In some cases, the target appeared as item B, sometimes as item A, and so on. As a result, no single location on a page should have been chosen more often when a response was incorrect than any other location on the page. Still, on each response page an individual had a 1 out of 4 chance (25%) to correctly identify the stimulus object by guessing. Because it is difficult to control how individuals perceive, understand, and encode individual items, no effort was made to control for item difficulty within each test, either of the target or distractor items. Therefore, it was expected that individual item error rates would vary within a test.

To avoid the problem of confusion resulting from repeated exposure of distractor items on multiple trials, response distractor items were not used more than once. In order to effect this, 48 target items and 432 individual distractor items had to be created. The RM Designs, consisting of 24 target items and 216 individual distractors, is half the length of the RM Pictures and the RM Words. Therefore, only two booklets are needed to contain this set of subtest materials: one for the practice items and two learning trials, and one for the retention trial. A sample of the three types of stimulus materials from the Recognition Memory Battery is shown in Figure 17.

 Insert Figure 17 about here

Stimulus page

Response page

Recognition Memory for Pictures

Stimulus page

Response page

Recognition Memory for Words

Stimulus page

Response page

Recognition Memory for Designs

Figure 17. Sample items from Recognition Memory Battery.

The MMSE was administered to screen out those with broad cognitive deficits. Because level of depression is also known to correlate with memory scores (Gainotti & Marra, 1994; Koenig, Cohen, Blazer & Krishnan, 1993), the Geriatric Depression Scale (GDS; Yesavage, Brink, Rose, Lum, Adey, & Leirer, 1983) was also administered. This test was designed for use with an older population, but when comparing the performance of young (20 to 39) and old (ages 70-102) individuals, no significant differences were found (Koenig et al., 1993). As a result, it was used for the entire sample in order to allow direct comparisons.

The following cognitive tests were also included. Because of the findings from Pilot Studies #4 and #5, and because of the high correlations found between memory and IQ (Fagan, 1984; Powell, 1979; Rosenberg & Heilbronner, 1984), the Vocabulary subtest of the WAIS-R was administered and scores used to estimate intellectual ability. The Trail Making Test (part of the Army Individual Test Battery, 1944) was used to assess visual conceptual and visual scanning ability. A measure of naming ability, the Boston Naming Test (Kaplan, Goodglass & Weintraub, 1978), was administered to assess each individual's ability to name common objects. Rosen's (1980) procedure was used to assess verbal fluency. This consists of administering the Controlled Oral Word Association Test (Benton & Hamsher, 1976; Benton, Hamsher, Varney & Spreen, 1983) using the letters F-A-S (Spreen & Straus), and the portion of the Set Test (Isaacs & Kennie, 1973) which allows one minute to name different animals as quickly as possible. Finally, additional measures of verbal auditory and visual memory were administered. These included

subtests from the G-LAMB (Hubley, 1995): a 15-Word List, a 31-item Paragraph, and a Simple Figure.

Procedure

Participants were sought throughout the Ottawa-Carleton community. Tests were administered individually. At the beginning of the testing session, each volunteer signed an informed consent form. This was followed by an interview to record age, gender, and years of education. In addition, a medical history was taken to record those who were currently taking medications, paying particular attention to those producing cholinergic or sedative effects. Information was also sought about past or present head injury, psychiatric conditions, and/or neurological or medical conditions that might lead to memory deficits. A total of 11 participants were rejected due to cognitive and/or medical complications.

The order of administration within the testing session(s) can be found in Table 1. In the case of some of the older participants, two separate testing sessions were undertaken because of fatigue or time limitations. As a result, there was some attrition, resulting in unequal numbers of participants for some of the measures.

- - - - -
 Insert Table 1 about here
 - - - - -

Administration of each Recognition Memory Battery test followed the same format (see Appendix B for a copy of administration directions). Two practice items were

Table 1

Order of presentation of tests

Informed consent
 Mini-Mental Status Examination
Picture Recognition Memory
 G-LAMB Word List
 G-LAMB Simple Figure
 Subjective Age Questionnaire
 Chronological Age Questionnaire
 Geriatric Depression Scale
 Trails A & B
Picture Recognition delay
 Word List delay & recognition
 Simple Figure delay & recognition

 G-LAMB Paragraph¹
RM Designs
RM Words
 Word Fluency
 Vocabulary subtest from WAIS-R
 Paragraph delay
RM Designs delay
RM Words delay
 Boston Naming Task
 Feedback and debriefing sheet

¹ Break was allowed, if required. If two testing sessions were conducted, this portion of the tests was given during the second session.

presented before the beginning of the first trial. Each stimulus on the learning trials was exposed for 3 sec until all items were presented. Then participants were requested to select the correct item (picture, word, or abstract drawing) on each response page by asking them to "Point to the item I showed you before.." Participants were permitted unlimited time to respond, but after 15 s, they were encouraged to respond by being instructed to "guess if you are not sure."

A copy of the administration and scoring directions the other memory tests used in this study can be found in Appendix C. Appendix D contains directions and scoring instructions as they were used for the Word Fluency task, Trails A & B, and the Boston Naming Test. Test materials were interspersed in such a way to assure a minimum 20 min delay period between the final learning trial of each memory task and its retention trial. Following testing, the participant received verbal and written feedback about the project.

Analyses

Throughout the results section, a minimum alpha level of 0.05 was used. All correlations reported were two-tailed Pearson Correlations. Because of its stringent criteria, *post hoc* comparisons were conducted using Bonferroni's correction, which divides the significance level ($p < .05$) among the number of comparisons made. When appropriate, Levene's Test of Homogeneity of Variance was used, again with a probability level of 0.05.

When developing the normative tables for each of the tests, cumulative percentiles rather than percentiles were used. Because these tests were designed to be used in a clinical setting, it is important that these scores can be compared to scores on other types of tests used to determine cognitive functioning. Since each cognitive test consists of a

different number of possible correct answers, normative scores are typically presented in the form of percentages rather than using raw scores. Two types of percentages are generally used: percentiles (the percentage of scores that fall *below* a given raw score), or cumulative percentiles (the percentage of scores that fall *at or below* a given raw score). If the scores are normally distributed, percentages and cumulative percentages should be nearly equal. But in instances where a skewed distribution occurs, cumulative percentiles should be used, rather than percentiles.

To demonstrate the relative advantage of using cumulative percentiles in the present study, a comparison of the normative performance on Trial 2 from the Recognition Memory for Pictures test (maximum score = 48) is shown below:

<u>raw score</u>	<u>f</u>	<u>cumulative percentile</u>	<u>percentile</u>
48	308	100	55.0
47	23	9.9	6.6
46	6	3.2	2.3
45	3	1.5	1.0
44	1	1.0	< 1
43	0	< 1	< 1
42	0	< 1	< 1
41	0	< 1	< 1
40	1	< 1	< 1
39	0	----	----

As with most tests involving learning tasks with more than one trial, these scores do not represent a normal learning curve, but are negatively skewed. Also, the scores are truncated, with the lowest score earned on any trial being 29 correct out of a possible 48

(or, in the above case, 40 correct). As can be seen in this table, a score of 48 places a person at the 100th *cumulative* percentile, but only at the 55th *percentile*, even though no errors are made. Therefore, it was decided that the norms from the present data are better represented by cumulative percentiles rather than by percentiles because cumulative percentiles better represent the data.

The standard error of measurement (SEM) will be used for each test as an estimate of how widely scores might vary for an individual tested an infinite number of times. Rather than report the obtained score as an estimate of the true score, the SEM presents the range in which a true score lies, taking into consideration the reliability of the measure under consideration. The SEM is calculated using the following formula:

$$SEM = s \sqrt{1 - r_{tt}}$$

where s is the standard deviation of the raw scores and r_{tt} is the measure of reliability (e.g., test-retest reliability coefficient, internal consistency coefficient) (Aiken, 1994, p. 92). For each of the three tests, r_{tt} is represented by the internal consistency for a given measure. The SEM is assumed to be normally distributed: approximately 68 times out of 100 the true score will fall between ± 1 SEM of the obtained score; approximately 95 times out of 100 it will fall between ± 2 SEM; and approximately 99 times out of 100, it will fall between ± 2.58 SEM.

RESULTS

Data from a total of 370 participants were used in analysing the demographic information for the sample. However, because of experimenter error, missing data, and attrition due to the necessity of testing older participants on a call-back basis, a small portion did not complete all three recognition memory tests. As a consequence, slightly different *n*'s were used for each test. Analyses for the entire sample will be presented here, but because there were minor changes in group means, descriptive and inferential statistics for the demographic and screening variables are reported separately in the Appendices for each recognition memory tests.

Demographic Information on Original Ten Age Groups

Demographic information and data for the screening variables for the final sample of participants are shown in Table 2. In this table, information is presented for the original 10 age groups used to recruit participants. Mean scores and standard deviations for age, education, vocabulary levels, MMSE and GDS are presented for each age group and for the total sample. As well, the gender composition is provided.

 Insert Table 2 about here

In the overall sample, 45.9% of the participants were male and 54.1% were female. A similar male-female proportion was present up to the age of 65. However, comparable numbers of male and female participants were found in the older three age groups. The mean

Table 2
Demographic Information for Ten Age Groups for Entire Sample

Demographics	Age Groups (number of subjects)										ALL (n=370)
	16-20 (n=28)	21-30 (n=30)	31-40 (n=30)	41-50 (n=48)	51-60 (n=54)	61-65 (n=43)	66-70 (n=29)	71-75 (n=41)	76-80 (n=33)	81-85 (n=34)	
Age											
Mean	18.36	24.07	35.27	45.10	56.46	63.19	68.83	72.00	77.79	82.94	55.57
(s.d.)	(1.42)	(3.05)	(2.72)	(2.80)	(2.38)	(1.42)	(1.10)	(2.60)	(1.27)	(1.39)	(20.05)
Education											
Mean	11.50	13.27	13.57	13.40	12.93	12.63	13.10	12.83	12.36	12.47	12.84
(s.d.)	(0.88)	(1.34)	(2.11)	(2.67)	(2.50)	(2.49)	(2.61)	(2.39)	(2.28)	(2.44)	(2.34)
Vocab											
Mean	44.75	45.97	53.13	52.35	54.57	55.05	54.76	55.78	54.46	56.94	53.15
(s.d.)	(10.10)	(11.75)	(8.93)	(10.30)	(8.28)	(7.89)	(10.82)	(9.06)	(9.00)	(7.77)	(9.90)
MMSE											
Mean	28.89	28.70	28.97	29.04	28.80	28.91	28.52	28.32	28.30	27.88	28.65
(s.d.)	(1.07)	(0.95)	(1.25)	(1.11)	(1.14)	(0.97)	(1.38)	(1.17)	(1.36)	(1.30)	(1.21)
GDS											
Mean	6.54	5.87	6.07	3.85	4.22	3.53	3.55	3.41	4.30	4.53	4.45
(s.d.)	(3.74)	(3.47)	(3.64)	(2.91)	(4.12)	(3.01)	(2.82)	(2.95)	(3.40)	(3.05)	(3.46)
Gender											
Males	12 (42.9)	14 (46.7)	13 (43.3)	23 (47.9)	22 (40.7)	17 (39.5)	13 (44.8)	22 (53.7)	16 (48.5)	18 (54.9)	170 (45.9)
Females	16 (57.1)	16 (53.3)	17 (56.7)	25 (52.1)	32 (59.3)	26 (60.6)	16 (55.2)	19 (46.3)	17 (51.5)	16 (47.1)	200 (54.1)

Note. Education is reported in number of years of formal schooling; Vocab = raw scores on the Vocabulary subtest of the Wechsler Adult Intelligence Scale-Revised (maximum score = 70); MMSE = Mini-Mental State Examination (maximum score = 30); GDS = Geriatric Depression Scale (maximum score = 30); Gender is reported by the number of participants as well as the percentage.

level of education for the entire age range was 12.84 years. In this sample, 43.5% of the volunteers had some post-secondary education, while 56.5% had a high school diploma or less. Because the research design required similar numbers of participants with 'high' and 'low' education (described in the Procedures section), this sample has a higher percentage of individuals with formal post-secondary education than exists in the general population (Statistics Canada, 1993). Vocabulary scores (WAIS-R) were in the average range overall as well as within each age group.

Two screening tests (GDS and MMSE) administered during this study were included to evaluate effects due to depression and/or cognitive impairment. The 18 volunteers earning scores above the cut-off point on the GDS (13/14) were not used for the normative portion of this study. No volunteer scored below the cut-off point on the MMSE (23/24). Not surprisingly, therefore, only slight variations in the scores on the MMSE and the GDS existed among age groups. In addition, these scores encompassed narrow ranges that were above the limits of the commonly used cut-off points. Because of these factors, results from the MMSE and GDS were not included in further analyses.

Pearson correlations among the demographic variables are presented for the total sample in Table 3. No significant intercorrelations were found among age, education levels, and gender. The lack of a relationship among these variables probably stems from their use as a basis for recruiting. There was a notable lack of relationship between gender and most of the variables considered in this study. However, because gender differences are of interest in psychological research, it will be included in the analyses conducted for each of the Recognition Memory Battery tests. Vocabulary was correlated with both age and education.

Before undertaking regression analyses that would assist in determining which variables should be used for stratifying the normative data, the inter-relationships between education and age, education and vocabulary, and age and vocabulary were explored.

 Insert Table 3 about here

First considered was the relationship between education and age. An analysis of variance (ANOVA) revealed that the mean number of years of education was related to age, with fewer years of education in the younger age groups ($F [9,360] = 2.155, p < .05$). Inspection of the means indicated that this small, but significant, relationship was due to two factors, both of which served to reduce variability in the younger age groups as compared to the older ones.

One factor was that individuals in the younger two age groups were more apt to have continued into high school than were individuals in the older eight age groups (see Table 4). For example, no one in the youngest two age groups reported less than a grade 10 level of education, whereas the reported years of education in the older age groups began at grade 8. Second, in the younger two age groups, there is a reduced opportunity for higher levels of education. The highest level of education reported in the group below the age of 31 was 16 years, while individuals in the older eight age groups reported completing up to 21 years of education. These factors combined to produce a lower variability in the younger age groups than in the older ones. The reduced variability was supported by Levine's test, which revealed significant differences in the variability of education reported among the age groups

Table 3

Correlations Among Demographic Variables for Full RMB Sample

VARIABLES	Age	Gender	Education	Vocabulary
Age	1.00			
Gender	-0.03	1.00		
Education	-0.02	-0.08	1.00	
Vocabulary	0.33***	-0.04	0.36***	1.00

* $p < .05$; ** $p < .01$; *** $p < .001$

Note. Education is reported in number of years; Vocabulary is the raw score from the Vocabulary subtest of the Wechsler Adult Intelligence Scale-Revised.

(Homogeneity of Variance statistic $[9,360] = 3.23, p < .01$). The variable of years of education was then split into two components: the younger two age groups (ages 16 to 30), and the older eight age groups (ages 31-85). A second ANOVA performed on the older eight age groups found no significant differences in education ($F [8,315] = 1.05, p > .05$) and no significant differences in variability (Homogeneity of Variance statistic $[7,304] = 0.56, p = .79$).

 Insert Table 4 about here

The next set of variables examined was education and vocabulary. A positive correlation between education and vocabulary levels, as measured by the WAIS-R Vocabulary subtest, was found in this study ($r = .36, p < .001$; see Table 4). To further explore this correlation, the relationship between age and vocabulary was examined. WAIS-R Vocabulary levels are known to correlate positively with age in the general population. In fact, the WAIS-R manual includes age-rated scores for its subtests; among them is Vocabulary⁴. As with education levels, an ANOVA revealed significant differences among the 10 age groups on vocabulary ($F [9,360] = 6.03, p < .01$), with equal variances among these age groups (Homogeneity of Variance statistic = 1024, $p > .05$). Visual inspection revealed that mean vocabulary scores in this community sample reflected a large jump in scores around the age of 30 (see Note 1). An ANOVA comparing the youngest two (ages 16-30) and older eight (ages 31-85) age groups indicated there was a significant difference in vocabulary scores *between*

⁴ Unfortunately, these age-rated scores could not be used, as the WAIS-R manual does not include scores beyond the age of 75.

Table 4

Education levels for original ten age groups

Age Group	Mean Years of Education	Range of Years
16 - 20	11.5	10 - 13
21 - 30	13.3	11 - 16
31 - 40	13.6	8 - 20
41 - 50	13.4	8 - 21
51 - 60	12.9	8 - 19
61 - 65	12.2	8 - 18
66 - 70	12.8	9 - 20
71 - 75	12.8	8 - 18
76 - 80	12.4	9 - 18
81 - 85	12.2	8 - 16

these two age groups ($F [1,368] = 47.67, p < .001$). An examination of the vocabulary scores across all age groups suggested that lower minimum vocabulary scores within the younger two age groups were responsible for the difference.

Picture Recognition Memory (RM Pictures)

A total of 342 individuals from the original 370 volunteers completed all three trials of the Recognition Memory for Pictures (RM Pictures) test and were used for developing normative tables for this test. The demographic information is similar to that seen in Table 2 for the entire sample (i.e., a comparable number of males and females in each group, education levels as well as variability within each age group increase after age 20, vocabulary levels increase with age, and screening variables are within the cut-off limits) and is presented in Appendix E.01.

Pearson correlations among the demographic variables were conducted for the RM Pictures test. The correlations (see Appendix F.01) are essentially the same as those previously presented in Table 4 for the entire sample. That is, significant positive correlations occurred between vocabulary and age, and between vocabulary and education. Gender was not correlated with the other demographic variables.

RM Pictures Measure Scores Across Trials for Original 10 Age Groups.

The mean scores of each age group for the four measures examined in this study are found in Table 5. These scores include the number correct on the first learning trial (Trial 1), the number of correct responses on the second learning trial (Trial 2), the total

number correct across learning trials (Total Acquisition), and the number correct on last trial (Delay Trial).

 Insert Table 5 about here

A repeated measures ANOVA revealed a significant Age effect ($F [9,332] = 3.30$, $p > .01$), Trial effect ($F [2,664] = 229.71$, $p < .001$), and Age by Trials interaction ($F [2,18] = 2.65$, $p > .001$). *Post hoc* analyses indicated that the age effect was due to statistically lower scores earned by participants in the 21-30 year age groups compared with the participants in the two age groups from ages 51-65. The Trial effect was due to the statistically lower scores earned on Trial 1 compared with Trial 2 and Delay Trial scores. Finally, the Age by Trials interaction was due to a statistically greater increase in scores among those in the two younger age groups compared with increases found for all other age groups. It should be noted, however, that the differences in scores are not very large (less than 2 items) and certainly lack clinical significance because the variances in scores within groups far outweighed the differences in scores between the groups.

It has been pointed out that in order to obtain a true measure of delayed memory, an examination of the amount the individual actually remembers should be undertaken (Tombaugh & Schmidt, 1992). Comparing an individual's delayed memory score with the standardized "delayed memory" score of the group does not reflect the amount the individual originally learned. For example, if 36 out of the 48 items were remembered by an individual on the Trial 2 of the RM Pictures test, his or her memory score would reflect

Table 5

Scores by Age Groups on Recognition Memory for Pictures Measures

RM PICTURES MEASURE	Age Group (number of participants)										
	16-20 (n=28)	21-30 (n=30)	31-40 (n=30)	41-50 (n=48)	51-60 (n=44)	61-65 (n=38)	66-70 (n=30)	71-75 (n=36)	76-80 (n=30)	81-85 (n=28)	ALL (n=342)
Trial 1 Mean (s.d.)	44.21 (3.58)	43.83 (3.70)	44.47 (4.04)	45.44 (2.58)	46.16 (2.51)	46.45 (1.70)	45.87 (3.73)	46.11 (2.58)	45.80 (2.54)	45.82 (2.40)	45.49 (3.03)
Trial 2 Mean (s.d.)	47.68 (0.61)	47.50 (1.57)	47.73 (0.91)	47.96 (0.20)	47.98 (0.15)	47.95 (0.23)	47.97 (0.18)	47.89 (0.40)	47.87 (0.43)	47.64 (0.73)	47.84 (0.65)
Total Acquisition Mean (s.d.)	91.89 (3.95)	91.33 (4.16)	92.20 (4.51)	93.40 (2.66)	94.14 (2.56)	94.40 (1.69)	93.83 (3.89)	94.00 (2.82)	93.67 (2.83)	93.46 (2.66)	93.32 (3.29)
Delay Trial Mean (s.d.)	47.75 (0.52)	47.80 (0.55)	47.90 (0.31)	47.79 (1.17)	47.89 (0.39)	47.97 (0.16)	47.97 (0.18)	47.83 (0.45)	47.80 (0.61)	47.86 (0.45)	47.86 (0.58)

75% correct. However, if that individual earned a score of 36 on the Delay Trial, the score would still produce a memory score of 75% correct, even though 100% of the information previously learned by that individual was remembered. The 75% correct on the Delay Trial would reflect the amount of information acquired, but not the percent retrieved from storage.

In many situations, the amount recalled on the delayed learning trial does not differ significantly from the amount remembered on the last learning trial. This is particularly true in a community-based sample which screens out scores of those with known cognitive or emotional difficulties. However, it is important to verify that the two memory scores equate. For this reason, 'proportion remembered' (% Rem) scores on the Delay Trial were determined in order to examine the amount forgotten between Trial 2 and the Delay Trial. The %Rem score was calculated by comparing each individual's delay trial score with his or her final learning trial score ($\text{Delay trial} \div \text{Trial 2} \times 100$). It was found that for all age groups, mean scores on Trial 2 and the Delay Trial were nearly equal, indicating that for most individuals, information remembered after the second learning trial was retained after a 20 to 30 minute delay (see Appendix G). Therefore, %Rem will not be included in the remainder of the analyses for the RM Pictures test.

Norming the Recognition Memory for Pictures Test

In order to use a test in a clinical setting, three important steps must first be taken. First, norms must be derived in order to compare an individual's performance to that of an appropriate community-based sample. Second, reliability must be examined based on that

sample. And third, validity must be evaluated to determine if the test is assessing the intended concept.

Although normative information is not sufficient on its own to determine whether a test is valid, it is an essential component in examining the reliability of a test, and contributing information that subsequently can be used to establish the validity of scores obtained in a clinical setting. Developing the norms for the RM Pictures involved three steps: (1) Conducting regression analyses to determine appropriate stratifying variables; and (2) determining the composition of the normative groups as required; and (3) constructing norms.

Regression analyses. Regression analyses were performed on the first and last (second) learning trials, on the total acquisition scores, and on the delay scores to determine how the normative data should be stratified. The purpose of the regression analyses was to determine (a) which variables would predict performance on each measure, and (b) the amount of variance explained by each variable. All variables were allowed to enter freely in a stepwise fashion. As a result, the factors which most highly affect performance on the RM Pictures could be identified, and individual contributions to the total variance could be examined.

Examination of the data collected for the full sample suggested three variables that could be used for stratification: age, years of education, and vocabulary. Vocabulary level was not included in the present analysis, however, for two reasons. First, vocabulary was shown to be moderately correlated with years of education ($r = .34$, $p < .001$) and with age ($r = .35$, $p < .001$), and these latter two variables are already included in the regression

analysis. Second, the use of vocabulary scores as a stratification variable would make the interpretation of the norms dependent upon the administration of the WAIS-R Vocabulary subtest. This would impose a condition which would not be met in many clinical applications, thereby seriously restricting the clinical utility of the test. Finally, since gender differences have been reported in some memory studies (e.g., Cohen, 1988; Trahan & Quintana, 1990), this variable was included in the regression analyses.

Results of the regression analyses for the RM Pictures measures are shown in Table 6. Age, Gender and Education each contributed significantly to scores on Trial 1 and to Total Acquisition. Mean scores for males were slightly lower than those of females on Trial 1 (1 to 1.5 items). Age alone contributed to variance on Trial 2. On the Delay Trial, none of the three variables contributed significantly to explain the variance in scores. As was shown in Table 5, however, mean scores within each of the RM Pictures measures varied little across age groups (Trial 1 mean scores ranged from 44.21 to 46.45 words correct, with no more than one item difference between any two adjacent age groups) and what little variation there was did not appear to be systematic. For example, with the highest mean scores appearing in the 61-65 age group and the lowest in the 21-30 age group. On Trial 2, mean scores varied even less, ranging across the 10 age groups from 47.50 to 47.98 words correct (less than half of one picture). Added to this small variability in scores was the contribution of the large number of participants used in this study. Consequently, even though age was statistically significant, it was not judged to be of sufficient magnitude to warrant using it as a stratifying variable. The same comments may be made about gender and levels of education. Therefore, it was determined that the

normative data for this test could best be represented without stratifying variables. However, because education differences have been found in previous memory research, caution should be taken when interpreting these findings if low levels of education are reported for an individual patient.

Insert 6 about here

Using cumulative percentiles, norms for each trial score are presented in Table 7, and the norms for total acquisition scores are shown in Table 8.

Insert Table 7 & 8 about here

An examination of the present data revealed a negatively skewed distribution of scores, primarily attributable to the range of participants earning near-perfect scores, particularly on the second trial and retention trial.

Reliability

While normative information is essential for cognitive testing, the further question of the reliability of the norms must be addressed. A variety of reliability measures are available for the test developer, including inter-rater reliability, test-retest reliability, alternate forms reliability, and internal consistency.

Unfortunately, most measures of reliability are inappropriate for examining the RM Pictures test. In memory tests using a recognition format, inter-rater reliability is seldom a

Table 6

Summary of Stepwise Regressions for Recognition Memory for Pictures Measures with Age Entered as the First Block

RMP MEASURES	R^2	F	STEP	R^2_{ch}	r_{sp}
Trial 1	0.10	13.07***			
Age			1	0.06	0.25***
Gender			2	0.02	0.16**
Education			3	0.02	0.12*
Trial 2	0.02	2.69*			
Age			1	0.01	0.11*
Gender			(n.s.) ¹		0.06
Education			(n.s.)		0.07
Total Acquisition	0.11	13.48***			
Age			1	0.06	0.25***
Gender			2	0.03	0.16**
Education			3	0.02	0.12*
Delay Trial	(n.s.)				
Age					0.06
Gender					0.07
Education					0.04

* $p < .05$; ** $p < .01$; *** $p < .001$

Note. Education is reported in number of years; Trial 1 and Trial 2 are the two learning trials from the Recognition Memory for Designs test, while Total Acquisition represents a sum of the scores of these two trials, and Delay Trial is the delayed memory trial. 'n.s.' represents a non-significant contribution to the regression equation.

¹'n.s.' represents a non-significant contribution to the regression equation.

Table 7

Cumulative Percentiles for Recognition Memory for Pictures Learning and Delay Trials
for Normative Sample

Raw Scores	Cumulative Percentiles		
	Trial 1	Trial 2	Delay
48	100	100	100
47	70	10	10
46	49	3	2
45	37	2	1
44	26	1	< 1
43	19	< 1	< 1
42	14	< 1	< 1
41	11	< 1	< 1
40	8	< 1	< 1
39	5	< 1	< 1
38	3	< 1	< 1
37	2	< 1	< 1
36	2	< 1	< 1
35	2	< 1	< 1
34	1	< 1	< 1
33	1	< 1	< 1
32	< 1	< 1	< 1

Table 8

Cumulative Percentiles for Recognition Memory for Pictures Total Acquisition
Scores for Normative Sample

Raw Scores	Cum. %ile
96	100
95	70
94	51
93	38
92	27
91	21
90	15
89	12
88	9
87	7
86	4
85	4
84	3
83	2
82	2
81	1
80	1
79	1
78	1
77	1
76	< 1

necessity because of the objective scoring criteria. When pilot testing for the Recognition Memory Battery was undertaken, two raters occasionally recorded answers for the same participant at the same time. In these cases, no differences in scores occurred. Therefore, a measure of inter-rater reliability was felt to be unnecessary with the present data set. In addition, because of the large number of stimulus items required for the response trials (line drawings of 480 unique items), only one version of this test was developed; therefore, reliability between alternate forms was not possible. Finally, there are two reasons that argue against using test-retest procedures to measure the reliability of learning and memory tests (Crosson et al., 1988; Schmidt & Tombaugh, 1995). First, reliability estimates may be inflated because of practice effects. That is, because the information has been examined repeatedly during the test administration and because a large portion of learned information is retained for a relatively long period of time, a response bias is present for subsequent administrations. Most cognitive testing involves stimuli which are presented only one time. For example, subtests on the Wechsler Adult Intelligence Scales present auditory and/or visual information one time only, and no feedback is given as to the accuracy of the answer. Still, the WAIS-R Manual cautions against learning effects. During learning tests, the material is examined several times so that the individual can self-correct by re-examining the same stimulus material. Second, visual memory is particularly persistent over time, with the result that an appropriate test-retest interval long enough to counteract these learning effects was beyond the scope of this study. Therefore, test-retest reliability was not felt to be appropriate for evaluating this test.

Internal consistency of test items on the RM Pictures test was examined, however, on a measure-by-measure basis. Because the items involved multiple choice, Cronbach's Coefficient alpha was chosen to examine inter-item correlations. Cronbach's test examines the co-variance among all item scores used in the scale. The inter-item correlation for the RM Pictures trials are as follows: Trial 1, 0.73; Trial 2, 0.78; and Delay Trial, 0.76. Inter-item correlation for the Total Acquisition measure was 0.84.

Although it was assumed that all target items reflected the same level of difficulty, it is possible that some were easier and some were harder. Therefore, the relative difficulty of test items was examined by comparing the percentage of correct and incorrect responses for each item. To find mean error rates, the percent of participants who had an incorrect response for each item was found. For example, 9 of the 342 participants made incorrect choices on the first trial for the item <paint brush>. This represented 2.6% of the sample ($9 \div 342$). Error rates for individual target items ranged from 0.6% to 17.5% wrong on the first trial, 0% to 2.7% wrong on the second trial, and 0% to 3.5% wrong on the delay trial (see Table 9).

 Insert Table 9 about here

To further explore item difficulty, a comparison of 'easier' and 'harder' items was made across trials. To accomplish this, after determining the mean percent wrong for each item, the items on each trial were rank ordered. If any items shared the same rank, they were assigned the lowest rank value, and the next higher rank reflected the number

Table 9
Rank Order of Item Difficulty for Recognition Memory for Pictures Test

ITEM	TRIAL 1		TRIAL 2		DELAY TRIAL	
	RANK ORDER	'EASIER' / 'HARDER'	PERCENT WRONG	RANK ORDER	'EASIER' / 'HARDER'	PERCENT WRONG
baseball bat	1	easier	0.6	14	easier	0.3
lighter	2.5	easier	0.9	22.5	easier	0.6
muffin tin	2.5	easier	0.9	14	easier	0.3
acorn	5	easier	1.2	14	easier	0.3
spinning wheel	5	easier	1.2	14	easier	0.3
tin can	5	easier	1.2	14	easier	0.3
music staff	7	easier	1.5	22.5	easier	0.6
dart	8	easier	1.8	14	easier	0.3
kleenex box	9	easier	2.0	45.5	harder	1.5
whistle	10	easier	2.4	31	harder	0.9
flashlight	12	easier	2.6	39	harder	1.2
tennis racquet	12	easier	2.6	31	harder	0.9
				5	easier	0.0

Table 9 (cont'd)

ITEM	TRIAL 1		TRIAL 2		DELAY TRIAL	
	RANK ORDER	'EASIER' / 'HARDER' / PERCENT WRONG	RANK ORDER	'EASIER' / 'HARDER' / PERCENT WRONG	RANK ORDER	'EASIER' / 'HARDER' / PERCENT WRONG
toothbrush	12	easier 2.6	22.5	harder 0.6	17	easier 0.3
feather	14	easier 2.9	45.5	harder 1.5	5	easier 0.0
paper clip	15	easier 3.0	14	easier 0.3	17	easier 0.3
saxophone	16	easier 3.2	22.5	easier 0.6	17	easier 0.3
pendant	17.5	easier 3.8	39	harder 1.2	30.5	harder 0.6
toaster	17.5	easier 3.8	14	easier 0.3	17	easier 0.3
sewing machine	19	easier 4.1	1	easier 0.0	40	harder 0.9
rose	20	easier 4.4	39	harder 1.2	37	harder 0.9
scissors	21.5	easier 4.9	22.5	harder 0.6	30.5	harder 0.6
suitcase	21.5	easier 4.9	39	harder 1.2	45	harder 2.0
fishing net	23.5	easier 4.9	1	easier 0.0	5	easier 0.0
wrench	23.5	easier 5.0	1	easier 0.0	17	easier 0.3

Table 9 (cont'd)

ITEM	TRIAL 1		TRIAL 2		DELAY TRIAL	
	RANK ORDER	'EASIER' / 'HARDER'	PERCENT WRONG	RANK ORDER	'EASIER' / 'HARDER'	PERCENT WRONG
paint brush	25	harder	5.0	36	harder	1.2
life preserver	26.5	harder	5.5	48	harder	2.7
pail	26.5	harder	5.5	47	harder	1.9
wheelbarrow	28	harder	5.6	22.5	easier	0.6
mouse	29	harder	5.8	22.5	easier	0.6
lightbulb	30	harder	6.2	1	easier	0.0
birthday cake	31.5	harder	6.7	31	harder	0.9
pencil	31.5	harder	6.7	39	harder	1.2
stool	33	harder	7.0	39	harder	1.2
candle	34	harder	7.1	39	harder	1.2
umbrella	35	harder	7.3	39	harder	1.2
football	36	harder	7.6	31	harder	0.9
				44	harder	1.5

Table 9 (cont'd)

ITEM	TRIAL 1		TRIAL 2		DELAY TRIAL	
	RANK ORDER	'EASIER' / 'HARDER'	PERCENT WRONG	RANK ORDER	'EASIER' / 'HARDER'	PERCENT WRONG
carrot	37	harder	8.0	22.5	easier	0.6
stepstool	38.5	harder	8.5	14	easier	0.3
motorcycle	38.5	harder	8.5	1	easier	0.0
leaf	40	harder	9.5	1	easier	0.0
ice cream cone	41	harder	10.5	1	easier	0.0
iron	42	harder	10.6	14	easier	0.3
jack-in-the-box	43.5	harder	11.1	31	harder	0.9
stapler	43.5	harder	11.1	14	easier	0.3
roller skate	45	harder	11.7	22.5	easier	0.6
shopping cart	46	harder	12.0	22.5	easier	0.6
birdhouse	47	harder	12.6	31	harder	0.9
duck	48	harder	17.5	31	harder	0.9

contained in the next lower rank. For example, three items were found to have a 2.6% error rate on Trial 1. This fell at the 11th position in the ranking. They were each assigned a rank order of 12, and the next higher error rate (2.9%) was rank ordered at 14. When all items were rank ordered, the 24 items with the least errors ranking from 1 to 24 were assigned to the 'easier' group, and the other 24 items (ranking from 25 to 48) were assigned to the 'harder' group. This resulted in a half-way split occurring between a 5.0% error rate (included in the 'easier' group) and 5.3% error rate (included in the 'harder' group) on Trial 1, between 0.6% and 1.2% on Trial 2,⁵ and between 0.3% and 0.6% on the final delay trial.

While some of the items remained in the 'easier' and 'harder' halves across trials, this was not a consistent finding. For example, on Trial 1, 2.9% of the participants failed to select the target picture <feather>, resulting in a rank order of 14 and placing it in the 'easier' half of the items. On the second learning trial, however, <feather> was in the 'harder' half, with 1.5% of the participants incorrectly choosing an alternate picture, resulting in a rank order of 45 for that item. Then, on the delay trial, it was rank ordered #1, with no errors made when choosing that item. On the other hand, the target picture <lightbulb> was rank ordered #30 on the first trial with 6% choosing incorrectly, while on Trial 2 and the delay trial it was ranked #1 and #48, respectively, because of error rates at 6% and 3.5%, respectively. It can be seen in Table 9 that many of the items shifted between the 'easier' and 'harder' groups from trial to trial. This may reflect random

⁵ On Trial 2, item numbers 19 through 28 shared a rank order of 22.5, resulting in the half-way mark (24 easy/24 hard) falling somewhere in this rank. Rather than arbitrarily choosing 'easier' and 'harder' items from this rank, all ten items at this range (19) were included in the 'easier' group.

errors, errors due to the specific foils included for any given response trial, or, more likely, it may reflect the extremely limited range of errors made during the last two trials (a mean-errors maximum of 2.7 wrong on Trial 2, and 3.5 wrong on the Delay Trial).

Finally, the location on the response page of incorrectly chosen items was examined to determine if a location bias existed in responses (see Table 10). It can be seen that when stimulus item response locations are summed across all three trials, ordinal ranking varied across the three trials.⁶

 Insert Table 10 about here

Standard Error of Measurement (SEM). The SEM is reported in Table 11 for each of the four measures evaluated.

 Insert Table 11 about here

To illustrate the utility of the SEM in this case, consider the example of a 40 year old person who scores 45 out of 48 on Trial 2 of the RM Pictures. Thus, using the SEM of 0.42 provided in Table 11, it can be concluded with 68% confidence that with an observed score of 45, this person's true score falls between 44.58 and 45.42, with 95% confidence

⁶ For a more complete description of item location analysis, see Appendix H.01.

Table 10

Analysis of Item Responses by Location on Response Trial Page for Recognition Memory
for Pictures (n = 342)

LOCATION ON PAGE ¹	TRIAL 1	TRIAL 2	DELAY TRIAL	TOTAL WRONG
1 Number incorrect <i>Percent of Trial</i>	285 (2) ² 29.5%	15 (4) 13.9%	16 (4) 17.4%	316 27.1%
2 Number incorrect <i>Percent of Trial</i>	331 (1) 34.2%	26 (3) 24.1%	25 (2) 27.2%	382 32.7%
3 Number incorrect <i>Percent of Trial</i>	170 (4) 17.6%	31 (2) 28.7%	27 (1) 29.3%	228 19.5%
4 Number incorrect <i>Percent of Trial</i>	181 (3) 18.8%	36 (1) 33.3%	24 (3) 26.1%	241 20.7%
Total incorrect	967	108	92	1167

¹ Item in the top left corner is indicated as #1; top right, #2; bottom left, #3; bottom right, #4.

² Numbers in parentheses indicate ordinal sequence from most difficult (1) to least difficult (2) for that location of the item on the page.

Table 11

Variability (SEM) Within Age Groups for Each Recognition Memory for Pictures Trial
and for Total Acquisition

RMW MEASURES	STANDARD DEVIATION OF SCALE SCORES	ALPHA LEVEL	S.E.M.
TRIAL 1	6.89	0.78	3.25
TRIAL 2	3.61	0.58	0.42
TOTAL ACQUISITION	3.29	0.78	1.53
DELAY TRIAL	0.58	0.76	0.29

Note: S.E.M. = Standard Error of Measurement; calculated by multiplying the standard deviation of the mean raw scale scores times the square root of 1 minus the Cronbach's alpha level for that scale (from inter-item correlations)

that it will fall between 45.16 and 46.84, and with 99% confidence this score will fall between 43.91 and 46.08.

Correlated Measures

The first step in evaluating the validity of a test is to gather normative information in order to determine what a 'normal' range of scores is for that test. Although a test used in a clinical setting must also be evaluated with a clinical population, this study focused on examining test reliability and developing normative information in a cognitively intact population for the three tests included in the Recognition Memory Battery. Using this normative population, analyses were conducted to examine the relationships of RM Pictures test scores with various other cognitive measures to determine convergent and discriminant support.

Because the administration of this test battery took two to four hours depending on the individual, older participants generally required two testing sessions. Although 342 participants completed all three trials of the RM Pictures test, some did not complete the testing battery in its entirety. A subset of 328 participants who did complete all of the neuropsychological measures and 292 participants completing all G-LAMB measures were used to analyze convergent and discriminant support for the RM Pictures. This was felt to be a reasonable decision, as descriptive statistics for the G-LAMB sample were highly similar to those found for the RM Pictures sample.

Relationship with other memory measures. In order to examine concurrent support, the relationship was examined between scores on the RM Pictures and comparable scores on the Geriatric Learning and Memory Battery. Like the RM Pictures,

the G-LAMB tasks were designed to be relatively easy compared to tests used to assess memory deficits in the general population. The G-LAMB involves the learning and memory of a list of words (presented orally), a simple geometric figure (presented visually), and a short paragraph (presented both orally and visually).⁷ For each test, measures reviewed were the first and last learning trials, the total acquisition score, the delayed memory trial, and the percent remembered.

Pearson correlations between the RM Pictures measures and the G-LAMB Paragraph, Word List, and Rectangular Figure scores are found in Tables 12, 13, and 14, respectively. None of the RM Pictures measures was correlated with the G-LAMB Paragraph scores. Trial 1 scores on the RM Pictures were correlated with the last learning trial scores of the Word List, and with the Word List Total Acquisition scores. The Total Acquisition scores on the RM Pictures were likewise correlated with the final acquisition trial and total acquisition scores on the Word List. And finally, the Delayed Memory scores on the RM Pictures were correlated with Trial 1 scores on the Word List. It should be noted, however, that all of these correlations were relatively low, and that little variance appeared in the scores on any of the RM Pictures measures, making meaningful correlations unlikely. On the G-LAMB Rectangular Figure test, correlations were found between Trial 2 scores on the RM Pictures and Total Acquisition scores on the Rectangular Figure. In addition, scores on the RM Pictures delayed memory measure were correlated with scores on the Delay Trial of the Rectangular Figure. All other correlations failed to reach significance levels.

⁷ A more complete description of the G-LAMB may be found in Appendix C.

Insert Table 12, 13, & 14 about here

Relationship with other neuropsychological factors. RM Pictures scores were also examined in relation to results from a set of neuropsychological tests designed to assess other cognitive abilities. Two measures were used to examine verbal ability: word fluency (the Controlled Word Fluency and Category tests, which together will be referred to as Word Fluency) and confrontational naming (Boston Naming Test). In addition, a test to assess visual tracking and attention was used (Trails A and B). If memory tests are to be useful, they should be related only mildly to moderately with scores on these other standardized tools. Finally, it would be difficult to attempt a meaningful measure of visual memory when the individual taking the test had poor visual perception. As a result, scores from the copy trial of the G-LAMB Rectangular Figure test were used as a rough measure of visual perception.

It is difficult to determine the functional overlap of what is being measured when comparing memory tests with other neuropsychological measures, since memory ability is implicated to some extent in the performance on each of the tests included, except on the copy trial from the G-LAMB Rectangular Figure. However, correlations should be lower between the RM Pictures measures and the other neuropsychological measures than between the RM Pictures measures and other memory tests.

Pearson correlations between RM Pictures measures and the neuropsychological measures are found in Table 15. Significant correlations exist between the RM Pictures Delay Trial and scores on Word Fluency, and between the Boston Naming Test and the

Table 12

Pearson Correlations Between Comparable Recognition Memory for Pictures and G-LAMB Paragraph Measures (n = 300)

RMP MEASURES	G-LAMB PARAGRAPH			
	TRIAL 1	TRIAL 2	TOTACQ	DELAY TRIAL
TRIAL 1	0.03	0.01	0.02	0.04
TRIAL 2	0.04	0.03	0.04	0.03
TOTACQ	0.03	0.01	0.02	0.05
DELAY TRIAL	0.11	0.06	0.10	0.07

* $p < .05$; ** $p < .01$; *** $p < .001$

Note: Trial 1 is the first learning trial from each test; the second measure (Trial 'x') represents the score on the last learning trial of each test; TotAcq represents a sum of the scores of all learning trials for each test.

Table 13

Pearson Correlations Between Comparable Recognition Memory for Pictures Measures
and G-LAMB Word List Measures (n = 300)

RMP MEASURES	G-LAMB WORD LIST			
	TRIAL 1	TRIAL 5	TOTACQ	DELAY TRIAL
TRIAL 1	0.01	0.13*	0.18**	-0.04
TRIAL 2	0.12*	0.03	0.05	-0.04
TOTACQ	0.02	0.12*	0.17**	-0.04
DELAY TRIAL	0.13*	0.03	-0.08	0.09

* $p < .05$; ** $p < .01$; *** $p < .001$

Note: Trial 1 is the first learning trial from each test; the second measure (Trial 'x') represents the score on the last learning trial of each test; TotAcq represents a sum of the scores of all learning trials for each test.

Table 14

Pearson Correlations Between Comparable Recognition Memory for Pictures Measures and G-LAMB Rectangular Figure Measures (n = 300)

RMP MEASURES	G-LAMB RECTANGULAR FIGURE			
	TRIAL 1	TRIAL 3	TOTACQ	DELAY TRIAL
TRIAL 1	0.00	-0.02	0.01	-0.02
TRIAL 2	0.11	0.08	0.12*	0.11*
TOTACQ	0.01	-0.01	0.02	-0.01
DELAY TRIAL	0.08	0.10	0.13*	0.13*

* $p < .05$; ** $p < .01$; *** $p < .001$

Note: Trial 1 is the first learning trial from each test; the second measure (Trial 'x') represents the score on the last learning trial of the test; TotAcq represents a sum of the scores of all learning trials for each test.

total acquisition and delayed memory trials of the RM Pictures. These correlations, however, represent only a small portion of the variance on each of the RM Pictures measures.

- - - - -
 Insert Table 15 about here
 - - - - -

Another way used to assess the covariance of memory batteries with other neuropsychological tests is to conduct a factor analysis. However, factor analysis requires moderate to high zero-order correlations (over 0.30). Since all of the correlations were below 0.30, this procedure was not used.

Relationship with reported depressive symptomatology. Finally, the relationship of scores on the RM Pictures and on the Geriatric Depression Scale (GDS) was explored. Performance on most clinical memory tests is affected by depression (Delis, 1989). However, the effects of depression on performance on the RM Pictures should be minimal, since the task requires relatively little conscious effort. To test this hypothesis for the RM Pictures, each of the 24 participants whose GDS was above the cut-off level (13/14) was matched on a variety of demographic measures (age, gender, years of education, and vocabulary scores) with a participant whose GDS score was within the 'normal' range. The matching rules required that paired individuals were matched as closely as possible on age, gender education, and vocabulary level.

A description of these matched groups may be found in Appendix I.01. As would be expected in a matched sample, the 'high' and 'low' depression groups appeared similar

Table 15

Pearson Correlations Between Recognition Memory for Pictures Measures and Scores on Neuropsychological Factors (n = 229)

RMP MEASURES	NEUROPSYCHOLOGICAL MEASURES				
	WORD FLUENCY	TRAILS A	TRAILS B	BOSTON NAMING TEST	G-LAMB FIGURE COPY
TRIAL 1	0.02	0.06	-0.10	0.20	0.03
TRIAL 2	0.20	-0.05	-0.15*	0.10	0.03
TOTACQ	0.05	0.05	-0.11	0.20**	0.03
DELAY TRIAL	0.19**	-0.02	-0.09	0.19**	0.05

* $p < .05$; ** $p < .01$; *** $p < .001$

Note. RMW measures include Trials 1 and 2 (learning trials); TotAcq (sum of the scores of both learning trials); Word Fluency represents an additive score of words identified on F-A-S and Animals; Trails A and Trails B are measured in seconds to completion of the task; BNT (Boston Naming Test) is represented by the total number of items correctly named without cues; and G-LAMB Figure Copy is the score on the Copy Trial for the Rectangular Figure in the G-LAMB.

to each other in terms of age, gender, education, vocabulary levels, and MMSE scores. ANOVAs showed that there were no significant differences between the two groups on any of these demographic variables.

To determine whether depression scores affected memory ability on the RM Pictures test, an analysis of variance was undertaken for the Trial 1, Trial 2, Total Acquisition and Delay Trial scores on the RM Pictures using the 'high' and 'low' depression groups. There were no significant differences between the two groups on any of these measures, indicating that self-reported mood did not reflect performance on this test (see Table 16). Interestingly, it can be seen that there was a small, but non-significant trend for performance scores in the 'low depression' group to be lower on the RM Picture measures than those in the 'high depression' group. That is, RM Pictures scores on three of the four measures were higher for those in the 'high depression' group than for those in the 'low depression' group. These differences, however, represented little clinical significance, as scores varied by less than one point.

 Insert Table 16 about here

Next, G-LAMB scores for the 'high' and 'low' depression groups were examined. Mean scores for these measures may be found in Appendix I.02. An ANOVA was used to determine whether differences existed between the two groups on the G-LAMB Paragraph, Word List and Rectangular Figure. Differences between the two groups did not reach levels of statistical significance.

Table 16

Mean Scores Comparing Recognition Memory for Pictures Measures for 'High' and 'Low' Depression Groups

RMP MEASURES	'Low' Depression (GDS Score below 14)	'High' Depression (GDS Score 14 and above)
TRIAL 1		
<i>Mean</i>	44.83	45.42
<i>(s.d.)</i>	(3.34)	(3.34)
TRIAL 2		
<i>Mean</i>	47.88	47.79
<i>(s.d.)</i>	(0.43)	(0.66)
TOTAL ACQUISITION		
<i>Mean</i>	92.71	93.21
<i>(s.d.)</i>	(3.65)	(3.68)
DELAY TRIAL		
<i>Mean</i>	47.67	47.83
<i>(s.d.)</i>	(0.64)	(0.38)

* $p < .05$; ** $p < .01$; *** $p < .001$

Finally, an ANOVA was conducted to examine differences between the two groups on the neuropsychological measures (BNT, Trails A & B, and Word Fluency). Differences between these two groups on these measures did not reach statistically significant levels. However, a pattern of scores similar to differences on RM Picture measures was present, in that lower performance scores were found among those in the 'low depression' group than in the 'high depression' group (see Appendix I.03).

Word Recognition Memory (RM Words)

All three trials of the Recognition Memory for Words (RM Words) test were completed by a total of 340 individuals from the original 370 volunteers. These 342 participants were used for developing normative tables for this test. Demographic information for the participants used in this portion of the study, which is similar to that found for the entire sample, is presented in Appendix E.02. As with the Recognition Memory for Pictures test, participants ranged from 16-85 years with a comparable number of males and females in each group, education levels as well as variability within each age group increased after age 20, vocabulary increased with age and screening variables were within the cut-off limits.

Pearson correlations among the demographic and screening variables were conducted for the RM Words test. These correlations (see Appendix F.02) were essentially the same range as those seen in Table 4 for the entire sample. This included significant positive correlations between vocabulary and age, and between vocabulary and education. As before, gender was not correlated with other demographic variables.

RM Words measure scores for original 10 age groups. Descriptive details about the mean scores for each age group for the four measures in this study are found in Table 17. As before, scores include the number correct on the first learning trial (Trial 1), the number correct on the second learning trial (Trial 2), the total number correct across learning trials (Total Acquisition), and the number correct on last trial (Delay Trial).

- - - - -
 Insert Table 17 about here
 - - - - -

A visual inspection of the data shows that there was a slight decrease in mean scores on each of the four RM Words measures as age increased. Therefore, a repeated measures ANOVA was conducted which revealed a significant Trial effect ($F [2, 660] = 751.87.22, p < .001$). *Post hoc* analyses indicated that the Trial effect was due to significantly higher scores on the second learning trial compared to initial learning scores. The Age by Trials interaction was not statistically significant ($F [2,18] = 0.87, p > .05$), nor was a significant Age effect was found ($F [9,330] = 1.30, p > .05$), indicating that scores between age groups and the rate of learning over the trials did not differ significantly.

As with the RM Pictures test, delayed memory scores were comparable to scores on the last learning trial (see Appendix G), making further examination of the proportion remembered unnecessary .

Table 17
Scores by Age Groups on Recognition Memory for Words Measures

RM PICTURES MEASURE	Age Groups (number of participants)										
	16-20 (n=28)	21-30 (n=30)	31-40 (n=30)	41-50 (n=48)	51-60 (n=43)	61-65 (n=38)	66-70 (n=27)	71-75 (n=39)	76-80 (n=30)	81-85 (n=27)	ALL (n=340)
Trial 1 <i>Mean</i> (<i>s.d.</i>)	38.18 (6.37)	38.77 (5.99)	36.97 (8.31)	37.48 (7.56)	39.42 (6.10)	37.50 (7.67)	36.04 (7.07)	36.15 (6.66)	36.10 (6.34)	36.74 (6.16)	37.23 (6.89)
Trial 2 <i>Mean</i> (<i>s.d.</i>)	45.96 (3.44)	44.70 (3.95)	46.10 (2.56)	46.19 (3.00)	46.23 (2.21)	45.32 (4.79)	45.56 (3.76)	45.08 (3.27)	43.43 (4.83)	44.81 (3.42)	45.41 (3.61)
Total Acquisition <i>Mean</i> (<i>s.d.</i>)	84.14 (8.96)	81.47 (9.02)	83.07 (10.19)	83.67 (9.77)	85.65 (7.89)	82.82 (11.54)	81.59 (10.09)	81.23 (9.38)	79.53 (10.42)	81.56 (8.67)	82.64 (9.66)
Delay Trial <i>Mean</i> (<i>s.d.</i>)	45.75 (3.41)	45.63 (3.44)	46.07 (3.18)	45.67 (3.81)	46.33 (2.41)	45.63 (3.63)	45.30 (4.20)	44.64 (4.63)	43.47 (4.93)	45.30 (3.48)	45.41 (3.79)

Norming the Recognition Memory for Words Test (RM Words)

Results from this sample were used to examine the nature of the normative information for the RM Words test. As with the RM Pictures test, developing the norms for the RM Words test involved three steps: (1) conducting regression analyses to determine appropriate stratifying variables; (2) determining the composition of the normative groups as required; and (3) constructing norms.

Regression analyses. Regression analyses were used to determine how the normative data should be stratified. Age, years of education, and gender were allowed to enter freely in a stepwise fashion in order to determine whether any of these variables would predict performance on each measure, and to examine the amount of variance explained by each of the three variables. Vocabulary levels were not included for reasons presented in the previous section (RM Pictures).

The results of the regression analyses for the RM Words measures indicated that none of the variables contributed significantly to the overall variance in scores (see Table 18). As shown in this table, performance on the four measures was highly similar for participants in all of the original ten age groups indicating that comparable scores would be earned on each of these measures at any given age. As with the RM Pictures scores, variance was not systematic. Therefore, the normative data was presented in a single table with no stratifying variables (see Tables 19 and 20).

 Insert Table 18, 19 and 20 about here

Table 18

Summary of Stepwise Regressions for Recognition Memory for Words Measures

RMD MEASURES	R^2	F	STEP	R^2_{ch}	r_{sp}
Trial 1	0.03	1.05 (n.s.) ¹			
Age			1	0.003	-0.06
Education			2	0.030	0.17**
Gender				n.s.	
Trial 2	0.05	3.50 (n.s.)			
Age			1	0.010	-0.01
Education			2	0.038	0.17**
Gender			n.s. ²		
Total Acquisition	0.04	2.04 (n.s.)			
Age			1	0.006	-0.08
Education			2	0.039	0.18**
Gender			n.s. ²		
Delay Trial	0.05	3.74 (n.s.)			
Age			1	0.01	-0.10
Education			2	0.04	0.18**
Gender			n.s. ²		

* $p < .05$; ** $p < .01$; *** $p < .001$

Note: Education is reported in number of years; Trial 1 and Trial 2 are the two learning trials from the Recognition Memory for Designs test, while Total Acquisition represents a sum of the scores of these two trials, and Delay Trial is the delayed memory trial. 'n.s.' represents a non-significant contribution to the regression equation.

¹ 'n.s.' represents a non-significant contribution to the regression equation.

Table 19

Cumulative Percentiles for Recognition Memory for Words Learning and Delay Trials for Normative Age Groups.

Raw Scores	Cumulative Percentiles		
	Trial 1	Trial 2	Delay
48	100	100	100
47	96	64	58
46	91	47	46
45	88	36	37
44	82	26	31
43	77	20	19
42	73	14	16
41	69	11	12
40	65	9	9
39	60	6	7
38	56	5	6
37	50	4	6
36	43	3	5
35	39	3	4
34	34	3	3
33	28	2	2
32	25	1	1

Raw Scores	Cumulative Percentiles		
	Trial 1	Trial 2	Delay
30	17	1	1
31	22	1	1
29	14	1	<1
28	12	<1	<1
27	10	<1	<1
26	7	<1	<1
25	5	<1	<1
24	4	<1	<1
23	3	<1	<1
20	1	<1	<1
19	1	<1	<1
18	1	<1	<1
17	1	<1	<1
16	<1	<1	<1
15	<1	<1	<1

Table 20

Cumulative Percentiles for Recognition Memory for Words Total Acquisition Scores for
Normative Age Groups.

Raw Scores	Cum. %ile	Raw Scores	Cum. %ile	Raw Scores	Cum. %ile
96	100	75	21	54	1
95	96	74	19	53	1
94	93	73	17	52	1
93	88	72	15	51	1
92	84	71	14	50	1
91	80	70	11	49	<1
90	75	69	10	48	<1
89	72	68	9	47	<1
88	68	67	7	46	<1
87	64	66	5	45	<1
86	60	65	4	44	<1
85	56	64	4	43	<1
84	52	63	4	42	<1
83	50	62	4	41	<1
82	47	61	4	40	<1
81	40	60	3	39	<1
80	36	59	2	38	<1
79	40	58	2	37	<1
78	36	57	1	36	<1
77	32	56	1	35	<1

A ceiling effect and truncated scores existed for each of the measure scores on the RM Words test. Therefore, cumulative percentiles are reported for all three trial measures (Trial 1, Trial 2, and Delay Trial - see Table 19) and for the total acquisition scores (see Table 20).

Reliability

As explained in the reliability section for the RM Pictures test, inter-rater reliability, alternate form reliability, and test-retest reliability are inappropriate for examining tests of learning and recognition memory. However, internal consistency of RM Words test measures was examined on a measure-by-measure basis using Cronbach's Coefficient Alpha. Inter-item correlation for the three Trial measures are as follows: $\alpha = 0.84$ (Trial 1), $\alpha = 0.80$ (Trial 2), and $\alpha = 0.87$ (Delay Trial). Inter-item correlation for the Total Acquisition measure was 0.89.

Although test items were not chosen specifically to reflect a similar level of difficulty, they were all concrete nouns. To examine difficulty levels of the target items, the percentage of correct and incorrect responses for each item was calculated by the method described in the RM Pictures reliability section. Mean error rates for individual target items ranged from 9.4% to 51.8% wrong on the first trial, 0.0% to 16.8% wrong on the second trial, and 1.2% to 19.8% wrong on the delay trial (see Table 21).

- - - - -
Insert Table 21 about here
- - - - -

Table 21

Rank Order of Item Difficulty for Recognition Memory for Words Test

ITEM	TRIAL 1			TRIAL 2			DELAY TRIAL		
	RANK ORDER	'EASIER' / 'HARDER'	PERCENT WRONG	RANK ORDER	'EASIER' / 'HARDER'	PERCENT WRONG	RANK ORDER	'EASIER' / 'HARDER'	PERCENT WRONG
calculator	1	Easier	9.4	4	Easier	1.2	3.5	Easier	1.8
brain	2	Easier	9.5	11	Easier	2.4	11.5	Easier	3.0
fossil	3	Easier	11.2	19.5	Easier	4.1	19.5	Easier	4.2
doughnut	4	Easier	11.4	13.5	Easier	3.0	11.5	Easier	3.0
harmonica	5	Easier	11.5	28	Harder	5.6	29.5	Harder	6.2
pizza	6	Easier	11.7	5.5	Easier	1.5	1.5	Easier	1.2
raft	7	Easier	13.8	35.5	Harder	7.4	33	Harder	6.8
flea	8	Easier	15.0	32	Harder	6.5	34	Harder	7.0
camera	9	Easier	15.2	37	Harder	7.6	19.5	Easier	4.2
blouse	10.5	Easier	15.3	47.5	Harder	16.8	6	Easier	2.1
valley	10.5	Easier	15.3	7.5	Easier	1.8	19.5	Easier	4.2
alligator	12	Easier	15.9	5.5	Easier	1.5	11.5	Easier	3.0

Table 21 (cont'd)

ITEM	TRIAL 1		TRIAL 2		DELAY TRIAL	
	RANK ORDER	'EASIER' / 'HARDER'	PERCENT WRONG	RANK ORDER	'EASIER' / 'HARDER'	PERCENT WRONG
trapeze	13	Easier	16.1	24	Easier	5.0
lobster	14	Easier	16.2	11	Easier	2.4
walnut	15	Easier	16.4	13.5	Easier	3.0
mattress	16	Easier	17.4	18	Easier	3.9
elevator	17	Easier	17.6	17	Easier	3.5
beach	18	Easier	17.7	35.5	Harder	7.4
dancer	19.5	Easier	18.9	42	Harder	9.7
saddle	19.5	Easier	18.9	25.5	Harder	5.1
garage	21	Easier	19.3	2.5	Easier	0.9
helmet	23	Easier	20.0	15.5	Easier	3.3
magnet	23	Easier	20.0	33.5	Harder	7.3
village	23	Easier	20.0	19.5	Easier	4.1
				25.5	Harder	5.1
				8	Easier	2.4
				29.5	Harder	6.2
				14.5	Easier	3.3
				6	Easier	2.1
				24	Easier	5.0
				17	Easier	4.1
				16	Easier	3.6
				1.5	Easier	1.2
				31	Harder	6.6
				27	Harder	5.9
				22	Easier	4.5

Table 21 (cont'd)

ITEM	TRIAL 1		TRIAL 2		DELAY TRIAL	
	RANK ORDER	'EASIER' / 'HARDER' PERCENT WRONG	RANK ORDER	'EASIER' / 'HARDER' PERCENT WRONG	RANK ORDER	'EASIER' / 'HARDER' PERCENT WRONG
ambulance	25	Harder 20.3	38	Harder 7.7	6	Easier 2.1
fort	27	Harder 20.9	47.5	Harder 16.8	35.5	Harder 7.1
hose	27	Harder 20.9	27	Harder 5.4	35.5	Harder 7.1
orchard	27	Harder 20.9	41	Harder 8.6	14.5	Easier 3.3
factory	29	Harder 21.0	11	Easier 2.4	39	Harder 8.5
calendar	30	Harder 22.9	2.5	Easier 0.9	3.5	Easier 1.8
pancake	31	Harder 23.6	22.5	Easier 4.8	11.5	Easier 3.0
burlap	32	Harder 24.0	22.5	Easier 4.8	19.5	Easier 4.2
shield	33	Harder 26.2	31	Harder 6.2	45	Harder 11.8
rock	34	Harder 26.4	15.5	Easier 3.3	40	Harder 8.5
flour	35	Harder 26.7	40	Harder 8.2	38	Harder 7.7
chimney	36	Harder 27.4	1	Easier 0	28	Harder 6.1

Table 21 (cont'd)

ITEM	TRIAL 1		TRIAL 2		DELAY TRIAL	
	RANK ORDER	'EASIER' / 'HARDER'	PERCENT WRONG	RANK ORDER	'EASIER' / 'HARDER'	PERCENT WRONG
forest	37	Harder	29.5	7.5	Easier	1.8
bacon	38	Harder	30.0	9	Easier	2.1
monkey	39.5	Harder	31.5	30	Harder	5.9
swan	39.5	Harder	31.5	21	Easier	4.4
graph	41	Harder	32.1	44	Harder	11.7
tire	42	Harder	33.8	43	Harder	10.6
ribbon	43	Harder	34.0	45	Harder	13.3
grass	44	Harder	37.1	25.5	Harder	5.1
spider	45	Harder	38.4	33.5	Harder	7.3
dollar	46	Harder	39.1	29	Harder	5.7
muzzle	47	Harder	41.2	39	Harder	7.9
tooth	48	Harder	51.8	46	Harder	16.1
				32	Harder	6.7
				9	Easier	2.7
				47	Harder	13.8
				40	Harder	8.5
				43.5	Harder	10.0
				42	Harder	9.1
				48	Harder	19.8
				25.5	Harder	5.1
				43.5	Harder	10.0
				23	Easier	4.7
				37	Harder	7.4
				46	Harder	13.6

Items were divided into 'easier' and 'harder' groups, following the procedure previously described. For the RM Words test, the division between 'easier' and 'harder' items fell between 20% and 20.2% wrong on Trial 1, between 4.7% and 4.9% on Trial 2, and between 5.1% and 5.3% on the Delay Trial. While 12 of the items stayed in the 'easier' half and 10 items remained in the 'harder' category across all three learning trials, the majority of items moved back and forth across the two halves. In addition, the difference in percentage points between 'easier' and 'harder' items was limited due to a ceiling effect after the first trial; that is, over 50% of the sample correctly identified 47 or 48 of the 48 words.

Finally, the location on the response page of incorrectly chosen items was examined to determine if a location bias existed in responses (see Table 22). This table identifies the location of all incorrectly chosen items across the three trials for the RM Words test. It can be seen that when the location of stimulus item response errors are summed across all three trials, ranking order of the location choice differed among trials.⁸ It can be concluded, therefore, that the location of the target item on the response pages did not significantly affect item choice.

 Insert Table 22 about here

⁸ For a more complete description of item location analysis, see Appendix H.02.

Table 22

Analysis of Item Responses by Location on Response Trial Page for Recognition Memory for Words (n = 340).

LOCATION ON PAGE ¹	TRIAL 1	TRIAL 2	DELAY TRIAL	TOTAL WRONG
1 Number incorrect <i>Percent of Trial</i>	871 (3) ² 23.6%	170 (4) 19.0%	184 (4) 19.7%	1225 20.8%
2 Number incorrect <i>Percent of Trial</i>	700 (4) 19.0%	207 (3) 23.1%	268 (1) 28.7%	1175 23.6%
3 Number incorrect <i>Percent of Trial</i>	970 (2) 26.3%	268 (1) 29.9%	215 (3) 23.0%	1453 26.4%
4 Number incorrect <i>Percent of Trial</i>	1143 (1) 31.0%	251 (2) 28.0%	267 (2) 28.0%	1661 29.2%
Total incorrect	3684	896	934	5514

¹ Item in the top left corner is indicated as #1; top right, #2; bottom left, #3; bottom right, #4.

² Numbers in parentheses indicate ordinal sequence from most difficult (1) to least difficult (2) for that location of the item on the page.

Standard Error of Measurement. Following the procedure described in RM Pictures, SEM is reported as a measure of internal consistency for each of the four measures evaluated. (see Table 23).

- - - - -
 Insert Table 23 about here
 - - - - -

Using the SEM of 1.62 for Trial 2 provided in this table, it can be concluded with 68% confidence that with an observed score of 43, this person's true score falls between 41.38 and 43.62, with 95% confidence, this score will fall between 39.76 and 46.24, and with 99% confidence, between 38.82 and 47.80. Because of the variability in performance for each measure, it would be difficult to predict an individual's score. However, it allows for evaluation of whether or not the individual can learn and recognize new information.

Correlated Measures

The relationships of RM Words test scores with various other cognitive measures was examined to determine convergent and discriminant support. A subset of 292 participants who completed all of the neuropsychological measures and 328 participants completing all G-LAMB measures were used.

Relationship with other memory measures. In order to examine concurrent support, the relationship was examined between scores on the RM Words and comparable scores on the Geriatric Learning and Memory Battery. For each test, measures reviewed

Table 23

Variability (SEM) Within Age Groups for Each Recognition Memory for Words Trial and for Total Acquisition

RMW MEASURES	STANDARD DEVIATION OF SCALE SCORES	ALPHA LEVEL	<u>SEM</u>
TRIAL 1	6.89	0.84	2.77
TRIAL 2	3.61	0.80	1.62
TOTAL ACQUISITION	9.66	0.89	3.26
DELAY TRIAL	3.79	0.87	1.39

Note: SEM = Standard Error of Measurement; calculated by multiplying the standard deviation of the mean raw scale scores times the square root of 1 minus the Cronbach's alpha level for that scale (from inter-item correlations)

were the first and last learning trials, the total acquisition score, and the delayed memory trial. Pearson correlations between the RM Words measures and the G-LAMB Paragraph scores can be found in Tables 24, with the G-LAMB Word List scores in Table 25, and with the G-LAMB Rectangular Figure scores in Table 26. As can be seen in these tables, the only significant correlation appeared between the initial learning trial of the G-LAMB Word List and Trial 1 of the RM Words test. This resulted in a positive correlation between the two tests on the total amount learned across both learning trials. It was noted that both of these correlations were relatively low, with little variance appearing on any of the RM Words measures. Similar to the RM Words test, the G-LAMB Word List involves the learning and memory of verbal stimuli in form of individual unrelated words. No correlations were found between RM Words measures and the Paragraph test, or between the RM Words measures and the Rectangular Figure test.

 Insert Tables 24, 25 & 26 about here

Relationship with other neuropsychological factors. As in the RM Pictures test, RM Words scores were examined in relation to results from the set of neuropsychological tests designed to reflect other cognitive abilities. These included tests involving word fluency, confrontational naming, visual perception, and visual tracking and attention.

Pearson correlations between RM Words measures and the neuropsychological measures are shown in Table 27. The only significant correlation found was between the RM Words delayed memory trial and the measure of word fluency, which required the

Table 24

Pearson Correlations Between Comparable Recognition Memory for Words Measures and G-LAMB Paragraph Measures (n = 298)

RMW MEASURES	G-LAMB WORD LIST			
	TRIAL 1	TRIAL 5	TOTACQ	DELAY TRIAL
TRIAL 1	0.08	0.03	0.06	0.03
TRIAL 2	0.09	0.06	0.18	0.06
TOTACQ	0.10	0.04	0.08	0.05
DELAY TRIAL	0.11	0.04	0.09	0.04

* $p < .05$; ** $p < .01$; *** $p < .001$

Note: Trial 1 is the first learning trial from each test; the second measure (Trial 'x') represents the score on the last learning trial of each test; TotAcq represents a sum of the scores of all learning trials for each test.

Table 25

Pearson Correlations Between Comparable Recognition Memory for Words Measures and G-LAMB Word List Measures (n = 298)

RMW MEASURES	G-LAMB WORD LIST			
	TRIAL 1	TRIAL 5	TOTACQ	DELAY TRIAL
TRIAL 1	0.14*	0.03	0.03	0.04
TRIAL 2	0.08	0.00	-0.03	0.08
TOTACQ	0.13*	0.02	-0.01	0.06
DELAY TRIAL	0.08	-0.03	-0.04	0.07

* $p < .05$; ** $p < .01$; *** $p < .001$

Note: Trial 1 is the first learning trial from each test; the second measure (Trial 'x') represents the score on the last learning trial of each test; TotAcq represents a sum of the scores of all learning trials for each test.

Table 26

Pearson Correlations Between Comparable Recognition Memory for Words Measures and G-LAMB Rectangular Figure Measures (n = 298)

RMW MEASURES	G-LAMB RECTANGULAR FIGURE			
	TRIAL 1	TRIAL 3	TOTACQ	DELAY TRIAL
TRIAL 1	0.03	0.00	0.02	-0.02
TRIAL 2	0.06	0.01	0.05	-0.03
TOTACQ	0.04	0.00	0.03	-0.03
DELAY TRIAL	0.03	-0.01	0.02	-0.03

* $p < .05$; ** $p < .01$; *** $p < .001$

Note. Trial 1 is the first learning trial from each test; the second measure (Trial 'x') represents the score on the last learning trial of each test; TotAcq represents a sum of the scores of all learning trials for each test.

quick initiation of words beginning with an assigned letter and of words from the given category of 'animals'.

Insert Table 27 about here

Relationship with reported depressive symptomatology. Finally, an ANOVA was undertaken to evaluate the relationship between RM Words measures and scores on the Geriatric Depression Scale (GDS) using the procedure described in the RM Pictures analyses (see page 135).⁹ No significant differences were found among any of these measures, indicating self-reported mood did not affect performance on the RM Words test (see Table 28).

Insert Table 28 about here

Abstract Design Recognition Memory (RM Designs)

A total of 338 individuals from the original 370 volunteers completed all three trials of the Recognition Memory for Designs (RM Designs) test and were used for developing normative tables for this test. Demographic information for the participants used in this portion of the study are presented in Appendix E.03. Similar to the previous two subject groups, participants ranged from ages 16-85 with a comparable number of

⁹ The same matched pairs of participants were used for all three Recognition Memory Battery tests. See Appendix 1.01 for a description of this sample.

Table 27

Pearson Correlations Between Recognition Memory for Words Measures and Scores on Neuropsychological Factors (n = 228)

RMW MEASURES	NEUROPSYCHOLOGICAL MEASURES				
	WORD FLUENCY	TRAILS A	TRAILS B	BOSTON NAMING TEST	G-LAMB FIGURE COPY
TRIAL 1	0.09	0.01	-0.02	-0.02	0.00
TRIAL 2	0.13	-0.10	-0.12	-0.06	-0.06
TOTACQ	0.11	-0.03	-0.06	0.01	-0.02
DELAY TRIAL	0.18**	-0.06	-0.11	0.09	-0.05

* $p < .05$; ** $p < .01$; *** $p < .001$

Note: RMW measures include Trials 1 and 2 (learning trials); TotAcq (sum of the scores of both learning trials); Word Fluency represents an additive score of words identified on F-A-S and Animals; Trails A and Trails B are measured in seconds to completion of the task; BNT (Boston Naming Test) is represented by the total number of items correctly named without cues; and G-LAMB Figure Copy is the score on the Copy Trial for the Rectangular Figure in the G-LAMB.

Table 28

Mean Scores comparing Recognition Memory for Words Measures for 'High' and 'Low' Depression Groups

RMW MEASURES	'Low' Depression (GDS Score below 14)	'High' Depression (GDS Score 14 and above)
TRIAL 1 <i>Mean</i> <i>(s.d.)</i>	36.00 <i>(5.72)</i>	39.67 <i>(5.72)</i>
TRIAL 2 <i>Mean</i> <i>(s.d.)</i>	44.79 <i>(3.82)</i>	46.00 <i>(4.97)</i>
TOTAL ACQUISITION <i>Mean</i> <i>(s.d.)</i>	80.79 <i>(8.76)</i>	85.67 <i>(11.35)</i>
DELAY TRIAL <i>Mean</i> <i>(s.d.)</i>	44.67 <i>(4.41)</i>	46.46 <i>(4.28)</i>

* $p < .05$; ** $p < .01$; *** $p < .001$

males and females in each group, education levels as well as variability in reported years of education within each age group increased after age 20, vocabulary increased with age, and screening variables were within the cut-off limits.

Pearson correlations among the demographic and screening variables were conducted for the RM Designs test. Results for all correlations (see Appendix F.03) were essentially the same as those seen in Table 4 for the entire sample. This included significant positive correlations between vocabulary and age, and between vocabulary and education. Gender was not correlated with any of the variables.

RMD measure scores for original 10 age groupings. Descriptive details about the mean scores earned by the entire sample for each of the four measures (Trial 1, Trial 2, Delay Trial), and the total number correct across learning trials (Total Acquisition) are found in Table 29. Because of the relative difficulty of the RM Designs materials, only 24 (rather than 48) stimulus items were included in this recognition memory test.

- - - - -
 Insert Table 29 about here
 - - - - -

Overall, it can be seen that scores improved with additional exposure. In addition, mean scores on each of the four RM Designs measures peaked between the ages of 20 and 40, and the oldest two age groups earned the lowest scores. This was borne out with an Age by Trials repeated measures ANOVA, which revealed a significant Age effect ($F [9,328] = 12.65, p < .001$), a significant Trials effect ($F [2,656] = 491.44, p < .001$), and a significant Age by Trials interaction ($F [2, 18] = 1.80, p < .05$).

Table 29

Scores by Age Groups on Recognition Memory for Designs Measures

AGE GROUPS	Age Groups										
	16-20 (n=28)	21-30 (n=30)	31-40 (n=30)	41-50 (n=48)	51-60 (n=49)	61-65 (n=34)	66-70 (n=31)	71-75 (n=30)	76-80 (n=30)	81-85 (n=28)	ALL (n=338)
Trial 1											
Mean	17.46	17.73	17.93	18.27	17.53	17.29	16.61	15.83	14.67	13.18	16.81
(s.d.)	3.51	3.36	4.14	3.23	3.90	3.46	3.75	3.49	4.18	3.69	3.92
Trial 2											
Mean	22.93	23.00	22.60	22.29	21.29	21.29	19.61	20.34	18.70	16.86	21.00
(s.d.)	1.46	1.23	1.75	2.16	3.37	2.50	4.76	3.06	3.85	4.59	3.54
Total											
Acquisition											
Mean	40.39	40.73	40.53	40.56	38.82	38.59	36.22	36.13	33.37	30.04	37.81
(s.d.)	4.56	4.23	5.45	4.69	6.90	5.21	8.04	5.96	7.53	7.59	6.86
Delay Trial											
Mean	22.82	22.70	22.53	22.00	20.88	20.82	19.45	19.53	18.07	16.32	20.62
(s.d.)	1.36	1.59	1.99	2.13	3.61	2.67	4.62	3.87	3.99	4.80	3.75

Post hoc analyses indicated that the Age effect was largely due on Trial 1 to significantly higher scores in the younger seven age groups (ages 16-79) compared with the older two age groups (76-81), and that differences in scores on the other two trials reflected higher scores in the younger five compared to the older five age groups. The Trials effect indicated an overall significant increase in scores from Trial 1 to Trial 2, while scores on the Delay Trial remained essentially the same within each age group. Finally, the interaction was due to a greater improvement in scores from Trial 1 to Trial 2 for those in the younger two age groups (16-29) compared with the older age groups.

Norming the Recognition Memory for Designs (RM Designs) Test

Results from this sample were used to examine the nature of the normative information for the RM Designs test. The same procedure was followed as noted in the previous two tests, including conducting regression analyses to determine appropriate stratifying variables, determining the composition of the normative groups, and constructing norms.

Regression analyses. Stepwise regression analyses were performed to determine appropriate stratification of the normative data. As before, age, years of education, and gender were entered in a stepwise fashion to determine which variables might predict performance on each measure. The results of these regression analyses are shown in Table 30. For each measure, significant age differences accounted for the greatest amount of variance. For the two learning trials, as well as for the total acquisition of items, education levels also predicted a significant, but small amount of the variance. On the Delay Trial, only age predicted a significant portion of the variance. Gender did not contribute to the

variance significantly on any measure. These findings were also supported by the semi-partial correlation (r_{sp}) results reported in this table. Darlington (1990) recommends the use of semi-partial correlations as an index of the relative importance of variables in a regression. The larger the r_{sp} for a variable, the greater the relative importance of that variable.

- - - - -
 Insert Table 30 about here
 - - - - -

After examining these analyses, age alone was chosen as the single stratifying variable for three reasons: (1) because of the three variables examined, it explained the largest amount of the variance; (2) the contribution of education was relatively small when the influence of age was removed, and (3) splitting the sample into too many subgroups (such as age and education) would make the number of participants within each group unacceptably low.

Composition of normative groups. The next step in developing normative information for interpretation of test results was to determine whether norms are best reported for each of the ten age groups, or if the data could be better represented using fewer age groups. Guided by statistical considerations, as well as clinical application, several criteria were considered in making this decision. First, age effects within the selected age groups should be minimized. That is, the majority of participants within an age group should be well-represented by the normative data reported for that group. Second, there should be no meaningful differences between age groups on the demographic variables. This reduces the possibility that differences in performance on the

Table 30

Summary of Stepwise Regressions for Recognition Memory for Designs Measures

RMD MEASURES	R^2	F	Step	R^2_{ch}	r_{sp}
Trial 1	.119	22.719 ¹			
Age			1	.080	.283
Education			2	.039	.198
Gender			n.s. ²		
Trial 2	.233	50.842			
Age			1	.209	.458
Education			2	.024	.153
Gender			n.s. ²		
Total Acquisition	.195	40.537			
Age			1	.158	.397
Education			2	.037	.192
Gender			n.s. ²		
Delay Trial	.036	12.567			
Age			1	.036	.190
Education			2		
Gender			n.s. ²		

Note: Education is reported in number of years; Trial 1 and Trial 2 are the two learning trials from the Recognition Memory for Designs test, while Total Acquisition represents a sum of the scores of these two trials, and Delay Trial is the delayed memory trial.

¹ All F scores in this table were at a .001 level of significance.

² 'n.s.' represents a non-significant contribution to the regression equation.

measures might be due to demographic factors other than age. Third, the sample sizes within the age groups should consist of at least 50 participants to ensure an adequate representation of that age group. Fourth, differences in performance should be apparent between the age groups. It should be noted, however, that although this consideration is desirable, it should not take precedence over minimizing age effects within an age group. Finally, the same age stratification should be used for all four measures. As presented in the previous set of analyses, age contributed significantly (albeit to various degrees) to the variance for each test measure. A comparable set of age-grouped norms for each test measure reduces the possibility that a clinician will use the wrong norms when interpreting one or the other measure on the RM Designs test.

Visual inspection of the mean scores for the RM Designs measures for the original 10 age groups suggested that if the scores were to be stratified by age, collapsing the data into fewer age groups would be beneficial because performance on the four measures was highly similar for some of the original ten age groups. This inspection of the data also suggested the most appropriate way to collapse the data was to split it into three groups: 16-50, 51-75, and 76-85. The following information supports this type of division: first, total acquisition scores begin declining at age 50; and second, although Trial 1 and Trial 2 scores decline again at age 65, this only represents a one or two item difference, on the average, whereas a larger reduction in scores on the two learning trials begins to show at age 75. A description of the demographic information for this age grouping may be found in Table 31.

 Insert Table 31 about here

The first criterion mentioned above (that normative data at each age group should well-represent the data reported for that group) was evaluated by examining the distribution of age within each age group. Although the sizes of the three age groups were discrepant (ranging from 144 to 58 participants), participants were represented by the full age spectrum within each age group, with only two individual ages remaining unrepresented in the sample (29 and 36 years).

An ANOVA compared the means of the demographic variables across the three normative age groups in order to evaluate the second criterion mentioned above. This revealed that statistical differences were not found among the three groups in the number of years of education ($F [2,335] = 2.01, p > .05$). In the general population, reported levels of education decrease as age increases. However, in this study, the number of years of education was considered when recruiting participants (i.e., a high school diploma was used as a cut-off between the two groups), resulting in relatively equivalent years of education for each age group (see Appendix E.03). Statistical differences were found among the three age groups in vocabulary levels but *post hoc* analysis indicated this was likely due to a difference between the youngest and the two older age groups ($F [2,335] = 12.63, p < .001$). These findings reflect differences recorded in the literature (e.g., Horn, 1982; Wechsler, 1981) which report that vocabulary ability increases with age as well as the fact that educational levels were foreshortened in the younger two age groups in this

Table 31

Demographic Information for Final Three Age Groups for Recognition Memory for Designs Test

DEMOGRAPHICS:	AGE GROUPS (number of participants)			
	16-50 (n=136)	51-75 (n=144)	76-80 (n=58)	ALL (n=338)
Age				
Mean	32.78	64.04	80.22	54.24
(s.d.)	(1.42)	(2.38)	(1.27)	(20.30)
Education				
Mean	13.01	12.70	12.29	12.76
(s.d.)	(2.14)	(2.48)	(2.39)	(2.34)
Vocab				
Mean	49.55	54.92	55.03	52.78
(s.d.)	(10.84)	(8.91)	(8.66)	(10.02)
MMSE				
Mean	28.92	28.60	28.07	28.64
(s.d.)	(1.09)	(1.19)	(1.32)	(1.21)
GDS				
Mean	5.34	3.63	4.40	4.45
(s.d.)	(3.53)	(3.33)	(3.35)	(3.49)
Gender				
Males	12 (42.9)	22 (40.7)	16 (48.5)	170 (45.9)
Females	16 (57.1)	32 (59.3)	17 (51.5)	120 (54.1)

Note. Education is reported in number of years; Vocab = raw scores from Wechsler Adult Intelligence Scale-Revised Vocabulary subtest (maximum score = 70); MMSE = Mini-Mental State Examination (maximum score = 30); GDS = Geriatric Depression Scale (maximum score = 30); Gender is reported by the number of participants as well as the percent.

study because of lessened opportunity. This grouping (ages 16-50, 51-75, 76-85), therefore, meets the second and third criteria set above for non-meaningful differences on the demographic variables and for sample size.

Performance on the Abstract Recognition Memory test over trials. Table 32 shows mean scores on the four measures examined for the RM Designs for each of the three normative age groupings. An Age by Trials repeated measures ANOVA was conducted and revealed a significant between-groups Age effect ($F [9,328] = 12.65, p < .001$) and a significant within-subjects Trial effect ($F [2,656] = 491.44, p < .001$). The Age by Trials interaction was also significant ($F [2,18] = 1.80, p < .05$), indicating that the rate of learning over the trials differed for the three age groups.

- - - - -
 Insert Table 32 about here
 - - - - -

Post hoc analysis indicated that on Trial 1, the two younger age groups performed at statistically significant higher levels than the 76-85 year age group. On Trial 2, all three groups differed from each other, with mean scores inversely correlated with age. Although mean scores on the Delay Trial were similar to Trial 2 scores within each age group, overall the Delay Trial scores again decreased with age. Finally, an examination of the data in Table 32 indicates that the significant interaction appears to be due to the greater improvement in performance by the youngest age group on Trial 2.

Normative data. Normative data for the RM Designs are reported separately for each of the three age groups: 16-50, 51-75, and 76-85. Because of the truncated scores

Table 32

Performance on Recognition Memory for Designs for Three Age Groups

RM DESIGNS MEASURES	AGE GROUPS (Number of Participants)			
	16-50 (n=136)	51-75 (n=144)	76-80 (n=58)	ALL (n=338)
Trial 1				
<i>Mean</i>	17.91	16.92	13.95	16.81
<i>(s.d.)</i>	(3.51)	(3.70)	(3.99)	(3.92)
Trial 2				
<i>Mean</i>	22.65	20.72	17.81	21.00
<i>(s.d.)</i>	(1.77)	(3.52)	(4.29)	(3.54)
Total Acquisition				
<i>Mean</i>	40.56	37.65	31.76	37.81
<i>(s.d.)</i>	(4.70)	(6.62)	(7.68)	(6.86)
Delay Trial				
<i>Mean</i>	22.44	20.28	17.22	20.62
<i>(s.d.)</i>	(1.86)	(3.74)	(4.45)	(3.75)

Note. Maximum score on all Abstract Recognition Memory trials is 24. Maximum score on Total Acquisition is 48.

and ceiling effect in a normal population for this test of learning and memory, cumulative percentiles are reported for all three trial measures (Trial 1, Trial 2, and Delay Trial - see Table 33) and for the total acquisition scores (see Table 34) scores.

 Insert Tables 33 and 34 about here

Reliability

As with RM Pictures, internal consistency of test items on the RM Designs test was examined on a measure-by-measure basis using Cronbach's Coefficient alpha. The inter-item correlation for the RM Designs trials were identified as $\alpha = 0.73$, 0.78 and 0.76 for Trial 1, Trial 2 and the Delay Trial, respectively. The inter-item correlation for the Total Acquisition measure was 0.84.

Items for the RM Designs test were not specifically designed to have a similar level of difficulty; they were all abstract line drawings of the same general proportions, with moderate complexity. Each distractor item has a similar shape and style to the corresponding target item. It is possible, however, that some target items were more difficult to identify than others. To determine whether item difficulty affected RM Designs test scores, further analysis of items was undertaken. First, mean error rates for individual target items were calculated. They ranged from 10.6% to 56.5% incorrect on the first trial, 4.5% to 27.2% incorrect on the second trial, and 2.1% to 35.6% incorrect on the delay trial (see Table 35).

 Insert Table 35 about here

Table 33

Cumulative Percentiles for Recognition Memory for Designs Learning and Delay Trials
for Three Age Groups.

Raw Scores	CUMULATIVE PERCENTILES								
	AGES 16-50			AGES 51-75			AGES 76-85		
	Trial 1	Trial 2	Delay	Trial 1	Trial 2	Delay	Trial 1	Trial 2	Delay
24	100	100	100	100	100	100	100	100	100
23	97	58	60	99	83	85	100	93	97
22	90	34	40	97	62	65	100	86	91
21	85	21	23	91	44	51	100	78	81
20	73	10	15	83	33	40	97	72	71
19	64	4	9	70	24	28	86	62	65
18	55	3	4	60	18	23	85	53	57
17	41	1	2	51	14	19	79	41	45
16	31	1	1	41	11	15	72	33	40
15	24	1	<1	31	10	13	67	26	35
14	19	1		28	7	9	52	19	24
13	13	1		19	6	8	47	14	19
12	8	<1		15	5	6	36	12	17
11	5			10	3	3	29	9	10
10	1			6	2	3	24	7	9
9	1			2	1	1	15	5	7
8	<1			1	1	1	9	4	5
7				<1	<1	1	5	3	4
6						1	2	3	3
5						<1			2

Table 34

Cumulative Percentiles for Recognition Memory for Designs Total Acquisition Scores for Three Age Groups

Raw Scores	CUMULATIVE PERCENTILES		
	Ages 16-50	Ages 51-75	Ages 76-85
48	100	100	100
47	97	99	100
46	92	98	100
45	87	94	100
44	76	90	98
43	69	83	93
42	63	76	91
41	50	67	88
40	43	59	87
39	37	52	86
38	31	43	81
37	26	38	78
36	19	34	69
35	15	31	64
34	13	27	60
33	8	21	55
32	7	19	51
31	4	17	47
30	3	15	43

Table 34 (con't)

CUMULATIVE PERCENTILES			
Raw Scores	Ages 16-50	Ages 51-75	Ages 76-85
29	2	12	40
28	1	10	38
27	<1	9	29
26	<1	8	26
25	<1	8	21
24	<1	7	17
23	<1	6	15
22	<1	4	13
21	<1	3	11
20	<1	3	9
19	<1	2	7
18	<1	2	6
17	<1	1	5
16	<1	1	3
15	<1	<1	3
14	<1	<1	2
13	<1	<1	2
12	<1	<1	2
11	<1	<1	<1

Table 35

Rank Order of Item Difficulty for Recognition Memory for Designs Test

ITEM	TRIAL 1		TRIAL 2		DELAY TRIAL		
	RANK ORDER	'EASIER'/'HARDER'	RANK ORDER	'EASIER'/'HARDER'	RANK ORDER	'EASIER'/'HARDER'	PERCENT WRONG
1	easier	10	easier	15	harder	12.7
2	easier	3	easier	5	easier	7.4
3	easier	8	easier	2	easier	8.4
4	easier	16	harder	11	easier	9.5
5	easier	2	easier	7	easier	8.3
6	easier	17	harder	2	easier	3.6

Table 35 (cont'd.)

ITEM	TRIAL 1		TRIAL 2		DELAY TRIAL	
	RANK ORDER	'EASIER' / 'HARDER'	RANK ORDER	'EASIER' / 'HARDER'	RANK ORDER	'EASIER' / 'HARDER'
7	easier	21.9	4 easier	9.5	easier
8	easier	24.5	15 harder	6	easier
9	easier	24.8	10 easier	14	harder
10	easier	25.5	1 easier	3	easier
11	easier	28.1	5 easier	12.5	easier
12	easier	19.9	24 harder	23	harder

Table 35 (cont'd.)

ITEM	TRIAL 1		TRIAL 2		DELAY TRIAL	
	RANK ORDER	'EASIER/ 'HARDER'	PERCENT WRONG	RANK ORDER	'EASIER/ 'HARDER'	PERCENT WRONG
13	harder	30.5	14	harder	10.3
14	harder	34.6	17	harder	8.9
15	harder	34.9	18	harder	16.0
16	harder	35.5	19	harder	16.6
17	harder	38.2	23	harder	24.0
18	harder	39.9	21	harder	19.2
				16	harder	13.3
				18	harder	16.3
				17	harder	15.3
				22	harder	27.5
				9.5	easier	9.2
				4	easier	4.5

Items were then identified as 'easier' and 'harder' through a procedure described in the RM Pictures section on reliability. The resultant split on the RM Designs measures fell between 24.5% and 25.5% wrong on Trial 1, between 9.5% and 9.8% on Trial 2, and between 10.4% and 11.3% on the Delay Trial.¹⁰ It can be seen in Table 35 that many of the items shifted between the 'easier' and 'harder' groups from trial to trial indicating that correct and incorrect responses were more likely to be due to variables other than inherent difficulty of the individual target stimulus.

In addition, the location on the response page of incorrectly chosen items was examined to determine if a location bias existed in responses (see Table 36). Similar to the RM Pictures and RM Words tests, it can be seen that when item response location was summed across all three trials, ranking order differed among trials.¹¹

 Insert Table 36 about here

As with the previous two tests, inter-rater reliability, alternate form reliability, and test-retest reliability were inappropriate for examining the RM Designs

Standard Error of Measurement. As a measure of internal consistency, the SEM is reported for each of the four measures evaluated. (see Table 37). Using the SEM of 1.81 for Trial 2 provided in Table 37, it can be concluded with 68% confidence that with an

¹⁰ On the delay trial, item numbers 12 and 13 shared a rank order of 12, resulting in the half-way mark (24 easy/24 hard) falling within this rank. Rather than arbitrarily choosing an 'easier' and 'harder' item from this rank, both items at this rank (12) were included in the 'easier' group.

¹¹ For a more complete description of item location analysis, see Appendix H3.

Table 36

Error Rate for Recognition Memory for Designs Item Responses by Location on Response Trial Page

LOCATION ON PAGE ¹	TRIAL 1	TRIAL 2	DELAY TRIAL	TOTAL WRONG
1 Number incorrect <i>Percent of Trial</i>	270 (3) ² 24.0%	91 (4) 19.7%	134 (2) 25.0%	495 23.3%
2 Number incorrect <i>Percent of Trial</i>	284 (2) 25.3%	98 (3) 21.2%	131 (3) 24.4%	513 24.2%
3 Number incorrect <i>Percent of Trial</i>	317 (1) 28.2%	107 (2) 23.2%	144 (1) 26.8%	568 26.8%
4 Number incorrect <i>Percent of Trial</i>	252 (4) 22.5%	166 (1) 35.9%	128 (4) 23.8%	546 25.7%
Total incorrect	1123	462	537	2122

¹ Item in the top left corner is indicated as #1; top right, #2; bottom left, #3; bottom right, #4.

² Numbers in parentheses indicate ordinal sequence from most difficult (1) to least difficult (2) for that location of the item on the page.

observed score of 18, this person's true score falls between 16.19 and 19.81, with 95% confidence, this score will fall between 14.38 and 21.62, and with 99% confidence between 13.33 and 22.67.

 Insert Table 37 about here

Correlated Measures

A portion of the 338 participants who completed all three trials of the RM Designs test did not complete the testing battery in its entirety. A subset of 328 participants who did complete all G-LAMB measures and 292 participants who completed all of the neuropsychological measures were used to analyze convergent and discriminant support for the RM Designs test. These same participants were used in evaluating the Pictures and RM Words tests.

Relationship with other memory measures. Performance on this recognition memory test reflected its more challenging nature when compared with high scores earned on the other two Recognition Memory Battery tests. For each of the G-LAMB tests, measures reviewed were the first and last learning trials, the total acquisition score, and the delayed memory trial. Because of the abstract nature of the RM Designs test and the G-LAMB Rectangular Figure subtest, higher correlations were hypothesized for these two tests than were expected between the RM Designs test and the other G-LAMB memory measures, which are verbal in content.

Table 37

Variability (SEM) Within Age Groups for Each Recognition Memory for Designs Trial and for Total Sample

RMD MEASURES:	STANDARD ERROR OF MEASUREMENT SCORES			
	AGES 16-50	AGES 51-75	AGES 76-85	TOTAL SAMPLE
TRIAL 1	1.82	1.93	2.07	2.04
TRIAL 2	1.81	1.50	2.02	1.66
DELAY TRIAL	0.91	1.83	2.18	1.84
TOTAL ACQUISITION	1.88	2.67	3.07	2.75

Note: SEM = Standard Error of Measurement; calculated by multiplying the standard deviation of the mean raw scale scores times the square root of 1 minus the Cronbach's alpha level for that scale (from inter-item correlations)

Pearson correlations between the RM Designs measures and the G-LAMB Paragraph scores can be found in Table 38, with the Word List scores in Table 39, and with the Rectangular Figure scores in Table 40. As can be seen in these tables, RM Designs measures showed the highest correlation with the Rectangular Figure test, with correlations ranging between 0.14 to 0.70 ($p < .01$). Slightly more moderate correlations were found between RM Designs measures and the Paragraph test, which also involves two learning trials. Finally, low to moderate correlations were found among measures of the RM Designs test and Word List Trial 1, and between RM Designs measures and Word List Total Acquisition scores, Delay Memory scores, and Proportion Remembered scores. The fifth, and final, learning trial on the Word list was not correlated with any RM Designs measures.

Insert Tables 38, 39 & 40 about here

Relationship with other neuropsychological factors. Following the procedure used in the RM Pictures and RM Words studies, RM Designs scores were examined in relation to results from the set of neuropsychological tests used to reflect other cognitive abilities. These included a test of word fluency, confrontational naming, visual perception, and visual tracking and attention. If memory tests are to be useful, they should be related only mildly to moderately to scores on these other standardized tools. Since the RM Designs test was developed to measure recognition of visual-spatial material, the scores from this test should correlate higher with scores from visual tasks, such as Trails A & B and the G-

Table 38

Pearson Correlations Between Comparable RM Designs and G-LAMB Paragraph Measures
 (n = 328)

RMD MEASURES	G-LAMB PARAGRAPH			
	TRIAL 1	TRIAL 2	TOTACQ	DELAY TRIAL
TRIAL 1	0.35***	0.32***	0.36***	0.37***
TRIAL 2	0.34***	0.40***	0.39***	0.40***
TOTACQ	0.37***	0.39***	0.40***	0.42***
DELAY TRIAL	0.32***	0.36***	0.36***	0.36***

* $p < .05$; ** $p < .01$; *** $p < .001$

Note: Trial 1 is the first learning trials from the each test; the second measure (Trial x) represents the score on the last learning trial of each test; TotAcq represents a sum of the scores of all learning trials.

Table 39

Pearson Correlations Between Comparable Recognition Memory for Designs and G-LAMB Word List Measures (n = 328)

RMD MEASURES	G-LAMB WORD LIST			
	TRIAL 1	TRIAL 5	TOTACQ	DELAY TRIAL
TRIAL 1	0.32***	0.05	-0.10	0.39***
TRIAL 2	0.33***	0.01	-0.21***	0.49***
TOTACQ	0.35***	0.02	-0.16**	0.47***
DELAY TRIAL	0.32***	-0.08	-0.24***	0.45***

* $p < .05$; ** $p < .01$; *** $p < .001$

Note. Trial 1 is the first learning trials from the each test; the second measure (Trial x) represents the score on the last learning trial of each test; TotAcq represents a sum of the scores of all learning trials.

Table 40

Pearson Correlations Between Comparable Recognition Memory for Designs Measures and G-LAMB Rectangular Figure Measures (n = 328)

RMD MEASURES	G-LAMB RECTANGULAR FIGURE			
	TRIAL 1	TRIAL 3	TOTACQ	DELAY TRIAL
TRIAL 1	0.55***	0.48***	0.59***	0.49***
TRIAL 2	0.59***	0.65***	0.68***	0.65***
TOTACQ	0.62***	0.61***	0.68***	0.62***
DELAY TRIAL	0.60***	0.64***	0.70***	0.63***

* $p < .05$; ** $p < .01$; *** $p < .001$

Note. Trial 1 is the first learning trials from each test; the second measure (Trial x) represents the score on the last learning trial of each test; TotAcq represents a sum of the scores of all learning trials.

LAMB Rectangular Figure copy trial than with more verbal tasks, such as Word Fluency and the Boston Naming Test (BNT). The BNT, however, is of particular interest, since visual stimuli on memory tests, even of abstract designs, are considered to be verbally mediated.

Pearson correlations between RM Designs measures and the neuropsychological measures are shown in Table 41. Significant correlations exist between the RM Designs acquisition and delayed memory measures and each of the neuropsychological measures, including the copy trial of the G-LAMB Simple Figure.

 Insert Table 41 about here

To further examine these relationships, a stepwise regression analysis was undertaken to determine shared variance by entering each of the neuropsychological tests as a separate variable and regressing them, in turn, on each of the RM Designs measures. These results can be seen in Table 42. In total, 20-36% of the variance in RM Designs scores was explained by these neuropsychological measures. The highest amounts of variance, ranging from 14% to 32%, were contributed by scores on Trails B, while the G-Lamb Copy Trial contributed another 3% to 5% to the shared variance, depending on the RM Designs measure examined. The Boston Naming test added another 2-3% to explaining the shared variance, while Word Fluency added less than 1% to variance on the Trial 2 measure. Trails A did not contribute significantly to the shared variance.

 Insert Table 42 about here

Table 42

Percent of Variance Explained Using Recognition Memory for Designs Measures and Scores on Neuropsychological Tests (N = 292)

RMD MEASURES	CONTRIBUTION TO SHARED VARIANCE					TOTAL SHARED VARIANCE
	TRAILS B	G-I AMB FIGURE COPY	BOSTON NAMING TEST	WORD FLUENCY	TRAILS A	
TRIAL 1	14%	5%	3%	---	---	22%
TRIAL 2	28%	5%	2%	<1%	---	36%
TOTACQ	24%	5%	2%	---	---	31%
DELAY TRIAL	32%	3%	<1%	---	---	36%

* $p < .05$; ** $p < .01$; *** $p < .001$

Note. RMD measures include Trials 1 and 2 (learning trials), TotAcq (sum of the scores of both learning trials), and Delay (number correct after 20 min. delay); Word Fluency represents an additive score of words identified on F-A-S and Animals; Trails A and Trails B are measured in seconds to completion of the task; Boston Naming Test is the total number of items named without cues. Empty cells represent non-significant contribution of a variable in this regression analysis.

Table 43

Percent of Variance Explained Using Recognition Memory for Designs Measures and Scores on Neuropsychological Tests, with Age and Vocabulary Levels Controlled
(N = 292)

RMD MEASURES	CONTRIBUTION TO SHARED VARIANCE						TOTAL SHARED VARIANCE
	AGE & VOCAB	TRAILS B	G-LAMB COPY TRIAL	BOSTON NAMING TEST	WORD FLUENCY	TRAILS A	
TRIAL 1	18%	3%	4%	--	--	--	25%
TRIAL 2	33%	7%	4%	--	--	--	44%
TOTACQ	29%	6%	4%	--	--	--	39%
DELAY TRIAL	31%	10%	3%	--	--	--	44%

* $p < .05$; ** $p < .01$; *** $p < .001$

Note. RMD measures include Trials 1 and 2 (learning trials), TotAcq (sum of the scores of both learning trials), and Delay (number correct after 20 min. delay); Word Fluency represents an additive score of words identified on F-A-S and Animals; Trails A and Trails B are measured in seconds to completion of the task; Boston Naming Test is the total number of items named without cues. Empty cells represent non-significant contribution of a variable in this regression analysis.

Each neuropsychological test used in this study has age-rated norms, representing the fact that performance on most of these measures is related to age. In this sample, scores on Trails B were significantly correlated with age ($r = 0.50$, $p < .001$) and BNT with vocabulary scores ($r = 0.45$, $p < .001$). For these reasons, another stepwise regression was undertaken, this time forcing age and vocabulary in on the first block, and allowing the four neuropsychological measures to compete in the second block. The results of this analysis are found in Table 43. It can be seen that age and vocabulary levels contributed approximately three-quarters of the explained variance for each trial, while Trails B and the G-LAMB Copy Trial combined to explain the remaining quarter. With age and vocabulary levels controlled, the amount of variance that the neuropsychological measures shared with scores on the learning trials was approximately one-third of that shared when age and vocabulary were not controlled.

 Insert Table 43 about here

Relationship with reported depressive symptomatology. Finally, the relationship of scores on the RM Designs test and the Geriatric Depression Scale (GDS) was explored. As in the analysis of RM Pictures and RM Word measures, matched pairs of individuals earning scores above and below the GDS cut-off were compared on RM Designs measures.¹² To determine whether depression scores affected memory ability on the RM Designs test, an ANOVA was undertaken for the Trial 1, Trial 2, Total Acquisition score

¹² The same matched pairs of participants were used for all three Recognition Memory Battery tests. See Appendix H for a description of this sample.

Table 43

Percent of Variance Explained Using Recognition Memory for Designs Measures and Scores on Neuropsychological Tests, with Age and Vocabulary Levels Controlled
(*N* = 292)

RMD MEASURES	CONTRIBUTION TO SHARED VARIANCE						TOTAL SHARED VARIANCE
	AGE & VOCAB	TRAILS B	G-LAMB COPY TRIAL	BOSTON NAMING TEST	WORD FLUENCY	TRAILS A	
TRIAL 1	18%	3%	4%	--	---	---	25%
TRIAL 2	33%	7%	4%	--	---	---	44%
TOTACQ	29%	6%	4%	--	---	---	39%
DELAY TRIAL	31%	10%	3%	---	---	---	44%

* $p < .05$; ** $p < .01$; *** $p < .001$

Note. RMD measures include Trials 1 and 2 (learning trials), TotAcq (sum of the scores of both learning trials), and Delay (number correct after 20 min. delay); Word Fluency represents an additive score of words identified on F-A-S and Animals; Trails A and Trails B are measured in seconds to completion of the task; Boston Naming Test is the total number of items named without cues. Empty cells represent non-significant contribution of a variable in this regression analysis.

and the Delay Trial score on the RM Designs test. There were no significant differences between the two groups on these measures, indicating that self-reported mood did not reflect performance on this test (see Table 44)

Insert Table 44 about here

Final Notes on the Three Recognition Memory Battery Tests

While the purpose of this study was not to directly compare the results of the three tests developed, two findings from this study which the tests have in common will be reported in this section.

First, the lack of forgetting during the delay period was notable across all three of the tests (see Figure 18). There was a slight, insignificant reduction on the RM Designs in the number correct on Trial 2 and the Delay Trial, but the average loss was less than one item for any given age group. The slopes representing the loss for each age group are parallel, indicating that the age differences were equivalent. Differences on the RM Pictures and RM Words tests were essentially nonexistent.

Insert Figure 18 about here

The second commonality of note among the three tests was that ceiling levels were reached on all three of the tests (see Tables 45 and 46). It was found that 100% of the

Table 44

Comparison of 'High' and 'Low' Depression Groups on Recognition Memory for Words Measures

RMW MEASURES	'Low' Depression (GDS Score below 14)	'High' Depression (GDS Score 14 and above)
TRIAL 1		
Mean	17.04	18.29
(s.d.)	(3.90)	(3.77)
TRIAL 2		
Mean	22.42	22.54
(s.d.)	(2.04)	(2.21)
TOTAL ACQUISITION		
Mean	39.46	40.83
(s.d.)	(5.51)	(5.64)
DELAY TRIAL		
Mean	22.25	22.13
(s.d.)	(2.85)	(2.38)

* $p < .05$; ** $p < .01$; *** $p < .001$

Figure 18. Rate of Forgetting for Recognition Memory Battery tests.

Note: For RM Pictures and RM Words, 2% equals approximately 1 item; for RM Designs, 4% equals approximately 1 item.

participants of every age correctly identified 80% or more of the RM Picture items, and over 93% from all age groups reached that level on the RM Words. On the RM Designs test, 97% of the youngest age group (ages 16-50) reached the 80% level by the second trial, 82% of the middle group (ages 51-75) and 46% (nearly half) in the oldest age group (ages 76-85) reached that level of proficiency. When that level was reduced to 70% correct for the RM Designs test, 99% of the subjects in the youngest age group, 88% in the middle age group, and 74% in the oldest age group reached that criteria.

- - - - -
Insert Tables 45 & 46 about here
- - - - -

Table 45

Percent Correct on Recognition Memory for Pictures and Recognition Memory for Words by Trials

Raw Scores	Percent Correct					
	RM Pictures			RM Words		
	Trial 1	Trial 2	Delay Trial	Trial 1	Trial 2	Delay Trial
48	30.4	90.1	90.1	3.8	36.2	42.4
47	20.2	6.7	7.6	4.7	16.5	14.1
46	12.3	1.8	1.8	3.8	11.5	9.4
45	11.1	0.9	0.3	5.9	10.3	5.6
44	6.7	0.3	0	5.0	5.6	9.4
43	5.3	0	0	3.8	5.6	3.2
42	3.2	0	0	3.8	2.9	3.5
41	2.9	0	0	4.4	2.6	3.2
40	2.6	0.3	0.30	5.0	2.4	2.4
39	2.3	0	0	4.1	1.2	0.6
38	1.2	0	0	5.3	1.5	0.6
37	0	0	0	6.8	0.9	0.3
36	0	0	0	4.1	0.3	1.2
35	0.6	0	0	5.6	0	0.9
34	0.3	0	0	5.6	0.6	1.2

Table 45 (cont'd)

Raw Scores	Percent Correct					
	RM Pictures			RM Words		
	Trial 1	Trial 2	Delay Trial	Trial 1	Trial 2	Delay Trial
33	0.3	0	0	3.5	0.6	0.6
32	0.3	0	0	2.6	0.9	0.3
31	0	0	0	4.7	0	0
30	0	0	0	3.2	0	0.9
29	0.3	0	0	2.4	0.3	0
28	0	0	0	2.1	0	0
27	0	0	0	2.9	0	0
26	0	0	0	1.5	0	0.3
25	0	0	0	1.2	0	0
24	0	0	0	1.2	0	0
23	0	0	0	1.2	0	0
22	0	0	0	0.3	0	0
21	0	0	0	0.3	0	0
20	0	0	0	0.3	0.3	0
<20	0	0	0	0.9	0	0

Table 46
Percent Correct on Recognition Memory for Designs by Trials for Three Age Groups

Raw Score	Percent Correct								
	Trial 1			Trial 2			Delay Trial		
	16-50	51-75	76-85	16-50	51-75	76-85	16-50	51-75	76-85
24	2.9	0.7	0	41.9	17.4	6.9	39.7	15.3	3.4
23	6.6	2.8	0	24.3	20.8	6.9	20.6	19.4	5.2
22	5.1	5.6	0	12.5	17.4	8.6	16.2	13.9	10.3
21	12.5	8.3	3.4	11.8	11.1	5.2	8.1	11.1	10.3
20	8.8	11.1	10.3	5.1	9.0	10.3	6.6	12.5	5.2
19	8.8	11.1	1.7	1.5	6.3	8.6	4.4	4.9	8.6
18	14.7	9.0	5.1	1.5	4.2	12.1	2.9	4.2	12.1
17	9.6	10.4	6.9	0.7	2.8	8.6	0	3.5	5.2
16	6.6	9.7	5.1	0	1.4	6.9	0.7	2.1	5.2
15	5.1	3.5	15.5	0	2.1	6.9	0.7	4.2	10.3
14	5.9	8.3	5.1	0	1.4	5.2	0	1.4	5.2
13	5.1	4.2	10.3	0	0.7	1.7	0	2.1	1.7
12	2.9	4.9	6.9	0.7	2.1	5.2	0	2.1	6.9
11	3.7	4.9	5.2	0	1.4	0	0	0.7	1.7
<11	1.5	5.6	17.2	0	2.1	6.8	0	2.8	8.5

DISCUSSION

In this dissertation, the need for a clinical memory test for evaluating memory abilities in individuals with known memory impairments was proposed. A review of the pertinent literature identified and described commonly used memory tests as well as evaluating their strengths and weaknesses. It was concluded that one or all of the following problems exist with current memory tests used in clinical settings: (1) They are often too difficult for the considered population, with floor effects interfering with the ability to assess retained information; (2) they are limited to verbal information and/or require constructional ability for non-verbal information; and (3) they often are limited to one learning trial, making it difficult to ascertain whether 'memory deficits' are due to difficulty with encoding or retrieval from long-term memory.

The current research was undertaken to address three difficulties. A battery of visual recognition memory tests was developed which employs verbal and non-verbal stimuli evaluated in a multiple forced-choice format. The tests include two learning trials and a delay trial. A recognition paradigm was used because it is the most sensitive procedure available to ascertain the amount of material accessible in memory. The result was a battery of tests entitled the "Recognition Memory Battery", which included Recognition Memory for Pictures (RM Pictures), Recognition Memory for Words (RM Words), and Recognition Memory for Abstract Designs (RM Designs). Using a community sample, this research project developed appropriate normative tables for each test, evaluated the reliability of each test, and examined the relationship of these tests with other memory tests and neuropsychological measures.

The following results were noted during the analysis of these data: (1) Significant but non-meaningful age effects were present for the RM Pictures test, no age effects were found for the RM Words test, and significant and meaningful age effects were reflected in performance on the RM Designs test; (2) scores improved significantly from Trial 1 to Trial 2 on each of these tests, indicating that learning took place despite truncated scores and a negative skew to the data; (3) little or no information was lost from the last learning trial to the delay trial on any of these tests; (4) ceiling effects were present on each test, especially after the first trial; and (5) the tests represented a range of difficulty, with the lowest set of scores found on the RM Designs test and the highest scores on the RM Pictures test.

Correlations were not found with free recall memory test scores for the RM Pictures and RM Words tests. Either these recognition memory tests are measuring a different construct than those represented by the G-LAMB, or else variability in performance was limited to the point where the results are not interpretable. Performance on each RM Designs measure, on the other hand, was correlated with each measure taken on the G-LAMB tests. This leads to speculation that the RM Designs test might be useful as a screening test for memory deficits when it is not possible to gather that information with currently available clinical tests, but only after validity studies are conducted with the appropriate groups of individuals. It is likely, however, that the RM Pictures and RM Words tests would not be useful for identifying memory deficits. It remains to be seen whether they can validly measure an ability to learn and retain information during the moderate to severe stages of dementia.

Of concern to the validity of the RM Designs test are the moderate correlations found with each of the neuropsychological measures representing verbal and visual ability included in this study. To be effective as a memory test, only mild to moderate correlations should be found with tests measuring other cognitive abilities. If high correlations exist and deficits in performance are found on a memory test, the clinician can not be certain whether that low performance is due to memory deficits or deficits in other cognitive areas.

While this study was not designed to test any specific theoretical hypotheses, three basic issues relating to the findings are explored in this discussion: (1) Age effects, (2) rates of forgetting, and (3) differences among the various types of stimuli. Following this, the relationship of these findings to the issue of recognition memory will be considered, clinical usefulness of the RMB tests will be discussed, and implications for future research will be explored.

Age Effects

One of the most consistent and well-recognized findings in the memory literature is that age-related declines exist in many kinds of memory performance (Craik, Anderson, Kerr & Li, 1995). The majority of these differences, however, occur when cognitive resources are stretched during encoding, and retrieval is assessed through free recall. Free recall relies on self-initiated manipulation and requires the most effort during retrieval. When sufficient support is available during encoding and retrieval processes, these deficits are less pronounced. One area where memory performance is unaffected or only slightly influenced by age is when environmental support is available at the time of retrieval. Such

is the case with a recognition procedure. On the Recognition Memory Battery, recognition memory for pictures and words did not reflect age differences in performance for either learning or memory. Only the RM Designs test reflected clear-cut age decrements in learning and memory performance on each trial.

Direct comparison of these results with previous research is troublesome, as very few studies could be found that comparably examined the range of various types of stimuli over age using recognition memory. As pointed out by McCormack (1984), it is difficult to meaningfully compare types of stimuli across studies, as research designs are usually disparate and often inadequately described. However, three topics from the literature relating to age differences will be discussed when considering the Recognition Memory Battery. These areas concern: (1) The type of stimuli which make up the episodic tests of this memory battery, (2) the theoretical differences between recall and recognition, and (3) various procedural matters relating to the format of the research, such as individual differences, the difficulty of the task, the degree to which the initial acquisition levels are equated, and the type of recognition tests administered.

When the literature is evaluated,¹³ it is seen that visual recognition memory can decline with age. Lower performance with age has been reported using a recognition format across a variety of test designs and for many types of visual stimuli, such as words (Dywan & Jacoby, 1990; Lehman & Mellinger, 1986; Poon & Fozard, 1980; West & Boatwright, 1983; Wickelgren, 1975), pictures or line drawings of everyday objects (Harwood & Naylor, 1969; Howell, 1972; Measso et al., 1990; Mitchell, 1989; Park,

¹³ Since this research involved collecting and evaluating normative information with a community sample, comparisons with previous research will focus on the non-clinical population. It should be noted, however, that the clinical literature would provide another level of support to these findings.

Puglisi & Smith, 1986; Park, Puglisi & Sovacool, 1984), faces (Ferris, 1980; Hulicka, 1967), abstract material such as abstract figures or geometric designs (Arenberg, 1978; Menich & Baron, 1990; Trahan, 1992; Trahan & Larrabee, 1992), and visuo-spatial material (Crook et al., 1990; Lipman, 1991; Rabbitt, 1989; Youngjohn & Crook, 1993).

Recognition memory for pictures of common objects or familiar scenes seems to be least affected by age, as long as the response item is identical to the target item (Bartlett & Leslie, 1986; Park et al., 1986; Till, Bartlett & Doyle, 1982). Recognition memory for pictures or line drawings of abstract designs, on the other hand, reflects significant age differences in performance (Menich & Baron, 1990; Trahan & Larrabee, 1992), as was the case here. Certainly, an abstract recognition task such as the RM Designs test requires the self-initiated, effortful processing noted by Craik et al. (1995) as the conditions least favourable to learning and memory performance in older adults.

Interpretation of results when word stimuli are used, on the other hand, is less straight-forward. It is suspected that the learning of verbal information tends to directly involve semantic processing for future retrieval. Semantic knowledge tends to reflect little change with age (see reviews by Salthouse, 1991; Light, 1991, 1992; Light & Burke, 1988). However, some semantic skills have been shown to decrease with age (Bowles, 1989; Cohen & Faulkner, 1986; Maylor, 1990). McIntyre & Craik (1987) reported greater deficits in memory among older Canadians compared with their younger counterparts when assessed for retrieval of new facts recently incorporated into semantic memory. Because the words presented in the present study could be easily visualized for encoding, as well as verbally processed, this task should have been relatively easy. Age

differences should be reflected, however, as visualizing verbal information in the absence of instructions requires self-initiated encoding, which demands greater effort. But no age differences appeared on the RM Words test. It is likely that in the present research, external support in the form of the use of common concrete nouns and a relatively long exposure time during encoding, as well as a multiple choice format during retrieval, were sufficient to eliminate any age effects in this episodic task.

The second area noted for discussion around age deficits is the differences in performance between free recall and recognition formats. For the most part, age differences are highlighted when a recall procedure is implemented but minimized during recognition procedures. However, recognition procedures accurately reflect the amount of information available in memory by providing environmental support at the time of retrieval. When recognition procedures are used, differences between the amount of newly learned episodic material available to older and younger individuals is not as great as free recall tests would lead one to believe (Craik et al., 1995; Huppert, 1991). Recall procedures provide little environmental support at the time of retrieval, while recognition not only supports retrieval, it bypasses a portion of encoding difficulties by providing the connections. If factors are not amenable to enriched encoding at the time of stimulus presentation (e.g., exposure time is limited or strategy guidance is not given), that material will be difficult for older individuals to access at the time of retrieval. "The art of memory...", William James tells us, "...is the art of thinking...our conscious effort should not be so much to impress or retain (knowledge) as to connect it with something already there." (quoted in Lockhart & Craik, 1990, p.99). It is this connection strategy that

appears to come less automatically as one ages. Free recall, therefore, is sensitive to what cannot be remembered (i.e., either has not been encoded or cannot be retrieved without further cues), and as such, is essential to clinical memory testing. However, there is no more effective way of determining if the information was learned in the first place than with a recognition memory procedure. In general, then, recognition procedures evaluate what an individual is capable of remembering, while recall procedures evaluate what that person is not capable of remembering. Since the Recognition Memory Battery was designed to evaluate retained knowledge, it was important to highlight memory ability, not memory deficits nor age differences.

In general, previous research has concluded that the majority of age differences appear during the encoding process. Moreover, age differences are greatly reduced or non-existent during retrieval when initial learning is equated (e.g., see Hart & O'Shanick, 1993; Park, Puglisi & Sovacool, 1984; Trahan & Larrabee, 1992; Wimer, 1960; Youngjohn & Crook, 1993). Craik et al. (1995) propose that age deficits in working memory may underlie differences in performance on encoding and retrieval tasks. They summarize three approaches to understanding the relationship between working memory and aging: individual differences, such as cognitive slowing (Salthouse, 1991, 1992), reduced processing resources (Craik, Morris & Gick, 1990; Salthouse, Babcock and Shaw, 1991), and disinhibition of extraneous stimuli (Hasher & Zacks, 1988). These factors are reminiscent of conditions often thought to produce age differences in free recall tasks, such as the short exposure time to the stimuli, disinclination to initiate appropriate strategies, and proactive and retroactive inhibition. The only way to determine whether

age differences in memory performance are due to differences in working memory or long-term memory ability is to assess working memory concomitantly.

Procedural factors of the test and the testing situation are a third area of consideration when analyzing age differences. As mentioned above, individual differences have been suggested as being responsible for such differences. When analyzing age differences in the learning and memory of new material, Huppert (1991) recommended a distinction be made between 'usual' vs. 'successful' aging. She noted that factors such as lower education levels can impact negatively on memory performance, but can be offset by leisure activities. The current study depended on volunteers, and all participants were screened for cognitive deficits and emotional complaints. It is likely that these factors resulted in the inclusion, for the most part, of older participants who belonged primarily to the 'successful' aging group.

Heterogeneity in memory performance among older adults was also pointed out by Dywan & Jacoby (1990), who recommended that it would be more useful to examine correlated changes in memory processing or memory functions, rather than the effects of age itself. Heterogeneity was found on the second and third trials of the RM Designs, the only test which reflected age differences in performance: variability in performance increased with age, and more than doubled from the youngest to the oldest age group. The development of age-appropriate norms is useful for minimizing misdiagnosis when variability is high.

Several times, the term 'environmental support' has been mentioned as a factor in age differences found during both encoding and retrieval stages. Some of the components

subsumed under this phrase are the complexity of the material presented and the degree to which cues are available, the amount of exposure time for each target stimulus and for identification of the target during the testing portion, the number of learning trials, and the type of recognition required (yes/no for individual items, or a forced multiple choice format with one or more distractor items).

There must be a degree of balance between simplicity and complexity of stimulus materials, depending on the purpose of the test in order for a test to be effective. The test materials used in this study were uncomplicated line drawings of everyday objects, large-print words representing common concrete items, and large and clear abstract designs. Park and colleagues (Park et al., 1986; Park et al., 1984) found that greater elaboration in a picture, a line drawing, or a cartoon resulted in better performance across all age groups in a community population (that is, provided more environmental support). Howell (1972) reported that complexity improved the recognition of stimulus materials. In general, the fewer the environmental cues, the less automatic the processing and the more age differences are apt to be present. The cues, however must be appropriate cues (Bartlett & Leslie, 1986).

Too many cues may also be ineffective. Stuart-Hamilton, Rabbitt, and Huddy (1988) reported that patients with severe memory disorders have difficulty identifying a target to which they have been trained to respond (such as a picture on the door identifying their room) when there is too much stimulation distracting from the target. They concluded that the current trend to provide high stimulation in long-term care facilities may interfere with the individual's ability to recognize the necessary cues. The

extraneous cues may cause a stimulus overload for an individual with limited attention and serious cognitive deficits. This was supported by Bird and Luszcz (1993), who restricted their experimental procedure to one type of encoding strategy and one type of retrieval strategy for patients with Alzheimer's disease (AD) because pilot testing indicated that more than one strategy was too confusing to provide any support for memory. While it is impossible and impractical to relieve the environment of all extraneous cues, these findings are important to those designing effective supporting environments for individuals with generalized cognitive deficits, particularly AD.

Since the tests in this study were designed to assess the ability of those with known memory losses and their ability to attend to visual cues and remember them for future use, the items purposely provided a good deal of environmental support. The pictures and words proved to be relatively easy to learn and remember, as evidenced by the high levels of performance on the second trial of both the RM Pictures and RM Words tests in this community sample. The RM Designs test also reached ceiling levels for most of the younger and some of the older participants, but they provided more of a challenge.

Other details affecting the difficulty of a memory task are the number of items presented, length of study time during encoding, and length of response time during retrieval. This study called for 48 words and pictures, and 24 abstract designs. Target stimuli were studied for 3 sec. and the paradigm called for an unlimited decision time. These relatively long examination and decision times were allowed because the time required to process information increases with age (Cohen, 1988; Hartley, 1992; Salthouse, 1985). If shorter exposure times were allowed, age differences may have been

greater. The number of trials chosen was meant to allow sufficient encoding without emphasizing the effects of overlearning. Performance increased significantly from Trial 1 to Trial 2 on each of the three Recognition Memory Battery tests, indicating that improvement was possible, but that some overlearning was occurring.

An additional factor which has been reported as effecting varying performance with age is the degree to which the initial information has been learned. It has often been reported that when criterion levels are met by all subjects, age differences are reduced in memory performance (e.g., see Harwood & Naylor, 1969; Crook et al., 1990; Youngjohn & Crook, 1993). (But on the other hand, see Huppert & Kopelman, 1989). To assess levels of remembering and forgetting in old vs. young subjects, Harwood and Naylor (1969) required that each subject reach a criterion level of 80% recognition at the time of an initial learning session. Hart and O'Shanick (1993) used a level of 65% for words, and a 75% criterion for words and designs. The lower level was chosen for pictures in order to avoid the effects of overlearning. Since a large number of participants in the current study exceeded these levels of learning, it is difficult to make direct comparisons with studies involving criterion levels, which, while trying to equalize performance levels, strive to avoid overlearning. However, it is interesting to note that the Recognition Memory Battery test in which performance was lowest (RM Designs) was the test producing age differences in performance on all trials.

A further test factor affecting the degree of age differences is the type of recognition test administered. Most researchers, employing a single item method which requires the subject to give a yes/no (or new/old) response to each item as it appears

singly, find distinct age differences at some point during the testing (e.g., see Harwood & Naylor, 1969; Park et al., 1986; Poon & Fozard, 1980; Trahan et al., 1986). This method consistently reflects lower performance on recognition memory than does a multiple choice format (Huppert & Piercy, 1976). In a study using both the yes/no procedure and a forced multiple-choice method, Freed, Corkin and Cohen (1987) noted that the multiple choice method is less difficult, less prone to response bias, and more reliable than the yes/no procedure. When using a yes/no procedure in a recognition memory test, there is always a concern that the number of correct answers will be artificially inflated if a score consists of the number of stimulus items correctly identified as 'previously shown.' A perfect score could be produced by an individual having poor memory ability if that person answered "yes" to all items, correct or incorrect. Signal detection methods are generally applied to this paradigm to counteract this possible response bias. A multiple choice format bypasses this by requiring a 'yes' choice every time, while at the same time reducing any possible location bias by counterbalancing correct items on the response pages.

Corkin and Cohen (1987) also suggested that different recognition procedures yield discrepant results when administered to clinical subjects and at different delay periods, and called for research conducted with a variety of recognition procedures at a number of delay intervals. It is possible that in a multiple item forced choice such as that used in this study, the distractor stimuli act to cue memory, allowing decisions to be made more on the basis of (eliminating) environmental cues, rather than having the decision depend only on the recognition of cues inherent to the single item. Since this test was

designed to be used in a clinical setting to test retained memory as opposed to identifying memory deficits, a multiple choice procedure is an appropriate format when considering the above-mentioned report of Freed et al. (1987).

Overall, then, despite reported age differences in the literature, two of the three tasks included in this study revealed no age differences in either acquisition or retrieval from long-term memory. The third task, memory for abstract designs, reflected the more usual reports of age differences in learning, but not in delayed memory. The findings described for the picture and word task likely reflect the research design used (multiple trials of multiple choice recognition memory), the screening methodology used for participants, and the varying difficulty of the types of stimulus material presented.

Rates of Forgetting

It has been reported that patterns of memory loss vary as a result of different neurological conditions (e.g., see Brandt, Corwin & Krafft, 1992; Huppert, 1991; Kopelman, 1985; Moss, Albert, Butters & Payne, 1986; Warrington, 1944) and between younger and older intact adults (Huppert & Kopelman, 1989; Park, Puglisi & Smith, 1986). As a result, one important area of memory research has focused on the rate of forgetting information. A variety of methods have been reported for determining rates of forgetting, including the equating of initial learning, comparison of the slopes of forgetting curves, and calculating difference scores and savings scores (percent retained). On the Recognition Memory Battery, there were no age differences in the rate of forgetting, as evidenced by the slopes of the retention curves for the RM Designs. In fact, percent remembered scores ($\text{trial 2 score} - \text{delay trial score} / 100$) revealed little or no forgetting

across any of the tests. Previous reports, however, would predict that with the test presented as it is normed (i.e., two learning trials at 3 sec. exposure per trial), some of the patient groups are likely to show a rate of forgetting different from the results found for these community-dwelling, cognitively intact adults. This possibility will be addressed in the final section of this discussion under clinical implications.

Previous studies examining rates of forgetting in normal adults generally compare rates of forgetting for younger and older adults, and fit one of two categories: normative information from clinical memory testing, or experimental data. Memory studies reported in clinical fields reveal that as age increases, greater acquisition deficits are found. Data from these sources on rates of forgetting, however, are ambiguous on the matter of retention deficits, with some studies reflecting equal rates of forgetting for young and old adults (Cullum, Butters, Troster & Salmon, 1990; Measso, Romani, Martinin & Zappala, 1990; Trahan, 1992; Trahan & Larrabee, 1992; Youngjohn & Crook, 1993), and some studies showing faster rates of forgetting with age (Cullum, Butters, Troster & Salmon, 1990; Ivnik, Tangalos, Petersen, Kokmen & Kurland, 1990). Clinical tests, however, tend to use only free and/or cued recall and do not equate initial learning, making it hard to tell if the information is due to encoding deficits, storage difficulties, or retrieval failure.

Experimental studies, on the other hand, often use a recognition format with pictures or line drawings as stimuli, which reduces encoding demands and minimizes retrieval failure. In addition, experimental studies tend to control or equate initial learning by choosing a criterion level (number correct) through the use of additional exposure time to the target stimuli, by matching subsets of subjects from younger and older groups,

and/or by using statistical techniques to reduce the effect of differences in initial levels of performance (Huppert & Kopelman, 1989; Park, Puglisi & Smith, 1986). This general paradigm equalizes encoding and storage problems, and assesses retrieval scores at various times post-learning to evaluate the amount of information lost from storage. These studies all reported faster rates of forgetting in older subjects. One study, however, reported equal levels of retention after delay intervals of 10 min., 2 hrs., and 48 hrs. (Rybarczyk, Hart & Harkins, 1987).

In order to determine a rate of remembering (or its inverse, the rate of forgetting) measures of retrieval must be taken at a minimum of two points in time. Clinical memory tests usually use delays up to 30 min., but experimental studies generally measure retention at three or more retrieval times, such as the 10 min., 1 day, and 7 days time intervals used by Huppert and Kopelman, or Rybarczyk et al.'s (1987) retention intervals of 10 min, 2 hr, and 48 hr. An alternative is to encourage ceiling effects by allowing (while still equating) overlearning. This procedure would permit an evaluation of forgetting for well-learned material, a much more realistic picture of the usefulness of training an individual with memory deficits to use cues. The current tests, which compared percent remembered scores for performance 5-10 sec. after the last learning trial and again 20 min. later, were closer to an overlearning procedure, as ceiling effects were present for all of the tests.

A small group of researchers have evaluated rates of forgetting by equating initial learning levels but altering the recognition memory format to include other types of stimulus materials. Results from these studies conclude that age differences for rates of

forgetting are reduced or non-existent when initial learning is equated (Youngjohn & Crook, 1993), when cognitive manipulation is implemented (Rabinowitz, 1989), or, as Huppert (1991) concluded, when the task is free of "memory load" (that is, lower processing demands or structured memory activities). Hasher and Zacks (1979) term this memory load as "effortful memory". This calls to mind the conclusions drawn by Craik et al. (1995), that the more difficult a memory task (at any step in the process), the more pronounced age differences will be, unless effective strategies are encouraged.

Type of Stimulus Material

Although this study was not designed to evaluate the interrelations of the three Recognition Memory Battery tests, it would be useful to examine what appears to be happening for each type of stimulus material. Because the intended purpose of these tests is to evaluate retained learning and memory ability when known memory deficits are present in the individual, the test materials and design were relatively uncomplicated and, to a large extent, parallel. The format was equated for the presentation and evaluation phases of each test, with the only difference being the halving the number of abstract designs compared to pictures and words.

As predicted, the three tests evaluated in this study reflected different patterns of results, thus requiring the development of normative tables specific to each test. Consistent with reports in the literature (Hart & O'Shanick, 1993; Howell, 1972; Measso et al., 1990; Millis & Dijkers, 1993; Noldy et al., 1990; Warrington, 1984; Youngjohn & Crook, 1993), performance was better on recognition memory for pictures than for words. In addition, performance on the RM Pictures was better than on the other two tests.

No attempt was made to equate the difficulty of the items comprising the three tests. In fact, familiar items were chosen for the line drawings, with unique distractor items for each of the three trials in order to avoid interference effects of viewing the 'wrong' item more than once. In addition, although response items could be easily named, each could be thought to represent a different semantic category (e.g., *purse* / *boot* / *dustpan* / *mouse*; or *birdbath* / *skis* / *tincan* / *horse*). The same was true for the words, which were all common concrete nouns (e.g., *elf* / *chimney* / *mule* / *string*; or *muzzle* / *heart* / *drawer* / *plaza*). Therefore, both the pictures and words allowed dual (verbal and non-verbal) encoding to occur, particularly since a relatively long 3 sec. presentation time was used in this project for each of the trials. This may account for the somewhat similar results found across all trials for these two sets of stimulus materials in terms of the absence of age effects as well as ceiling effects on all trials for the majority of participants, and essentially no forgetting when tested after a 20-minute delay.

But scores on the RM Words test were lower than on the RM Pictures test. Although dual encoding could take place for both verbal and non-verbal forms of common nouns, it may be more likely to occur for pictures than for words. Research has shown that given sufficient resources (time, capability to attend, familiarity of the items, etc.) humans attempt to assign verbal labels to visual information, even complex abstract designs (Kimura, 1963). Further supporting a dual processing hypothesis are reports that when verbal information is provided along with visual information (that is, a name attached to a face), scores are even higher than when faces or a list of connected words are learned and remembered (Measso et al., 1990; Youngjohn & Crook, 1993).

It may be that it is more difficult to conjure up a visual image of a word than it is to attach a verbal label to a picture. It has been shown that the type of words chosen as stimuli (e.g., low frequency vs. high frequency) may affect the level of performance (Poon & Fozard, 1980), and Lehman & Mellinger (1986) reported that when words are presented and recognized in the same mode (i.e., auditorally or visually), performance is lower than when they are recalled in the opposite mode (i.e., presented visually and recognized auditorally, or presented auditorally and recognized visually). Greater effort required at the time of encoding would leave fewer resources for deeper processing of the word. For example, when a picture of a roller-skate is presented, not only does the verbal label automatically spring to mind, but a category clue is also presented by the visual image (shoe-type vs. clamp-on, high-top vs. low boot, four-wheel vs. roller-blades). If, on the other hand, the word 'roller-skate' is presented, a visual image needs to be created in order for dual encoding to take place. Let us say the individual visualizes a 'shoe-type high-booted 4-wheel roller skate'. The imagined picture then has to be analyzed to develop those category clues which are automatically available when the presentation mode is visual. This more effortful process may ensure future recognition of the word 'roller-skate,' but 3 sec. does not always allow enough time to develop the visual image, categorize it, and then develop the associations required to ensure its subsequent retrieval from storage. Or perhaps a visual image does not naturally arise whereas a verbal labelling is an automatic reaction for humans as a language-oriented organism.

Although there are differences in performance scores on the RM Pictures and RM Words tests, the differences between performance on the RM Pictures and RM Designs

tests are more notable, including significant age differences and discrepant acquisition scores. Both tests provide visual images for encoding and later recognition, yet scores on the RM Designs test reflected greater difficulty with recognizing the stimuli than was experienced with the Pictures.

The RM Designs test varied from the other two tests in several ways. First, although the nature of the stimuli made it more difficult (but not impossible) to assign meaningful verbal labels, the multifaceted nature of many of the designs allowed verbal labeling of visual cues to be undertaken during each 3 sec. study period on each of the learning trials. In fact, when comments were made by the participants about the difficulty of the designs, they commented that they did attach verbal labels to many of the designs, usually to a specific portion of the design. Milberg, Hebban and Kaplan (1986) also reported that both verbal and visuo-spatial material is normally processed for detailed or feature information.

In a set of experiments using random shapes, Ellis and his colleagues (Daniel & Ellis, 1972; Ellis & Daniel, 1971; Ellis & Homan, 1968; Ellis & Muller, 1964) specifically assessed both recognition memory and free recall of labeling for those memories, if it occurred. Some subjects were trained to use verbal labels; others may or may not have used them automatically. These researchers indicated that verbal labeling helped at the time of encoding by focusing on visual details inherent in the item. The labels were subsequently forgotten, even though the shapes were recognized at recall. It may be that at the time of encoding or acquisition, processing of visual information is not solely automatic, particularly when sufficient time is available for some sort of elaboration.

However, once the image has been 'learned' (available at immediate recall), recognition is indeed more automatic, to the extent that the cues used to identify an item are not available, while the memory itself is present. When the individual professes to not know the correct answer, it may be because they are accessing their verbal hemisphere, and they cannot remember the verbal cues. But when non-verbal processing is accessed, the correct answer is available, even if it is not 'labelled.'

A second way in which the RM Designs task varied relates to the type of items used during the retrieval trials. Distractor items were in the same 'class' as the target, having the same general outline and many of the same features (see again Figure 17). These findings were different that those of Hart and O'Shanick (1993) who used pictures, words and designs in a within-subject design, and reported word stimuli to provide the largest challenge.¹⁴ They felt that this may be due to what they called an "all-or-nothing" approach for the encoding of the non-verbal material, while verbal material tends to be semantically encoded, which requires more effort. Gardiner and Java (1990) reported a higher level of 'know' responses for non-words than for words, and concluded that normal adults process non-word information semantically. This concept was further supported by Bartlett and Leslie (1986), who presented pictures of individual's faces for later recognition. Performance was reduced when the same face was presented in a slightly different profile compared with the presentation of the same face in the identical profile viewed during encoding. It seems that the stimulus item (face) was encoded as a visual image, not as a three-dimensional entity of a person with a unique presence. In much the

¹⁴ The 'designs' used by Hart and O'Shanick included some fairly common geometric shapes, such as a triangle and circle.

same way, the RM Designs items had little or no connection to the individual's previous experiences, and thus requiring the application of strategies to make these connections. RM Words and RM Pictures items did not require the same level of strategy development, as items were already familiar to the participants.

Besides the inherent differences among the types of stimuli, procedural factors may have been responsible for varied performance on these recognition memory tests. One possibility is the sequencing of the three tests. Proactive and retroactive interference are known to have an effect on memory performance, and a large amount of material was presented in the testing situation (see Table 1). The RM Pictures test, which had the highest scores, was the first of the three Recognition Memory Battery tests presented in the two-hour testing session. Its position may have served to avoid proactive interference for these stimuli compared with the material in the other two tests. However, since the RM Words test was the last of the three tests, it should have reflected the lowest performance of the three if proactive interference was to have an affect on the scores of a later-presented test. It did not.

Another consideration is that the number of items to be learned might relate to the degree of difficulty in a test. All else being equal, if less items need to be learned during a given test, that test should be easier. If this were true, however, the RM Designs test should have been the easiest since it had only half the number of items as the other two tests. However, on the RM Designs test, participants earned the lowest percent correct scores of the three tests.

A further possibility is that the distractors were harder for the RM Designs test. As noted before, the specific distractors used in a study may affect the levels of performance. In the case of these three tests, responses on a given item were not evenly distributed across the three alternate choices in some cases, but tended to center on one or another of the distractor items. It is not known whether this is because the distractor had a higher innate familiarity than the target or whether it was more unique than the other distractor items, but it indicated the choice was made on the basis of the specific distractor rather than an uncertain rejection of the target stimulus.

It appears, therefore, that differences noted in performance among these three recognition memory tests were either due to differences in the degree of difficulty of the three types of stimulus materials or to the type of distractors used (same-category versus different-category). As mentioned above, the test items were not calculated to be equally challenging, so this question is difficult to resolve.

Theoretical Implications

Gardiner and Parkin (1990) define recognition as "...what happens when an individual identifies a stimulus as having been encountered previously" (p. 579). It is not yet clear whether that identification is made on an implicit or explicit level; likely, it depends on the circumstances.

Researchers earlier believed all memory, including recognition to be a single process under a continuous dimension of response strength. The theory stated that the stronger the "memory trace" the less effort it takes to recall the event (e.g., see Anderson & Bower, 1972, 1974; Tulving, 1972). Tulving and Thompson (1973) spoke of this

representation as a unique 'trace' established at the time of encoding (or presentation of a stimulus). Recognition was presented as a phase of all memory that follows an evaluation involving decisions about the generation of 'lexical realizations' of ideas. The recognition phase consists of a judgment as to the 'correctness' of the idea or item. If the idea / item was presented in the environment (such as in a recognition memory test), the generation phase was not necessary, making recognition of the item a much easier process.

But as Gardiner and Java (1990) have pointed out, all strong memories are not necessarily conscious memories. More recent theories view recognition and recall as two separate, mutually activated processes. These processes are often termed 'know' and 'remember' (Gardiner & Java, 1990; Tulving, 1985). At the presentation of a to-be-recognized stimulus, these processes are called upon. If the stimulus is 'known' (recognized as being correct), the 'remember' process is not needed unless the individual or task requires more information (such as when or where it was previously encountered).

Mandler (1980) spoke of these two components as 'search' and 'familiarity,' and felt that the extent to which the two processes are used to make recognition judgments depends on the familiarity of the item and the circumstances of the memory task. Similar to Mandler's approach, if a stimulus is immediately recognized (often unconsciously), the search process may not be undertaken or completed.

Jacoby and his colleagues (Dywan & Jacoby, 1990; Jacoby et al., 1989; Jacoby, Woloshyn, & Kelley, 1989; Johnston et al., 1985) used the term 'perceptual fluency' to refer to the familiarity component of this dual process model, and also concluded that conscious recollection is a separate act from assessing familiarity. If perceptual fluency

(familiarity) is high, an individual will provide an identification in a recognition memory format. On the other hand, if perceptual fluency is low, a subsequent 'search' (presumably of storage), will be implemented or will continue. This 'search' involves seeking memory in a much more organized way.

These researchers reported that with the addition of a divided attention component during a recognition memory test, recognition memory performance was greatly reduced, but gains continued to be made in familiarity (Jacoby, Woloshyn & Kelley, 1989). The findings were supported later by Gardiner and Parkin (1990). Thus, it appears that some situations only require a familiarity component (e.g., determination of whether or not an item was previously encountered), whereas if an identification is required (was the encounter remembered the appropriate one? or, under what conditions was the item encountered?), the search component also may also required. The onset of these two processes may be simultaneous or sequential. Or, perhaps, a simultaneous onset occurs, but if immediate recognition occurs, if the situation only calls for 'knowing' the item, or if the search component is unsuccessful, that search is terminated.

A variety of situational factors affect the two components, aside from the identification of the object. For example, recognition is enhanced by integration of details within an item or surrounding a situation (greater context), while recall is enhanced by interrelating items (Griffith, 1975; Mandler, 1980). Organizational variables increase, but do not necessarily speed up recognition performance. For example, additional item exposure increases the recognition probability by increasing its familiarity (Hogan & Kintsch, 1971; Noldy et al., 1990). And finally, while familiarity is more of an automatic

process and often can be successful when the search process is unsuccessful (i.e., an individual recognizes something, but does not necessarily recognize its source or details about it), familiarity decays much more quickly than information that is accessed through a memory search (Hogan & Kintsch (1971; Noldy, Stelmack & Campbell, 1990). In addition, recognition failure can also occur. This involves items which can be recalled, but not recognized (Rabinowitz, 1984).

Jacoby, Kelley, Brown, & Jasechko (1989) found that unconscious errors in recognition can be influenced by recollecting the source of the memory, raising the criterion level for judgments, or changing from decision making grounded on familiarity to a more analytic basis for making judgments. It is likely that in the tests comprising the Recognition Memory Battery, the presence of two trials and fairly long exposure and decision making times allowed at least several of these aids to be put into place by the individuals as long as motivation was present to perform in an optimal way.

These dual component theories do not negate current comprehensive theories of memory. For example, the depth of processing concept of Craik and Lockhart (Craik & Lockhart, 1972; Lockhart & Craik, 1990) remains valid. Deeper processing during encoding increases familiarity, allows the 'know' response to be more easily triggered, and requires less of a search component. Deeper processing would be aided by additional exposure time and more context cues.

Dual processing theories of recognition memory also fit in with Tulving's concepts of episodic and semantic memory. Most recognition memory tests are episodic by nature, as they present new items for learning, often in an out-of-context manner (individual

words, line drawings with little detail, abstract designs). But although recognition memory is considered to be easier than free recall, episodic memories are more difficult to access than are semantic items. In a set of experiments evaluating the relationship between recognition memory and conscious awareness, Gardiner (1988) found that levels of processing and generation effects (both features of episodic memory) occurred only when recognition memory was accompanied by conscious recollection. He concluded that within the recognition memory process, 'know' responses, which involve a more implicit component, are accessed through semantic, or even procedural memory. A 'remember' response, which involves more elaboration and/or conscious recollection, is affected by episodic phenomena.

In terms of Squire's (1992) model of declarative/nondeclarative memory, recognition memory appears at first glance to belong with the nondeclarative system. Nondeclarative memory relies on implicit learning, which enhances performance even when the actual recall of the learning experience is unavailable. Implicit learning tends to depend on the characteristics of the stimuli and the environment, and less so on depth of processing or degree of attention paid during the encoding period.

It is clear when observing recognition memory testing that either much of the learning is fairly automatic, or else it is processed more deeply in a manner that is not accessible to verbal explanation. While participants were making a conscious choice (performing a conscious act of pointing to one or another of the stimuli), that choice often appeared to be based on unconscious, or subconscious decisions. Even when participants were reluctant to choose an answer because they were unsure of the correct response, the

appropriate stimulus item was nearly always chosen. Yet, explicit processing of some sort was (also) going on at the time of encoding, as evidenced by comments such as "Oh yes, there are the skis" to one of the designs, or "I thought those were really old-fashioned skates" to the picture of a roller-skate, or (when pointing to the pennant), "I don't know what that's called, but I remember it points in that direction." Interestingly, few such unsolicited comments were made during the RM Words test.

It has been reported that effective learning depends less on the conscious intent to learn than on the degree of enrichment applied during encoding, whether that enrichment is explicit or implicit (Craik & Tulving, 1975; Kliegl, Smith & Baltas, 1989). It is likely that a combination of two types of processes are occurring during a recognition memory test: both intentional learning and incidental learning, both automatic and effortful processing, and both explicit retrieval and implicit recognition.

In summary, although this thesis was not planned to test theoretical aspects of recognition memory, it can be seen as supporting the two-process model of recognition memory. Furthermore, it appears that although processing was fairly automatic for the pictures and words, it was somewhat effortful for designs. The effort, however, appeared to be taking place during the acquisition phase of the experiment, as responses were quick and retrieval scores often reached ceiling levels. Comments made by the participants during the testing indicated that some of the answers, at least, were automatic at the time of recall, appearing to access a 'know' response, while others were more effortful, accessing a 'remember' response.

Clinical Usefulness

This study focused on developing a normative base for the application of the Recognition Memory Battery in a clinical setting. Although data have not yet been collected in such a clinical setting, it is useful at this point to present some of the features of these tests which would support further research to examine its validity. It is understood that validity studies would have to be conducted with the appropriate populations before accurate interpretation could be applied. For example, the comparison of subjects who were above and below the cut-off points on the GDS did not include any subjects over the age of 60. This is of particular concern when considering that the target population is individuals who have known memory deficits, and because the majority of those deficits would likely be related to dementias such as that found in Alzheimer's disease, which are usually found in the oldest age groups. Although it has been shown that older depressed adults exhibit more memory complaints than memory deficits, the reverse has been shown for older persons with moderate to severe dementia (Christensen, Jorm, Henderson, Mackinnon, Korten & Scott, 1994; Dawe, Procter & Philpot, 1992; Flicker, Ferris & Reisberg, 1993; Siu, 1991). That is, they no longer express concern about memory problems, but tend to demonstrate memory impairment.

Caution should be taken when generalizing these findings to those in poor health but who are cognitively intact. Performance in this group may differ from that of the healthy sample gathered for this study. In addition, education levels were higher in this sample than those found in the average elderly population. On the other hand, education levels reported for this study do not differ from those up to the age of 50, and increasingly

will represent the average aging population in the future if current trends in education are followed. The above concerns are particularly important when considering the use of the RM Designs test as a test predicting memory deficits. Since lower levels of education and poorer health status are represented by lower memory ability, comparing scores with this normative data may underestimate one's ability to recognize information. Considering the relatively high scores reported in this study for the RM Pictures and RM Words tests along with reports in the literature that recognition memory improves retrieval of most newly learned episodic information, it is likely that results from further normative populations would also reflect notable retrieval of the target stimuli. Because of relatively lower scores on the RM Designs test, it would be especially important to add to the normative information with patient populations.

However, based on the normative findings, the following predictions could be made. It is clear that at least one of these tests (RM Designs) may prove to be useful as an easily administered screening measure of recognition memory in a clinical population. The test is portable, requires relatively minimal administration time, and because of its interactive format, is engaging. The distribution of scores reflects normally found age effects. Because of the ceiling effects found in this normative study, if an individual earned a score in the normal range during either of the acquisition stages but any reduction occurred in delayed memory, a memory deficit should be suspected and further explored. If the individual's performance on the delayed memory task was in the normal range (little or no loss of information), but scores on the initial learning trials were lower than the normative values for that age group, encoding deficits should be suspected. It may be

useful to combine it with the RM Pictures test, however, so the individual could be drawn into the positive reinforcement of doing well with little effort on a task which initially appears to be arduous because of the large number of items, but which, in fact, is not difficult when memory is intact. Of course, if performance on the RM Pictures test was deficit, there would be little point in administering the RM Designs test as a screening tool.

Depression is a typical concomitant of dementia (Brandt & Rich, 1995), and depression itself is correlated with retrieval deficits (Ouellette-Hughes, 1990). Initial findings from the current study on the range of self-report scores on the depression scale indicated that for those who feel depressed, symptomatology would not interfere with performance on the tests in this memory battery. Intact memory would be indicated for those with performance in the normal range who have been identified with clinical depression. Before scores in the below-average range could be accurately interpreted for this population, validity studies would be need to be conducted using a clinically depressed sample. As a recognition test, however, these measures would bypass the retrieval difficulties usually found with depression. Furthermore, it is engaging and interactive, so it would give an indication of memory ability for this population where other, more effortful tasks may fail.

It was initially thought that the test may be useful for examining verbal memory deficits through the use of the RM Words test in situations where standard verbal recall tests could not be applied. Although ceiling effects were high and no age differences appeared, a range of scores was identified in the normative population. It still may prove

to be useful when screening tests for verbal memory are not available in the individual's primary language, or when aphasia does not allow the individual to express verbally what he or she has learned in the testing situation. Certainly, a comparison of performance on the RM Words test and the RM Designs test would provide helpful information in such a clinical situation.

When a unilateral vascular lesion exists, it has been shown that recognition of words tests differentially help to identify memory loss resulting from left-hemisphere lesions, whether or not the individual is aphasic (Boller & De Renzi, 1969; De Renzi, 1968; Diesfeldt, 1990; Millis & Dijkers, 1993; Trahan et al., 1990; Warrington, 1984). Interestingly, design recognition tests (or face recognition tests) do not provide the same differential identification of right-hemisphere lesions, but act much as the word tests in identifying that brain damage exists. Therefore, if the RM Pictures and Designs tests were interpreted together, one could be fairly certain that left-hemisphere damage would be reflected in deficit performance on both tests, but that when right hemisphere damage exists, RM Words would reflect normal performance, but RM Designs would reflect reduced performance.

Finally, information from the literature concerning Alzheimer's disease (AD) indicates that encoding deficits are a universal feature of AD, whether one follows the stage model (Constantinidis, 1978; Reisberg, Ferris & Crook, 1982) or the subgroup model (Hart, Kwentis, Taylor & Harkins, 1987; Martin, Cox, Brouwers & Fedio, 1985). Indications are, however, that at least in the early stages, the disorder involves deficits in short-term or working memory that interfere with the ability to connect episodic memories

with previously stored information (Baddeley, 1992; Baddeley, Bressi et al., 1991; Craik et al., 1995}. When sufficient time and support is provided for effectively encoding newly learned information, it is retrievable at a later time (Diesfeldt, 1978; Hart et al., 1987; Huppert, 1991; Kopelman, 1985; Moss, Albert, Butters & Payne, 1986). The Recognition Memory Battery tests would reflect these deficits in episodic memory, but because of the longer exposure times than standard tests, it would allow for more encoding than the standard clinical recognition memory test in order to reflect levels of learning that are still available.

Rates of forgetting have been shown to be greater for those with AD than their intact peers in some studies (e.g., see Hart, Kwentis, Harkins & Taylor, 1987; Moss et al. 1986), but equivalent in others (e.g., Becker et al., 1987; Butters, Miliotis, Albert & Sax, 1984; Huppert & Piercy, 1987; Martone, Butters & Trauner, 1986; Volpe, Holtzman & Hirst, 1986). Discrepancies between these two sets of findings related to procedural differences among the designs, such as whether rates are presented as a percent remembered curve over two time periods or as an absolute difference between two scores. Furthermore, subcortical dementias, such as Huntington's Disease and Parkinson's Disease, appear to be characterized by near-normal recognition scores when free recall performance is severely compromised (Butters, Wolfe, Granholm & Martone, 1986; Massman et al., 1990). In addition, a double dissociation has been reported for implicit memory tasks for AD and PD patients, and for AD and HD patients (Bondi & Kasniak, 1991; Harrington, Haaland, & Yeo & Marder, 1990; Salmon, Shimamura, Butters &

Smith, 1988). A use may exist, therefore, for a set of well-normed recognition memory test that includes various semantic and non-semantic materials.

Stuart-Hamilton et al. (1988) recommends that clinicians look at memory loss as secondary to other cognitive deficits in AD patients. Hart et al. (1987) concluded that although a generalized memory deficit appears in all AD patients, different patterns of cognitive impairment represent different regions of cortical dysfunction, and when these original losses are re-assessed one to two years later, they represent greater deficits in overall cognitive dysfunction. Therefore, the use of all three Recognition Memory Battery tests when examining memory with known or suspected dementia would allow for two things: (1) The type of stimulus materials which may prove most useful as memory aids to the individual, and (2) the area of memory likely to experience the greatest amount of loss as the disease progresses.

Overall, it appears that these tests would be useful with a variety of clinical groups for their primary purpose, that of identifying retained memory in an effort to provide appropriate strategies for retaining the quality of life for those with known memory impairment.

Implications for Future Research

An immediate need should be met for gathering information concerning performance on these tests from related populations. For example, age groups should be extended, data from patient groups without known cognitive deficits should be gathered, exploration should be made into the effects of depression on test performance, particularly with older populations. In addition, it should be determined whether the RM Designs test

is of any use in these populations as a test of available retrieval, let alone as a screening test for memory deficits. Data collected on a similar version of the RM Pictures format (Tombaugh, 1996) indicates that the RM Pictures test, at least, should not present any difficulty to patient groups in a hospital setting with or without dementia. Since research by Hart and O'Shanick has indicated that word stimuli present the greatest difficulty of the three types of materials used in this study, this test should be further explored in a clinical setting with patients not identified with dementia before deciding if it would be useful in a population with dementia.

A further area of exploration relates to rates of forgetting for the information on this test. Previous research has consistently shown that rates of forgetting begin to appear even in an intact population after 24 hours and that they are higher for those with cognitive deficits. This is of particular concern as the purpose of developing this test was to evaluate learning and memory ability to maintain quality of life. It would not be of much use to know an individual could learn to use cues and remember what they mean in order to keep a scrapbook, or find the bathroom, or understand how to use a phone if that information was not retained beyond a fairly limited amount of time.

ENDNOTES

1. Although participants were recruited for the present study with the aim of a mean level of education of approximately grade 12-13, the two youngest age groups consisted of participants with lower levels of education and lower WAIS-R Vocabulary scores than appeared in the eight older age groups. These differences, however, can be explained by the correlation between vocabulary and levels of education found in the general population. As education levels increase, vocabulary increases. But since, because of age alone, those in the younger age groups did not have the same chance to achieve the range of educational experiences possible in older age groups, these differences are not necessarily related to various intellectual ability or to sampling variability. As a result, it is likely that ability in these two younger groups from the sample population could be expected to be similar to those in the older age groups. Finally, it is likely that the method of seeking participants so that the same mean level of education occurred between the differing age groups did have some effect, since the variability in all age groups was equivalent.

REFERENCES

Anderson, J.R. & Bower, G.H. (1972). Recognition and retrieval processes in free recall. Psychological Review, 79, 97-123.

Anderson, J.R. & Bower, G.H. (1974). A propositional theory of recognition memory. Memory and Cognition, 2, 406-412.

Army Individual Test Battery. Manual of directions and scoring (1944). Washington, D.C.: War Department, Adjutant General's Office.

Atkinson, R.C. & Juola, J.F. (1974). Search and decision processes in recognition memory. In D.H. Krantz, R.C. Atkinson, R.D. Luce & P. Suppes (Eds.), Contemporary Developments in Mathematical Psychology: Learning, Memory and Thinking, I. San Francisco: Freeman.

Baddeley, A.D. (1978). The trouble with levels: a reexamination of Craik & Lockhart's framework for memory research. Psychological Review, 85, 139-152.

Baddeley, A.D. (1992). Working memory. Science, 255, 556-559.

Baddeley, A.D. (1995). The psychology of memory. In A.D. Baddeley, B.A. Wilson & F.N. Watts (Eds.), Handbook of Memory Disorders (pp. 3-25). Toronto: John Wiley & Sons Ltd.

Baddeley, A.D., Bressi, S., Della Sala, S., Logie, R. & Spinnler, H. (1991). The decline of working memory in Alzheimer's disease: a longitudinal study. Brain, 114, 2521-2542.

Bahrick, H.P. (1970). Two-phase model for prompted recall. Psychological Review, 77, 212-222.

Bak, J.S. & Greene, R.L. (1980). Changes in neuropsychological functioning in an aging population. Journal of Consulting & Clinical Psychology, 48, 395-399.

Bartlett, J. & Leslie, J.E. (1986). Aging and memory for faces versus single views of faces. Memory and Cognition, 14, 371-381.

Bartlett, J., Till, R. & Levy, J. (1980). Retrieval characteristics of complex pictures: Effects of verbal encoding. Journal of Verbal Learning and Verbal Behavior, 19, 430-449.

Beck, A.T., Ward, C.H., Mendelson, M., Mock, J., & Erbaugh, J.K. (1961). An inventory for measuring depression. Archives of General Psychiatry, 4, 561-571.

Becker, J.T., Boller, F., Saxton, J. & McGonigle-Gibson, K.L. (1987). Normal rates of forgetting of verbal and non-verbal material in Alzheimer's disease. Cortex, 23, 59-72.

Benton, A.L. (1974). The Revised Visual Retention Test (4th ed.). New York: Psychological Corporation.

Benton, A.L. & Hamsher, K. DeS. (1976). Multilingual Aphasia Examination. Iowa City, IA: University of Iowa. (Manual, Revised, 1978).

Benton, A.L., Hamsher, K. DeS., Varney, N.R. & Spreen, O. (1983). Contributions to neuropsychological assessment. New York: Oxford University Press.

Berry, D.T.R. & Carpenter, G.S. (1992). Effect of four different delay periods on recall of the Rey-Osterrieth Complex Figure by older persons.

- Bigler, E., Rosa, L., Schultz, F., Haas, T. & Harris, J. (1989). Rey-Auditory Verbal Learning and Rey-Osterreith Complex Figure design performance in Alzheimer's Disease and Closed Head Injury. Journal of Clinical Psychology, 45, 277-280.
- Bird, M. & Luszcz, M. (1993). Enhancing memory performance in Alzheimer's disease: acquisition assistance and cue effectiveness. Journal of Clinical and Experimental Neuropsychology, 15, 921-932.
- Black, F.W. (1973). Cognitive and memory performance in subjects with brain damage secondary to penetrating missile wounds and closed head injury. Journal of Clinical Psychology, 29, 441-442.
- Boller, F. & De Renzi, E. (1967). Relationship between visual memory defects and hemispheric locus of lesion. Neurology, 17, 1052-1058.
- Boller, F. & Spinnler, H. (1967). Visual memory for colors in patients with unilateral brain damage. Cortex, 7, 141-145.
- Bondi, M.W. & Kasniak, A.W. (1991). Implicit and explicit memory in Alzheimer's disease and Parkinson's disease. Journal of Clinical and Experimental Neuropsychology, 13, 339-358.
- Bornstein, R.A. & Chelune, G.J. (1989). Factor structure of the Weschler Memory Scale-Revised in relation to age and educational level. Archives of Clinical Neuropsychology, 4, 15-24.
- Bornstein, R.A. & Chelune, G.J. (1988). Factor structure of the Weschler Memory Scale-Revised. Clinical Neuropsychologist, 2, 107-115.

Bornstein, R.A., Pakalnis, A., Drake, M.E., Jr., & Suga, L.A. (1988). Effects of seizure type and waveform abnormality on memory and attention. Archives of Neurology, 45, 884-887.

Bowles, N.L. (1989). Age and semantic inhibition in word retrieval. Journal of Gerontology, 44, 88-90.

Branconnier, R.J., Cole, J.O., Spera, K.F. & DeVitt, D.R. (1982). Recall and recognition as diagnostic indices of malignant memory loss in senile dementia. Experimental Aging and Research, 8, 189-193.

Brandt, J., Corwin, J., & Krafft, L. (1992). Is verbal recognition memory really different in Huntington's and Alzheimer's disease? Journal of Clinical and Experimental Neuropsychology, 14, 773-784.

Brandt, J. & Rich, J.B. (1995). Memory disorders in the dementias. In A.D. Baddeley, B.A. Wilson & F.N. Watts (Eds.), Handbook of Memory Disorders (pp. 245-270). Toronto: John Wiley & Sons Ltd.

Brinkman, S.D., Largen, J.W. Jr., Gerganoff, G. & Pomara, N. (1983). Russell's revised WMS in the evaluation of dementia. Journal of Clinical Psychology, 39, 989-993.

Brooker, A.E. & George, J.C. (1984). Visual recognition memory of severely head-injured patients. Perceptual & Motor Skills, 59, 249-250.

Brooks, D.N. (1974). Recognition memory, and head injury. Journal of Neurology, Neurosurgery, and Psychiatry, 37, 794-801.

Brown, G.G., Sawyer, J.D., Nathan, A. & Shatz, M.W. (1987). Effects of lateralized cerebral functions on the Continuous Paired-Associate Test. Journal of Clinical and Experimental Neuropsychology, 9, 680-698.

Brouwers, P., Cox, C., Martin, A., Chase, T., & Fedio, P. (1984). Differential perceptual-spatial impairment in Huntington's and Alzheimer's dementias. Archives of Neurology, 41, 1073-1076.

Buschke, H. (1973). Selective reminding for analysis of memory and learning. Journal of Verbal Learning and Verbal Behaviour, 12, 543-550.

Buschke, H. (1984). Cued recall in amnesia. Journal of Clinical Neuropsychology, 6, 433-440.

Buschke, H., & Fuld, P.A. (1974). Evaluating storage, retention, and retrieval in disordered memory and learning. Neurology, 11, 1019-1025.

Butters, N., Miliotis, P., Albert, M.S. & Sax, D.S. (1984). Memory assessment: Evidence of the heterogeneity of amnesic symptoms. In G. Goldstein (Ed.), Advances in Clinical Neuropsychology, 1, pp. 127-159). New York: Plenum Press.

Butters, N., Wolfe, J., Granholm, E. & Martone, M. (1986). An assessment of verbal recall, recognition and fluency abilities in patients with Huntington's disease. Cortex, 22, 11-32.

Cauthen, N.R. (1977). Extension of the Wechsler Memory Scale norms to older age groups. Journal of Clinical Psychology, 33, 208-211.

Cermak, L.S. & Tarlow, S. (1978). Aphasic and amnesic patients' verbal vs. Nonverbal retentive abilities. Cortex, 14, 32-40.

Christensen, H., Jorm, A.F., Henderson, A.S., Mackinnon, A.J. Korten, E.E. & Scott, L.R. (1994). The relationship between health and cognitive functioning in a sample of elderly people in the community. Age & Ageing, 23, 204-212.

Cohen, G.D. (1988). The Brain in Human Aging. New York: Springer Publishing Company.

Cohen, G. & Faulkner, D. (1986). Memory for proper names: age differences in retrieval. British Journal of Developmental Psychology, 4, 187-197.

Constantinidis, J. (1978). Is Alzheimer's disease a major form of senile dementia? Clinical, anatomical, and genetic data. In R. Katzman, R.D. Terry & K.L. Bick (Eds.), Alzheimer's disease: Senile dementia and related disorders (Aging, Vol. 7, pp. 15-25), New York: Raven Press.

Craik, F.I.M. (1977). Age differences in human memory. In J.E. Birren & K.W. Schaie (Eds.), Handbook of The Psychology of Aging (pp. 384-420). New York: Van Nostrand Reinhold.

Craik, F.I.M., Anderson, N.D., Kerr, S.A. & Li, K.Z.H. (1995). Memory changes in normal ageing. In A.D. Baddeley, B.A. Wilson & F.N. Watts (Eds.), Handbook of Memory Disorders (pp. 211-241). Toronto: John Wiley & Sons Ltd.

Craik, F.I.M. & Lockhart, R.S. (1972). Levels of processing: a framework for memory research. Journal of Verbal Learning and Verbal Behavior, 11, 671-684.

Craik, F.I.M., Morris, R.G. & Gick, M.L. (1990). Adult age differences in working memory. In G. Vallar & T. Shallice (Eds.), Neuropsychological Impairments of Short-Term Memory (pp. 247-267). Cambridge: Cambridge University Press.

Craik, F.I.M., Morris, L.W., Morris, R.G. & Loewen, E.R. (1990). Relations between source amnesia and frontal lobe functioning in older adults. Psychology and Aging, 5, 148-151.

Craik, F.I.M. & Simon, E. (1980). Age differences in memory: the roles of attention and depth of processing. In L.W. Poon, J.L. Fozzard, L.S. Cermak, D. Arenberg, & L.W. Thompson (Eds.), New Directions in Memory and Aging: Proceedings of The George A. Talland Memorial Conference (pp. 95-112). Hillsdale, NJ: Erlbaum.

Craik, F.I.M. & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. Journal of Experimental Psychology: General, 104, 268-294.

Crook, T.H., III, Youngjohn, J.R. & Larrabee, G.J. (1990). The Misplaced Objects Test: A measure of everyday visual memory. Journal of Clinical and Experimental Neuropsychology, 12, 819-833.

Crosson, B., Novack, T.A., Trenerry, M.R. & Craig, P.L. (1988). California Verbal Learning Test performance in severely head-injured and neurologically normal adult males. Journal of Clinical and Experimental Neuropsychology, 10, 754-768.

Cullum, C.M., Butters, N., Troster, A.I., & Salmon, D.P. (1990). Normal aging and forgetting rates on the Wechsler Memory Scale-Revised. Archives of Clinical Neuropsychology, 5, 23-30.

Daniel, T.C. & Ellis, H.C. (1972). Stimulus codability and longterm recognition memory for visual form. Journal of Experimental Psychology, 93, 83-89.

Dawe, B., Procter, A. & Philpot, M. (1992). Concepts of mild memory impairment in the elderly and their relationship to dementia - a review. International Journal of Geriatric Psychiatry, 7, 473-479.

Delaney, R.C., Rosen, A.J., Mattson, R.H. & Novelly, R.A. (1980). Memory function in focal epilepsy: a comparison of non-surgical unilateral temporal lobe, and frontal lobe samples. Cortex, 16, 103-117.

Delis, D.C. (1989). Neuropsychological assessment of learning and memory. In F. Boller & J. Grafman (Eds.), Handbook of Neuropsychology, 3, (pp. 3-33). New York: Elsevier Science Publishers.

Delis, D., Cullum, C.M., Butters, N., Cairns, P. & Prifitera, A. (1988). Wechsler Memory Scale-Revised and the California Verbal Learning Test: Convergence and Divergence. The Clinical Neuropsychologist, 2, 188-196.

Delis, D.C., Freeland, J., Kramer, J.H. & Kaplan, E. (1988). Integrating clinical assessment with cognitive neuroscience: Construct validation of the California Verbal Learning Test. Journal of Consulting and Clinical Psychology, 56, 123-130.

Delis, D.C., Kramer, J.H., Kaplan, E. & Ober, B.A. (1987). The California Verbal Learning Test - Research edition. New York: Psychological Corporation.

Delis, D.C., Massman, P.J., Butters, N., Salmon, D.P., Cermak, L.S. & Kramer, J.H. (1991). Profiles of demented and amnesic patients on the California Verbal Learning Tests: Implications for the assessment of memory disorders. Psychological Assessment, 3, 19-26.

De Renzi, E. (1968). Nonverbal memory and hemispheric side of lesion. Neuropsychologia, 6, 181-189.

De Renzi, E. & Spinnler, H. (1966). The influence of verbal and non-verbal defects on visual memory tasks. Cortex, 2, 322-336.

Diesfeldt, H.F.A. (1978). The distinction between long-term and short-term memory in senile dementia: An analysis of free recall and delayed recognition. Neuropsychologia, 16, 115-119.

Diesfeldt, H.F.A. (1990). Recognition memory for words and faces in primary degenerative dementia of the Alzheimer type and normal old age. Journal of Clinical and Experimental Neuropsychology, 12, 931-945.

Duley, J.F., Wilkins, J.W., Hamby, S.L., Hopkins, D.G., Burwell, R.D. & Barry, N.S. (1993). Explicit scoring criteria for the Rey-Osterrieth and Taylor Complex Figures. The Clinical Neuropsychologist, 7, 29-38.

Dywan, J. & Jacoby, L. (1990). Effects of aging on source monitoring: Differences in susceptibility to false fame. Psychology and Aging, 5, 379-387.

Ellis, H.C. & Daniel, T.C. (1971). Verbal processes in longterm stimulus-recognition memory. Journal of Experimental Psychology, 90, 18-26.

Ellis, H.C. & Homan, L.E. (1968). Implicit verbal responses and the transfer of stimulus predifferentiation. Journal of Experimental Psychology 76, 486-489.

Ellis, H.C. & Muller, D.G. (1964). Transfer in perceptual learning following stimulus predifferentiation. Journal of Experimental Psychology, 68, 388-395.

Elwood, R.W. (1991a). The Wechsler Memory Scale-Revised: Psychometric characteristics and clinical application. Neuropsychology Review, 2, 179-201.

Elwood, R.W. (1991b). Factor structure of the Wechsler Memory Scale-Revised in a clinical sample: A methodological reappraisal. Clinical Neuropsychologist, 5, 329-337.

Erber, J.T. (1974). Age differences in recognition memory. Journal of Gerontology, 29, 177-181.

Erickson, R.C. & Scott, M.L. (1977). Clinical memory testing: a review. Psychological Bulletin, 84, 1130-1149.

Fagan, J.F. (1984). Recognition memory and IQ: High visual recognition memory and IQ. Intelligence, 8, 31-36.

Ferris, S.H., Crook, T., Clark, E., McCarthy, M., & Rae, D. (1980). Facial recognition memory deficits in normal aging and senile dementia. Journal of Gerontology, 35, 707-714.

Flicker, C., Ferris, S.H. & Reisberg, B. (1993). A longitudinal study of cognitive function in elderly persons with subjective memory complaints. Journal of the American Geriatrics Society, 41, 1029-1032.

Flicker, C., Ferris, S.H., Crook, T. & Bartus, R.T. (1987). A visual recognition memory test for the assessment of cognitive functioning in aging and dementia. Experimental Aging Research, 13, 127-132.

Folstein, M.F., Folstein, S.E. & McHugh, P.R. (1975). "Mini-Mental State"; A practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research, 12, 189-198.

Foster, N.L., Chase, T.N., Fedio, P., Patronas, N.J., Brooks, R.A., & Di Chiro, G. (1983). Alzheimer's disease: focal cortical changes shown by positron emission tomography. Neurology, 33, 961-965.

Freed, D.M., Corkin, S. & Cohen, H.J. (1987). Forgetting in H.M.: A second look. Neuropsychologia, 25, 461-471.

Gainotti, G. & Marra, C. (1994). Some aspects of memory disorders clearly distinguish dementia of the Alzheimer's type from depressive pseudo-dementia. Journal of Clinical and Experimental Neuropsychology, 16, 65-78.

Gardiner, J.M. (1988). Functional aspects of recollective experience. Memory and Cognition, 16, 309-313.

Gardiner, J.M. & Java, R.I. (1990). Recollective experience in word and nonword recognition. Memory and Cognition, 18, 23-30.

Gardiner, J.M. & Parkin, A.J. (1990). Attention and recollective experience in recognition memory. Memory and Cognition, 18, 579-583.

Glosser, G., Goodglass, H. & Biber, C. (1989). Assessing visual memory disorders. Psychological Assessment: A Journal of Consulting & Clinical Psychology, 1, 82-91.

Gordon, S.K. & Clark, W.C. (1975). Adult age differences in word and nonsense syllable recognition memory and response criterion. Journal of Gerontology, 29, 659-665.

Griffith, D. (1975). Comparison of control processes for recognition and recall. Journal of Experimental Psychology, 104, 223-228.

Grober, E., Buschke, H., Crystal, H., Bang, S. & Dresner, R. (1988). Screening for dementia by memory testing. Neurology, 38, 900-903.

Haist, Shimamura, & Squire (1992). On the relationship between recall and recognition memory. Journal of Experimental Psychology: Learning, Memory & Cognition, 18, 691-702.

Halstead, W.C. (1947). Brain and Intelligence. Chicago: University of Chicago Press.

Hannay, J., Levin, H. & Grossman, R. (1979). Impaired recognition memory after head injury. Cortex, 15, 269-283.

Harkins, S.W., Chapman, C.R., & Eisendorfer, C. (1979). Memory loss and response bias in senescence. Journal of Gerontology, 34, 55-72.

Hart, R.P., Kwentus, J.A., Harkins, S.W. & Taylor, J.R. (1988). Rate of forgetting in mild Alzheimer's-Type Dementia. Brain and Cognition, 7, 31-38.

Hart, R.P., Kwentus, J.A., Taylor, J.R., & Harkins, S.W. (1987). Rate of forgetting in dementia and depression. Journal of Consulting and Clinical Psychology, 55, 101-105.

Hart, R.P. & O'Shanick, G.J. (1993). Forgetting rates for verbal, pictorial, and figural stimuli. Journal of Clinical and Experimental Neuropsychology, 15, 245-265.

Harrington, D.L., Haaland, K.Y., Yeo, R.A. & Marder, E. (1990) Procedural memory in Parkinson's disease: impaired motor but not visuoperceptual learning. Journal of Clinical and Experimental Neuropsychology, 12, 323-339.

Hartley, A.A. (1992). Attention. In F.I.M. Craik & T.A. Salthouse (Eds.), The brain and human aging (pp.3-49), Hillsdale, N.J.: Lawrence Erlbaum Assoc.

Harwood, E. & Naylor, G.F.K. (1969). Recall and recognition in elderly and young subjects. Australian Journal of Psychology, 21, 251-257.

Hasher, L. & Zacks, R.T. (1979). Automatic and effortful processes in memory. Journal of Experimental Psychology: General, 108, 356-388.

Heilbronner, R.L. (1992). The search for a "pure" visual memory test: pursuit of perfection? The Clinical Neuropsychologist, 6, 105-112.

Heilbronner, R.L., Buck, P. & Adams, R.L. (1989). Factor analysis of verbal and nonverbal clinical memory tests. Archives of Clinical Neuropsychology, 4, 299-309.

Herman, D.O. (1988). Development of the Weschler Memory Scale-Revised. The Clinical Neuropsychologist, 2, 102-106.

Hogan, R.M. & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. Journal of Verbal Learning and Verbal Behavior, 10, 562--567.

Horn, J.L. (1982). The Theory of Fluid and Crystallized Intelligence in Relation to Concepts of Cognitive Psychology and Aging in Adulthood. In F.I.M. Craik & S. Trehub (Eds.), Aging and Cognitive Processes, pp 237-278. New York: Plenum.

Howell, S.C. (1972). Familiarity and complexity in perceptual recognition. Journal of Gerontology, 27, 364-71.

Hubley, A.M. (1995). Geriatric Learning and Memory Battery: Test Development, Psychometric Evidence, and Normative Data. Unpublished doctoral dissertation. Carleton University: Ottawa, Canada.

Hulicka, I.M. (1967). Age differences in retention as a function of interference. Journal of Gerontology, 22, 180-184.

Huppert, F.A. (1991). Age-related changes in memory, learning, and remembering new information. In F. Boller & J. Grafman (Eds.), Handbook of Neuropsychology, 5. New York: Elsevier Science Publishers.

Huppert, F.A. & Kopelman, M.D. (1989). Rates of forgetting in normal aging: Comparison with dementia. Neuropsychologia, 27, 849-860.

Huppert, F.A. & Piercy, M. (1977). Recognition memory in amnesic patients: A defect of acquisition? Neuropsychologia, 15, 643-652.

Isaacs, B. & Kennie, A.T. (1973). The Set Test as an aid to the detection of dementia in old people. British Journal of Psychiatry, 123, 467-470.

Ivnik, R.J., Tangalos, E.G., Petersen, R.C., Kokmen, E. & Kurland, L.T. (1990). The Auditory-Verbal Learning Test (AVLT): Norms for ages 55 years and older. Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2, 304-312.

Jacoby, L.L., Kelley, C., Brown, J. & Jasechko, J. (1989). Becoming famous overnight: Limits on the ability to avoid unconscious influences of the past. Journal of Personality and Social Psychology, 36, 326-338.

Jacoby, L.L., Woloshyn, V. & Kelley, C. (1989). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. Journal of Experimental Psychology: General, 118, 115-125.

Johnston, W.A., Dark, V.J. & Jacoby, L.L. (1985). Perceptual fluency and recognition judgments. Journal of Experimental Psychology, 11, 3-11.

Jubel, J. (1991). Spatial abilities and individual differences in visual information processing. Intelligence, 15, 117-137.

Kaplan, E.F., Goodglass, H., & Weintraub, S. (1978). The Boston Naming Test. Boston: E. Kaplan & H. Goodglass.

Kapur, N. (1987). Some comments on the technical acceptability of Warrington's Recognition Memory Test. British Journal of Clinical Psychology, 26, 144-146.

Kimura, D. (1963). Right temporal-lobe damage: perception of unfamiliar stimuli after damage. Archives of Neurology, 8, 264-271.

Kintsch, W. (1978). More on recognition failure of recallable words: implications for generation-recognition models. Psychological Review, 85, 470-473.

Kliegl, R., Smith, J. & Baltes, P.B. (1989). Testing-the-limits and the study of adult age differences in cognitive plasticity of a mnemonic skill. Developmental Psychology, 25, 247-256.

Klonoff, H. & Kennedy, M. (1965). Memory and perceptual functioning in octogenarians and non-agenerians in the community. Journal of Gerontology, 20, 328-333.

Klonoff, H. & Kennedy, M. (1966). A comparative study of cognitive functioning in old age. Journal of Gerontology, 21, 239-243.

Kopelman, M.D. (1985). Rates of forgetting in Alzheimer-type dementia and Korsakoff's syndrome. Neuropsychologia, 23, 625-638.

Korsnes, M.S. & Gilinsky, A.S. (1993). Aging and serial list picture memory. Perceptual & Motor Skills, 76, 1011-1014.

Kramer, J.H. & Delis, D.C. (1991). Interference effects on the California Verbal Learning Test: A construct validation study. Psychological Assessment, 2, 299-302.

Larrabee, G.J., Kane, R.L., Schuck, J.R. & Francis, D.J. (1985). Construct validity of various memory testing procedures. Journal of Clinical & Experimental Neuropsychology, 7, 395-405.

Larrabee, G.J., Kane, R.L. & Schuck, J.R. (1983). Factor analysis of the Wechsler Adult Intelligence Scale and Wechsler Memory Scale: an analysis of the construct validity of the Wechsler Memory Scale. Journal of Clinical Neuropsychology, 5, 159-168.

Lee, G., Loring, D.W. & Thompson, J.L. (1989). Construct validity of material-specific memory measures following unilateral temporal lobe ablations. Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1, 192-197.

Lezak, M.D. (1983). Neuropsychological Assessment (second ed.). New York: Oxford University Press.

Light, L.L. (1991). Memory and aging: four hypotheses in search of data. Annual Review of Psychology, 42, 336-376.

Light, L.L. (1992). The organization of memory in old age. In F.I.M. Craik & T.A. Salthouse (Eds.), The Handbook of Aging and Cognition (pp. 244-272). Hillsdale, NJ: Earlbaum.

Lockhart, R.S. & Craik, F.I.M. (1990). Levels of processing: a retrospective commentary on a framework for memory research. Canadian Journal of Psychology, 44, 87-112.

Logue, P. & Wyrick, L. (1979). Initial validation of Russell's Revised Wechsler Memory Scale: A comparison of normal aging versus dementia. Journal of Consulting & Clinical Psychology, 47, 176-178.

Loring, D.W. (1989). The Wechsler Memory Scale-Revised, or the Wechsler Memory Scale-Revisited? Clinical Neuropsychologist, 3, 59-69.

Loring, D.W., Lee, G.P., Martin, R.C. & Meador, K.J. (1989). Verbal and visual memory index discrepancies from the Wechsler Memory Scale-Revised: cautions in interpretation. Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1, 198-202.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. Psychological Review, 87, 252-271.

Massman, P.J., Delis, D.C., Butters, N., Levin, B.E. & Salmon, D.P. (1990). Are all subcortical dementias alike? Verbal learning and memory in Parkinson's and Huntington's disease patients. Journal of Clinical and Experimental Neuropsychology, 12, 729-744.

Martin, A., Cox, C., Brouwers, P. & Fedio, P. (1985). A note on different patterns of impaired and preserved cognitive abilities and their relation to episodic memory deficits in Alzheimer's patients. Brain and Language, 26, 181-185.

Mattis, S. (1976). Mental status examination for organic mental syndrome in the elderly patient. In L. Bellak & T.B. Karasu (Eds.), Geriatric psychiatry. New York: Grune & Stratton.

Mayes, A. (1987). Human Organic Memory Disorders. In H. Beloff and A.M. Colman (Eds.), Psychology Survey. Leicester, UK: The British Psychological Society.

Maylor, E.A. (1990). Age and prospective memory. Quarterly Journal of Experimental Psychology, 42A, 471-493.

McCarty, S.M., Logue, P.E., Power, D.G. Ziesat, H.A. & Rosenstiel, A.K. (1980). Alternate-form reliability and age-related scores for Russell's Revised Wechsler Memory Scale. Journal of Consulting & Clinical Psychology, 48, 296-298.

McCormack, P.D. (1984). Aging and recognition memory: Methodological and interpretive problems. Experimental Aging Research, 10, 215-219.

McIntyre, J.S. & Craik, F.I.M. (1987). Age differences in memory for item and source information. Canadian Journal of Psychology, 41, 175-192.

McNeil, J.E. & Warrington, E.K. (1993). Prosopagnosia: A face-specific disorder. Quarterly Journal of Experimental Psychology, Human Experimental Psychology, 46A, 1-10.

Measso, G., Romani, L., Martini, E. & Zappalà, G. (1990). Preliminary analysis of effect of "normal" aging on different memory processes and abilities. Perceptual & Motor Skills, 71, 395-401.

- Menich, S.R. & Baron, A. (1990). Age-related effects of recognition memory: consistent versus varied stimulus-response relations. Journal of Gerontology: Psychological Sciences, 45, P88-93.
- Meudell, P. & Mayes, A. (1981). A similarity between weak normal memory and amnesia in two and eight choice word recognition: a signal detection analysis. Cortex, 17, 19-30.
- Milberg, W.P., Hebben, N. & Kaplan, E. (1986). The Boston process approach to neuropsychological assessment. In I. Grant & K.M. Adams (Eds.), Neuropsychological assessment of neuropsychiatric disorders (pp.65-86). New York: Oxford University Press.
- Millis, S.R. & Dijkers, M. (1993). Use of the Recognition Memory Test in traumatic brain injury: Preliminary findings. Brain Injury, 7, 53-58.
- Milner, B. (1968). Visual recognition and recall after right temporal lobe excision in man. Neuropsychologia, 6, 1191-209.
- Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. British Medical Bulletin, 27, 272-277.
- Milner, B. (1975). Psychological aspects of focal epilepsy and its neurological management. In D.P. Purpura, J.K. Penry, & R.D. Walter (Eds.), Advances in neurology (Vol. 8). New York: Raven Press.
- Mitchell, D.B. (1989). How many memory systems: Journal of Experimental Psychology: Learning Memory and Cognition, 15, 31-49.
- Morris, R.B. & Kopelman, M.D. (1986). The memory deficits in Alzheimer-type dementia: A review. The Quarterly Journal of Experimental Psychology, 38A, 575-602.

Moss, M.B., Albert, M.S., Butters, N. & Payne, M. (1986). Differential patterns of memory loss among patients with Alzheimer's Disease, Huntington's Disease, and Alcoholic Korsakoff's syndrome. Archives of Neurology, 43, 239-246.

Murphy, M.D., Sanders, R.E., Gabrielski, A.S. & Schmitt, F.A. (1981). Metamemory and the aged. Journal of Gerontology, 36, 185-193.

Muter, P. (1984). Recognition and recall of words with a single meaning. Journal of Experimental Psychology, 10, 198-202.

Naugle, Chelune, Cheek, Lunders (1993). Detection of changes in material-specific memory following temporal lobectomy using the Wechsler Memory Scale-Revised. Archives of Clinical Neurology, 8, 381-395.

Newcombe, F. (1969). Missile wounds of the brain. London: Oxford University Press.

Nielson, H., Knudsen, L. & Daugbjerg, O. (1989). Normative data for eight neuropsychological tests based on a Danish sample. Scandinavian Journal of Psychology, 30, 37-45.

Noldy, N.E., Stelmack, R.M. & Campbell, K.B. (1990). Event-related potentials and recognition memory for pictures and words: The effects of intentional and incidental learning. Psychophysiology, 27, 417-428.

Osborne, D.P., Jr., Brown, E.R. & Randt, C.T. (1982). Qualitative changes in memory function: Aging and dementia. Aging, 19, 165-169.

Ouelette-Hughes, N.M.G. (1990). An investigation of the impact of clinical depression on the acquisition and retrieval of verbal and visual material. Unpublished master's thesis, Carleton University, Ottawa, Ontario.

Park, D.C., Puglisi, J.T., Smith, A.O. (1986). Memory for pictures: Does an age-related decline exist? Psychology and Aging, 1, 11-17.

Park, D.C., Puglisi, J.T., Smith, A.O. & Dudley, W.N. (1987). Cue utilization and encoding specificity in picture recognition by older adults. Journal of Gerontology, 42, 423-425.

Park, D.C., Puglisi, J.T., & Sovacool, M. (1984). Picture memory in older adults: Effects of contextual detail at encoding and retrieval. Journal of Gerontology, 39, 213-215.

Park, D.C., Royal, D., Dudley, W. & Morrell, R. (1988). Forgetting of Pictures over a long retention span in young and older adults. Psychology and Aging, 3, 94-95.

Perlmutter, M. (1979). Age differences in adults' free recall, cued recall and recognition. Journal of Gerontology, 34, 533-539.

Pheley, A.M. & Klesges, R.C. (1986). The relationship between experimental and neuropsychological measures of memory. Archives of Clinical Neuropsychology, 1, 231-241.

Poon, L.W. & Fozard, J.L. (1980). Age and word frequency effects in continuous recognition memory. Journal of Gerontology, 35, 77-86.

Postman, L. (1963). One trial learning. In C.F. Cofer & B.S. Musgrave (Eds.), Verbal behavior and learning (pp. 295-332). New York: McGraw-Hill.

Powell, G.E. (1979). The relationship between intelligence and verbal and spatial memory. Journal of Clinical Psychology, 35, 335-340.

Prigatano, G.P. (1978). Wechsler Memory Scale: A selective review of the literature. Journal of Clinical Psychology, 34, 816-832.

Rabbitt, P. (1989). Inner-city decay? Age changes in structure and process in recall of familiar topographical information. In L.W.Poon, D.C. Rubin & B.A. Wilson (Eds.), Everyday Cognition in Adulthood and Late Life. Cambridge: Cambridge University Press.

Rabinowitz, J.C. (1984). Aging and recognition failure. Journal of Gerontology, 39, 65-71.

Rabinowitz, J.C. (1989). Judgments of origin and generation effects: Comparisons between Young and elderly adults. Psychology and Aging, 4, 259-268.

Rabinowitz, J.C., Mandler, G., & Barsalou, L.W. (1979). Generation-recognition as an auxiliary retrieval strategy. Journal of Verbal Learning and Verbal Behavior, 18, 57-72.

Randt, C.T. & Brown, E.R (1983). The Randt Memory Test: Administration Booklet. Bayport, NY: Life Sciences Associates.

Randt, C.T., Brown, E.R. & Osborne, D.P. Jr. (1980). A memory test for longitudinal measurement of mild to moderate deficits. Clinical Neuropsychology, 2, 184-194.

Rey, A. (1941). L'examen psychologique dans les cas d'encephalopathie traumatique. Archives de Psychologie, 28, 286-340.

Rixecker, H. & Hartje, W. (1980). Kimura's Recurring Figures Test: A normative study. Journal of Clinical Psychology, 36, 465-467.

Riddell, S.A. (1962). The performance of elderly psychiatric patients on equivalent forms of tests of memory and learning. British Journal of Social and Clinical Psychology, 1, 70-71.

Riege, W. & Inman, V. (1980). Age differences in nonverbal memory tasks. Journal of Gerontology, 36, 51-58.

Ritchie, K. (1988). The screening of cognitive impairment in the elderly: a critical review of current methods. Journal of Clinical Epidemiology, 41, 635-643.

Rosen, W.G. (1980). Verbal fluency in aging and dementia. Journal of Clinical Neuropsychology, 2, 135-146.

Roth, M., Tym, E., Mountjoy, C.W., Huppert, F.A., Hendrie, H., Verma, S., & Goddard, R. (1986). CAMDEX: A standardized instrument for the diagnosis of mental disorder in the elderly with special reference to the early detection of dementia. British Journal of Psychiatry, 149, 698-709.

Russell, E.W. (1975). A multiple scoring method for the assessment of complex memory functions. Journal of Consulting and Clinical Psychology, 43, 800-809.

Russell, E.W. (1988). Renorming Russell's version of the Wechsler Memory Scale. Journal of Clinical and Experimental Neuropsychology, 10, 235-249.

Ryan, J.J., Rosenberg, S.J. & Heilbronner, R.L. (1984). Comparative relationships of the Wechsler Adult Intelligence Scale-Revised (WAIS-R) and the Wechsler Adult

Intelligence Scale (WAIS) to the Wechsler Memory Scale (WMS). Journal of Behavioral Assessment, 6, 37-43.

Rybarczyk, B.D., Hart, R.P. & Harkins, S.W. (1987). Age and forgetting rate with pictorial stimuli. Psychology & Aging, 2, 404-406.

Salmon, D.P., Shimamura, A.P., Butters, N. & Smith, S. (1988). Lexical and semantic priming deficits in patients with Alzheimer's disease. Journal of Clinical and Experimental Neuropsychology, 10, 477-494.

Salthouse, T.A. (1991). Theoretical Perspectives on Cognitive Aging. Hillsdale, NJ: Erlbaum.

Salthouse, T.A. (1992). Influences of processing speed on adult age differences in working memory. Acta Psychologica, 79, 155-170.

Salthouse, T.A., Babcock, R.L. & Shaw, R.J. (1991). Effects of adult age on structural and operational capacities in working memory. Psychology and Aging, 6, 118-127.

Shimamura, A.P. (1986). Priming effects in amnesia: Evidence for a dissociable memory function. Quarterly Journal of Experimental Psychology, 384, 619-644.

Shimamura, A.P. & Squire, L.R., (1988). Long-term memory in amnesia: Cued recall, recognition memory, and confidence ratings. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14, 763-770.

Silverstein, A.B. (1962). Perceptual, motor, and memory functions in the Visual Retention Test. American Journal of Mental Deficiency, 66, 613-617.

Siu, A.L. (1991). Screening for dementia and investigating its causes. Annals of Internal Medicine, 115, 122-132.

Skilbeck, C.E. & Woods, R.T. (1980). The factorial structure of the Wechsler Memory Scale: Samples of neurological and psychogeriatric patients. Journal of Clinical Neuropsychology, 2, 293-300.

Snodgrass, J.G. & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. Journal of Experimental Psychology: Human Learning and Memory, 6, 174-215.

Spreen, O. & Strauss, E. (1991). A Compendium of Neuropsychological Tests. New York: Oxford University Press

Squire, L.R. (1987). Memory and brain. New York: Oxford University Press.

Squire, L.R. (1992). Declarative and nondeclarative memory: multiple brain systems supporting learning and memory. Journal of Cognitive Neuroscience, 4, 232-243.

Statistics Canada (1993). Educational attainment and school attendance. Ottawa: Industry, Science and Technology Centre.

Storandt, M., Botwinick, J. & Danzinger, W.L. (1986). Longitudinal changes: Patients with mild DAT and matched healthy controls. In L.W. Poon (Ed.), Clinical Memory Assessment of Older People, (pp. 277-284). Washington, DC: American Psychological Association.

Stuart-Hamilton, I.A., Rabbitt, P.M.A. & Huddy, A. (1988). The role of selective attention in the visuo-spatial memory of patients suffering from dementia of the Alzheimer type. Comprehensive Gerontology Bulletin, 2, 129-134.

Taylor, L.B. (69). Localization of cerebral lesions by psychological testing. Clinical Neurosurgery, 16, 269-287.

Taylor, L.B. (79). Psychological assessment of neurosurgical patients. In T. Rasmussen & R. Marino, (Eds.), Functional Neurosurgery. New York: Raven Press.

Till, R.E., Bartlett, J.C. & Doyle, A.H. (1982). Age differences in picture memory with resemblance and discrimination tasks. Experimental Aging Research, 8, 179-184.

Tombaugh, T.N. & Hubley, A.M. (1991). Four studies comparing the Rey-Osterrieth and Taylor Complex Figures. Journal of Clinical and Experimental Neuropsychology, 13, 587-599.

Tombaugh, T.N. & McIntyre, N.J. (1992). The Mini-Mental State Examination: A Comprehensive Review. Journal of the American Geriatrics Society, 40, 922-935.

Tombaugh, T.N. & Schmidt, J.P. (1992). The learning and memory battery (LAMB): Development and standardization. Psychological Assessment, 4, 193-206.

Tombaugh, T.N., Schmidt, J.P. & Faulkner, P. (1992). A new procedure for administering the Taylor Complex Figure: Normative data over a 60-year age span. The Clinical Neuropsychologist, 6, 63-79.

Trahan, D.E., Larrabee, G.J. & Levin, H.S. (1986). Age-related differences in recognition memory for pictures. Experimental Aging Research, 12, 147-150.

Trahan, D.E., Larrabee, G.J. & Quintana, J.W. (1990). Visual recognition memory in normal adults and patients with unilateral vascular lesions. Journal of Clinical and Experimental Neuropsychology, 12, 857-872.

Trahan, D.E., Larrabee, G.J., Quintana, J.W., Goethe, K.E., & Willingham A.C. (1989) Development and clinical evaluation of an Expanded Paired Associate Test with delayed recall. The Clinical Neuropsychologist, 3, 169-183.

Trahan, D.E., Quintana, J.W., Goethe, K.E., & Willingham, A.C. (1988). The Visual Reproduction Subtest: Standardization and clinical validation of a delayed recall procedure. Neuropsychology, 2, 29-39.

Treat, N.J. & Reese, H.W. (1976). Age, pacing and imagery in paired-associate learning. Developmental Psychology, 12, 119-124.

Tulving, E. (1984). Précis of Elements of episodic memory. The Behavioral and Brain Sciences, 7, 223-268.

Tulving, E. & Thomson, D.M. (1973). Encoding specificity and retrieval processes in episodic memory. Psychological Review, 80, 352-373.

Vanderplas, J.M. & Garvin, E.A. (1959). The association value of random shapes. Journal of Experimental Psychology, 57, 147-154.

Warrington, E.K. (1984). Recognition Memory Test Manual. Windsor, England: NFER-Nelson.

Warrington, E.K. & James, M. (1967). An experimental investigation of facial recognition in patients with unilateral cerebral lesions. Cortex, 3, 317-326.

Watkins, M.J. & Gardiner, J.M. (1979). An appreciation of generate-recognize theory of recall. Journal of Verbal Learning and Verbal Behavior, 18, 687-704.

Wechsler, D.A. (1945). A standardized memory scale for clinical use. Journal of Psychology, 19, 87-95.

Wechsler, D.A. (1981). Wechsler Adult Intelligence Scale-Revised Manual. New York: Psychological Corporation.

Wechsler, D.A. (1987). Wechsler Memory Scale-Revised Manual. New York: Psychological Corporation.

Wilson, B., Cockburn, J., Baddeley, A. & Hiorns, R. (1989). The development and validity of a test battery for detecting and monitoring everyday memory problems. Journal of Clinical and Experimental Neuropsychology, 11, 855-870.

Wimer, R.E. (1960). A supplementary report on age differences in retention over a twenty-four hour period. Journal of Gerontology, 15, 417-418.

Winograd, E., Smith, A.D. & Simon, E.W. (1982). Aging and the picture superiority effect in recall. Journal of Gerontology, 37, 70-79.

Wood, F.B., Ebert, V., & Kinsbourne, M. (1982). The episodic-semantic memory distinction in memory and amnesia: clinical and experimental observations. In L. Cermak (Ed.), Memory and amnesia. Hillsdale, NJ: Lawrence Erlbaum Associates.

Youngjohn, J.R. & Crook, T.H., III (1993). Learning, forgetting, and retrieval of everyday material across the adult life span. Journal of Clinical and Experimental Neuropsychology, 15, 447-460.

Appendix A**Description of Recruiting Cells for Recognition Memory Battery Study**

AGE GROUP	High School Diploma or Less		Beyond High School Diploma		Total <i>n</i>
	Males	Females	Males	Females	
16-20	11 ¹	15	1	1	28
21-30	5	7	9	9	30
31-40	5	6	8	11	30
41-50	8	15	15	10	48
51-60	12	17	10	15	54
61-65	12	16	5	10	43
66-70	4	12	9	4	29
71-75	11	12	11	7	41
76-80	10	11	6	6	33
81-85	11	9	7	7	34

¹ Number of participants

Appendix B

Administration Procedure for Recognition Memory Battery Tests

Appendix B.01 Administration Directions

- allow a 3 sec exposure time for each target item

Trial 1: "This is a test to see how well you can recognize pictures (words/ drawings). You will have 2 chances to do this. I will show you 48 pictures (words/drawings). Then I will give you a test to see if you can pick out the pictures (words/drawings) I have shown to you. Let's try a sample first. Look carefully at each one."

- present sample items
- if incorrect, repeat sample
- if still incorrect, give correct answer

In your own words, convey to the subject that although they may feel during the test that they will never remember all the items, "**...we know people will remember much more than they think.**" Therefore, they should try not to become discouraged. The best strategy is to concentrate on looking at each picture.

During the test, if subject appears to become discouraged or give up, encourage them to keep looking at each picture (word/design).

(Response portion) "**Point to the picture (word/design) I showed you before.**"

- Subject must give a response (approx. 15 sec)
- If response is not given by 15 sec, encourage subject to guess if they don't know the answer
- Subject must point to the answer, not give the number by the item
- Mark down the beginning time and ending time for each trial

Trial 2: "I'm going to show you the same pictures (words/designs) again. Look carefully at each one and try to remember it."

Delay Trial: "Remember the book of pictures (words/designs) that I showed you earlier? I'm going to show you some pictures (words/designs) and I want you to point to the picture (word/design) I showed you before."

Appendix B.02 Recognition Memory for Pictures - Target Items

Appendix B.03 Recognition Memory for Words - Target Items

GRAPH	WALNUT	FLOUR
HOSE	TRAPEZE	MAGNET
CALENDAR	LOBSTER	ROCK
ELEVATOR	FACTORY	BLOUSE
AMBULANCE	SHIELD	SADDLE
TOOTH	PANCAKE	RAFT
HARMONICA	HELMET	SWAN
SPIDER	FORT	CALCULATOR
GARAGE	VALLEY	DOUGHNUT
ORCHARD	BEACH	MUZZLE
CAMERA	MATTRESS	FOSSIL
BRAIN	RIBBON	GRASS
ALLIGATOR	BURLAP	DOLLAR
DANCER	FLEA	MONKEY
VILLAGE	PIZZA	CHIMNEY
TREE	FOREST	BACON

Appendix B.04 **Recognition Memory for Abstract Designs - Target Items**

Appendix C

Supplementary Memory Tests Used In This Study

Appendix C.01 G-LAMB Word List

WORD LIST:	<u>Stimulus Word</u>	<u>Cued Recall Stimulus Cue</u>
	TIGER	Animal
	GREEN	Colour
	DESK	Furniture
	HAND	Part of Body
	SWORD	Weapon
	NURSE	Profession
	SWEATER	Clothing
	BASEMENT	Part of Building
	GUITAR	Musical Instrument
	SPINACH	Vegetable
	PIGEON	Bird
	CANOE	Boat
	SISTER	Relative
	DAY	Unit of Time
	TENNIS	Sport

ADMINISTRATION:

Trial 1: "This is a test to see how well you can remember several words. You will be given 5 chances to learn the words. This will be the first time. I am going to read you a list of 15 words. Then I will ask you to tell me all the words you can remember. The order you remember them doesn't matter."

-- read word list (2 sec. intervals)

Trial 1 (Free Recall): "Tell me all the words you can remember."

Trial 2 (Cued Recall): "I'm going to help you with the words you missed. Which word was a (type of)..."

- use category list for cues (for missed words only)
- say "correct" or "incorrect," but do not provide correct answer
- on answer sheet, number order in which words were recalled

(con't on next page)

(Appendix C.01: G-LAMB Word List, continued)

Trials 2, 3, 4, & 5: "I'm going to read the list of words again. Then I will ask you to tell me all the words you can remember, including the ones you remembered correctly before."

-- read the entire word list (2 sec. intervals)

Free Recall (FR): "Tell me all the words you can remember."

Cued Recall (CR): "I'm going to help you with the words you missed. Which word was a (type of)..."

- use category list for cues (for missed words only)
- say "correct" or "incorrect," but do not provide correct answer
- on answer sheet, number order in which words were recalled
- discontinue after 5 trials, or after two consecutive FR trials with 15 correct (100%)

.....

Delay Trial (FR): "Remember the list of 15 words that I read to you? Tell me all the words you can remember. The order you remember them does not matter."

-- if the person cannot begin, say, "Remember one word was an animal."

Delay Trial (CR): "I'm going to help you with the words you missed. Which word was a (type of)..."

- use category list for cues (for missed words only)
- say "correct" or "incorrect," but do not provide correct answer
- on answer sheet, number order in which words were recalled

.....

Recognition: "On this next sheet, I want you to circle all 15 of the words that I read to you."

-- subject should circle exactly 15 words, even if they have to guess or erase some.

(Appendix C.01: G-LAMB Word List, continued)

G-LAMB WORD LIST - RECOGNITION TRIAL

Please circle the 15 words that you remember in the following list. Do not circle more than 15 words.

sweater	foot	river
drum	basement	tennis
coffee	mile	teacher
red	tiger	carrot
spoon	baseball	hand
roof	guitar	oil
desk	pear	yacht
pigeon	hour	hour
cabin	lion	canoe
chair	pepper	green
day	aunt	tulip
eagle	spinach	knife
sword	whiskey	moth
hammer	nurse	wind
wool	shirt	sister

Appendix C.02 G-LAMB Paragraph

PARAGRAPH: Carol Simpson, aged 57, is a bookkeeper at Murphy's Garage. She lives in a brown house at 14 Willow Avenue, Toronto. Every morning at 6:15, she walks her black poodle, Brandy, in the park. When she passes the Dutch bakery she buys three ginger cookies. Last Monday, at 2 p.m., she took a bus to Michigan to visit her nephew's farm. While there, she planted carrots, went fishing for perch, and read the Country Digest magazine.

ADMINISTRATION:

Trial 1: "This is a test to see how well you can remember a story. You will be given 2 chances to learn it. This is the first time. I will read the story to you. Then, I will give you 30 sec. To read and study the story yourself. After that, I will remove the story and ask you to tell me all you can remember about it."

- read paragraph
- S may read out loud or silent

"Now, tell me all you can remember about the story."

- circle number of response or record S's responses verbatim
- at end, ask if there is anything else that they can recall about the story, even if it is not in the right order.

.....

Trial 2: "I will read the story to you again. Then, I will give you 30 sec. To read and study the story yourself. After that, I will remove the story and ask you to tell me all you can remember about it."

- read paragraph
- S may read out loud or silent

(Appendix C.02: G-LAMB Paragraph, continued)

"Now, tell me all you can remember about the story."

- circle number of response or record S's responses verbatim
 - at end, ask if there is anything else that they can recall about the story, even if it is not in the right order.
-

Delay Trial: "Remember the story that I read and you studied? Tell me all you can remember about it."

- if S cannot remember, say it was a story about a woman
 - circle number of response or record S's responses verbatim
 - at end, ask if there is anything else that they can recall about the story, even if it is not in the right order.
-

Cued Recall and Recognition:

Directions and scoring sheet for Cued Recall and Recognition trials can be found on the next page.

(Appendix C.02: G-LAMB Paragraph, cont'd)

LAMB PARAGRAPH #2

CR & Recognition: "I'm going to give you some help on the answers you didn't know or left out. I'm going to ask you some questions about the story. If you don't know the answer, I'll give you four choices. You then tell me which one is correct."

- say "correct" or "incorrect", but give no answers

<u>CUED RECALL</u>	<u>RECOGNITION</u>	<u>CR</u>	<u>REC</u>
1. Person's first name?	Carol, Karen, Ann, Cathy	—	—
2. Person's last name?	Andrews, Murphy, Martin, Simpson	—	—
3. Person's age?	42, 53, 57, 69	—	—
4. Person's occupation?	secretary, mechanic, teacher, bookkeeper	—	—
5. Type of place she worked?	garage, bakery, dept. store, factory	—	—
6. Name of her employer?	Richmond, Murphy's, Young's, Mike's	—	—
7. Colour of her house?	green, white, grey, brown	—	—
8. Number of her street address?	3, 7, 14, 18	—	—
9. Name of street where she lived?	Erindale, Queen, Willow, York	—	—
10. Name of the city where she lived?	Seattle, Vancouver, Boston, Toronto	—	—
11. Time she left in the morning?	6:15, 6:45, 7:15, 7:45	—	—
12. Colour of her dog?	grey, black, brown, white	—	—
13. Type of dog she had?	terrier, poodle, collie, beagle	—	—
14. Name of her dog?	Rover, Champ, Peanut, Brandy	—	—
15. Where she walked the dog?	park, woods, farm, field	—	—
16. Where did she buy something?	bakery, market, pharmacy, butcher	—	—
17. Type/nationality of bakery?	Italian, Dutch, German, French	—	—
18. What did she buy at the bakery?	muffins, cookies, doughnuts, biscuits	—	—
19. What type of cookies did she buy?	sugar, oatmeal, raisin, ginger	—	—
20. How many cookies did she buy?	1, 2, 3, 4	—	—
21. What day did she go visit someone?	Monday, Wednesday, Thursday, Saturday	—	—
22. What time did she leave on Monday?	1:30, 2:00, 2:30, 3:00	—	—
23. Type of transportation she used?	plane, car, bus, train	—	—
24. What state did she go to?	Kansas, Ohio, Missouri, Michigan	—	—
25. Who did she go to visit?	brother, son, nephew, niece	—	—
26. What did her nephew own?	company, cottage, ranch, farm	—	—
27. What did she plant?	parsnips, carrots, potatoes, corn	—	—
28. Type of activity/sport she did?	gardening, swimming, fishing, jogging	—	—
29. Type of fish she was fishing for?	sunfish, perch, trout, pickerel	—	—
30. First name of magazine she read?	Country, Valley, Outdoor, Western	—	—
31. Second name of magazine she read?	Gazette, Times, Living, Digest	—	—

FR+CR= _____

FR+CR+REC= _____

Appendix C.03

G-LAMB Simple Figure (Rectangle)

ADMINISTRATION:

Trial 1: "This is a test to see how accurately you can remember a geometric design. You will have 3 chances to do this. This will be the first time. I will given you 30 sec. to study the design. Then I'll remove the design, and ask you to draw it from memory."

- show design for 30 sec. Do not allow the subject to rotate the design. While subject is studying the design, place one blank sheet of paper in front of the subject, folded in half vertically and with the number of the trial written in the top right corner.
- after 30 sec., remove the design

(con't on next page)

(Appendix C.03: G-LAMB Simple Figure, continued)

“Now, draw the geometric design I showed you as accurately as you can. Make it approximately the same size.”

- maximum time limit of 2 min., record time taken
 - subject may make erasures, but cannot use a ruler
-

Trial 2 & 3: “I am going to show you the same design again for 30 sec. Then I’ll remove the design, and ask you to draw it from memory.”

- maximum time limit of 2 min., record time taken
 - subject may make erasures, but cannot use a ruler
-

Delay Trial: “Remember the figure I showed you earlier? Please draw it as accurately as you can. Make it approximately the same size.”

- maximum time limit of 2 min., record time taken
 - subject may make erasures, but cannot use a ruler
-

Copy Trial: “This time, I want you to copy the design. Do not use your memory to do this. Look at the design to make sure you are copying it accurately.”

- place design in front of subject (no rotation)
 - maximum time limit of 3 min., record time taken
 - S may make erasures, but cannot use a ruler
-

Appendix D

Clinical Tests Used In This Study

Appendix D.01 Instructions for Word Fluency Test

Directions: "I will say a letter of the alphabet. Then I want you to give me as many words that begin with that letter as quickly as you can. For instance, if I say 'B,' you might give me 'bad, battle, bed...' I do not want you to use words that are proper names, such as 'Boston, Bob, or Betty.' Also, do not use the same word again with a different ending such as 'bike,' 'biking,' or 'bikers.' Any questions?

"Begin when I say the letter. The first letter is 'F' - go ahead."

-- You must give the full 60 sec. Mark when each 15 sec. is completed.

.....

"Now we are going to do the same thing, but with a different letter this time. Tell me as many words as you can think of that begin with the letter 'A' - go ahead."

-- You must give the full 60 sec. Mark when each 15 sec. is completed.

.....

"This time, I want you to tell me as many words as you can think of that begin with the letter 'S' - go ahead."

-- You must give the full 60 sec. Mark when each 15 sec. is completed.

.....

"This time I want you to tell me as many different animals as you can think of as quickly as you can - go ahead."

-- You must give the full 60 sec. Mark when each 15 sec. is completed.

Appendix D.02 Instructions for Trail Making Test A & B

Trails A Sample: "On this page are some numbers. Begin at #1 and draw a line from 1 to 2, 2 to 3, 3 to 4, and so on until you reach the end. Draw the lines as fast as you can without lifting the pencil from the paper. Ready? Begin."

- point to the first three numbers as you say them, then point to the last number
- person must keep pencil on paper
- if correct, say good; if mistakes are made, correct immediately
- use another sample sheet if need to do over

Trails A: "On this page, there are numbers from 1 to 25. Begin at #1 and draw a line from 1 to 2, 2 to 3, 3 to 4, and so on until you reach the end. Draw the lines as fast as you can without lifting the pencil from the paper. Ready? Begin."

- point to the first three numbers as you say them, then point to the last number
- person must keep pencil on paper
- time subject
- if mistakes are made, stop the person and set them straight again, but do not stop timing

Trails B Sample: "On this page are some numbers and letters. Begin at #1 and draw a line from 1 to A, A to 2, 2 to B, B to 3, 3 to C, and so on until you reach the end. So I want you to alternate, a number then a letter, then a number, then a letter, and so on. Draw the lines as fast as you can without lifting the pencil from the paper. Ready? Begin."

- point to first three number/letter combinations as you say them, then point to the last number
- person must keep pencil on paper
- if correct, say good; if mistakes are made, correct immediately
- use another sample sheet if need to do over

Trails B: "On this page are also some numbers and letters. Begin at #1 and draw a line from 1 to A, A to 2, 2 to B, B to 3, 3 to C, and so on until you reach the end (#13). So I want you to alternate, a number then a letter, then a number, then a letter, and so on. Do not skip around, but go from one circle to the next in the proper order. Draw the lines as fast as you can without lifting the pencil from the paper. Ready? Begin."

- point to first three number/letter combinations as you say them, then point to the last number
 - person must keep pencil on paper
 - time subject
 - if mistakes are made, stop the person and set them straight again, but do not stop timing
-

Appendix D.03 Instructions for Boston Naming Test

1. There are 60 pictures to be presented in order. Begin by saying, "What is this?" Subjects has up to 20 sec to make an initial response (although they need not take the entire 20 sec).
2. Record subjects' responses verbatim. If the person says "four-poster bed" for item 1 (bed), then record that. If they say "box" or "you know, you sleep in it", record that. This applies to any responses, whether it is correct without cue or whether a stimulus or phonemic cue is used.
3. Time how long it takes the subject to respond. This is entered under "Latency Secs" - record in seconds.
4. The stimulus cues are printed in brackets under each response. These are only given when the subject either (a) does not know the answer (i.e., lack of recognition), or (b) their response reflects a misperception. An example of (b) would be if the person thought that item 19 (pretzel) was a snake - they have misperceived the picture. If the person thought item 29 (beaver) was an "otter", this is incorrect, but not a misperception.
5. The phonemic cues refer to the underlined portion of each response. These help the subject by providing the opening sound of the target word. These cues are given either (a) after a stimulus cue has not resulted in the correct response, or (b) when the subject has given an incorrect response and a stimulus cue is not appropriate. For example, if the subject said item 29 (beaver) was an "otter", you would give a phonemic cue, rather than a stimulus cue.
6. Although one typically stops after six consecutive failures on this test, we won't be following that rule in this study. We would like to see performance on each item. Start at item 1 and try to go all the way to item 60. Only stop if you are beyond six consecutive failures and the person is becoming visibly upset. Use your judgment.
7. See the last page of the scoresheet for the scoring instructions.
8. There are no equivalent answers (e.g., 'mouth organ' for 'harmonica' or 'toadstool' for 'mushroom').
9. If a person gives an incorrect response, but has a chance to self-correct before you can give the next cue, mark that as a correct answer.

Appendix D.03 (cont'd)

10. If a person gives correct response after you have told them their initial response is incorrect but before you have given a cue, score it as a spontaneous response, but place a star beside it. E.g. #29, Beaver 'otter', (no) 'beaver'.
 11. If a person answers in a questioning manner (e.g., "Is it a compass?") ask them if that is their response (use a neutral voice).
 12. If a person gives 2 answers - "it's a compass or a protractor", make them choose a response.
-

Appendix E

Demographic Information for Ten Age Groups for Recognition Memory Tests

Appendix E.01 Recognition Memory for Pictures

Demographics	Age Group (number of participants)										ALL (n=347)
	16-20 (n=28)	21-30 (n=30)	31-40 (n=30)	41-50 (n=48)	51-60 (n=44)	61-65 (n=38)	66-70 (n=30)	71-75 (n=36)	76-80 (n=30)	81-85 (n=28)	
Age Mean (s.d.)	18.36 (1.42)	24.07 (3.05)	35.27 (2.72)	45.10 (2.80)	56.32 (2.55)	63.24 (1.42)	68.57 (1.33)	72.83 (1.36)	77.63 (1.22)	82.93 (1.44)	54.59 (20.33)
Education Mean (s.d.)	11.50 (0.88)	13.27 (1.34)	13.57 (2.11)	13.40 (2.67)	13.16 (2.42)	12.82 (2.58)	12.87 (2.60)	12.94 (2.45)	12.47 (2.35)	12.29 (2.32)	12.88 (2.33)
Vocab Mean (s.d.)	44.75 (10.10)	45.97 (11.75)	53.13 (8.93)	52.35 (10.30)	55.48 (7.22)	55.50 (8.13)	54.97 (10.60)	55.44 (9.57)	55.53 (8.42)	57.14 (7.29)	53.22 (9.94)
MMSE Mean (s.d.)	28.89 (1.07)	28.70 (0.95)	28.97 (1.25)	29.04 (1.11)	28.84 (1.12)	28.97 (0.97)	28.57 (1.36)	28.17 (1.16)	28.33 (1.40)	27.86 (1.38)	28.67 (1.22)
GDS Mean (s.d.)	6.54 (3.74)	5.87 (3.47)	6.07 (3.64)	3.85 (2.91)	3.59 (3.77)	3.39 (3.11)	3.47 (2.79)	3.53 (3.00)	4.33 (3.54)	4.86 (3.09)	4.32 (3.45)
Gender Males	12 (42.9)	14 (46.7)	13 (43.3)	23 (47.9)	19 (43.2)	13 (34.2)	13 (43.3)	18 (50.0)	15 (50.0)	14 (50.0)	154 (45.0)
Females	16 (57.1)	16 (53.3)	17 (56.7)	25 (52.1)	25 (56.8)	25 (65.8)	17 (56.7)	18 (50.0)	15 (50.0)	14 (50.0)	188 (55.0)

Note: Education is reported in number of years of formal schooling, Vocab = raw scores on Wechsler Adult Intelligence Scale-Revised Vocabulary subtest (maximum score = 70), MMSE = Mini-Mental State Examination (maximum score = 30), GDS = Geriatric Depression Scale (maximum score = 30). Gender is reported by the number of participants as well as the percent.

Appendix E.02 Recognition Memory for Words

Demographics	Age Groups (Number of Participants)										
	16-20 (n=28)	21-30 (n=30)	31-40 (n=30)	41-50 (n=48)	51-60 (n=43)	61-65 (n=38)	66-70 (n=27)	71-75 (n=19)	76-80 (n=30)	81-85 (n=27)	All (n=340)
Age											
Mean (s.d.)	18.36 (1.42)	24.07 (3.05)	35.27 (2.72)	45.10 (2.80)	56.23 (2.52)	63.24 (1.42)	68.85 (1.06)	72.31 (2.26)	77.63 (1.23)	83.00 (1.41)	54.50 (20.34)
Education											
Mean (s.d.)	11.50 (0.88)	13.27 (1.34)	13.57 (2.11)	13.40 (2.67)	13.09 (2.41)	12.82 (2.58)	13.04 (2.64)	12.82 (2.40)	12.47 (2.35)	12.30 (2.37)	12.88 (2.33)
Vocab											
Mean (s.d.)	44.75 (10.10)	45.97 (11.75)	53.13 (8.93)	52.35 (10.30)	55.19 (7.04)	55.50 (8.12)	54.70 (11.20)	55.59 (9.24)	55.53 (8.42)	56.93 (7.33)	53.14 (9.92)
MMSE											
Mean (s.d.)	28.89 (1.07)	28.70 (0.95)	28.97 (1.25)	29.04 (1.11)	28.81 (1.12)	28.97 (0.97)	28.52 (1.42)	28.23 (1.13)	28.33 (1.40)	27.81 (1.39)	28.66 (1.22)
GDS											
Mean (s.d.)	6.54 (3.74)	5.87 (3.47)	6.07 (3.64)	3.85 (2.91)	3.65 (3.80)	3.39 (3.11)	3.63 (2.88)	3.41 (2.92)	4.33 (3.54)	4.70 (3.04)	4.41 (3.45)
Gender											
Males	12 (42.9)	14 (46.7)	13 (43.3)	23 (47.9)	18 (41.9)	13 (34.2)	11 (40.7)	20 (51.3)	15 (50.0)	14 (51.9)	153 (45.0)
Females	16 (57.1)	16 (53.3)	17 (56.7)	25 (52.1)	25 (58.1)	25 (65.8)	16 (59.3)	19 (48.7)	15 (50.0)	13 (48.1)	187 (55.0)

Note. Education is reported in number of years of formal schooling; Vocab = raw scores on Wechsler Adult Intelligence Scale-Revised Vocabulary subtest (maximum score = 70); MMSE = Mini-Mental State Examination (maximum score = 30); GDS = Geriatric Depression Scale (maximum score = 30); Gender is reported by the number of participants as well as the percent.

Appendix F.03 Recognition Memory for Designs

Demographics:	Age Group (number of participants)											Visual Recognition Memory	
	16-20 (n=28)	21-30 (n=30)	31-40 (n=30)	41-50 (n=48)	51-60 (n=49)	61-65 (n=34)	66-70 (n=31)	71-75 (n=30)	76-80 (n=30)	81-85 (n=28)	A.L. (n=338)		
Age Mean (s.d.)	18.36 (1.42)	24.07 (3.05)	35.27 (2.72)	45.10 (2.80)	51.40 (1.40)	63.29 (1.38)	68.32 (1.45)	72.97 (1.40)	77.79 (1.27)	82.94 (1.39)	84.24 (20.30)		
Education Mean (s.d.)	11.50 (0.88)	13.27 (1.34)	13.57 (2.11)	13.40 (2.67)	12.92 (2.60)	12.15 (2.16)	12.84 (2.60)	12.37 (2.33)	12.36 (2.28)	12.47 (2.44)	12.76 (2.34)		
Vocab Mean (s.d.)	44.75 (10.10)	45.97 (11.75)	53.13 (8.93)	52.35 (10.30)	54.43 (8.59)	55.68 (7.33)	55.68 (10.11)	55.20 (9.06)	54.46 (10.04)	56.94 (7.77)	52.78 (10.02)		
MMSE Mean (s.d.)	28.89 (1.07)	28.70 (0.95)	28.97 (1.25)	29.04 (1.11)	28.73 (1.15)	28.85 (0.96)	28.58 (1.36)	28.32 (1.17)	28.10 (1.21)	27.88 (1.30)	28.64 (1.27)		
GDS Mean (s.d.)	6.54 (3.74)	5.87 (3.47)	6.07 (3.64)	3.85 (2.91)	4.22 (4.13)	3.24 (2.64)	3.26 (2.72)	3.41 (2.95)	3.47 (3.11)	4.53 (3.05)	4.45 (3.49)		
Gender Males	12 (42.9)	14 (46.7)	13 (43.3)	23 (47.9)	21 (42.9)	13 (31.2)	14 (45.2)	15 (50.0)	15 (50.0)	16 (57.1)	156 (46.2)		
Females	16 (57.1)	16 (53.3)	17 (56.7)	25 (52.1)	28 (57.1)	21 (61.8)	17 (54.8)	15 (50.0)	15 (50.0)	12 (42.9)	182 (53.8)		

Visual Recognition Memory

Note. Education is reported in number of years of formal schooling, Vocab = raw scores on Wechsler Adult Intelligence Scale-Revised Vocabulary subtest (maximum score = 70), MMSE = Mini-Mental State Examination (maximum score = 30), GDS = Geriatric Depression Scale (maximum score = 30). Gender is reported by the number of participants as well as the percent

Appendix F

Correlations Among Demographic Variables

Appendix F.01 Recognition Memory for Pictures

VARIABLES	Age	Gender	Education	Vocabulary
Age	1.00			
Gender	-0.01	1.00		
Education	-0.01	-0.09	1.00	
Vocabulary	0.35***	-0.03	0.35***	1.00

* $p < .05$; ** $p < .01$; *** $p < .001$

Note. Education is reported in number of years; Vocabulary = raw scores from Vocabulary subtest of Wechsler Adult Intelligence Scale-Revised.

Appendix F.02 Recognition Memory for Words

VARIABLES	Age	Gender	Education	Vocabulary
Age	1.00			
Gender	-0.02	1.00		
Education	-0.01	-0.09	1.00	
Vocabulary	0.34***	-0.03	0.35***	1.00

* $p < .05$; ** $p < .01$; *** $p < .001$

Note. Education is reported in number of years; Vocabulary = raw scores from Vocabulary subtest of the Wechsler Adult Intelligence Scale-Revised.

Appendix F.03 Recognition Memory for Designs

VARIABLES	Age	Gender	Education	Vocabulary
Age	1.00			
Gender	-0.03	1.00		
Education	-0.04	-0.06	1.00	
Vocabulary	0.31***	-0.04	0.37***	1.00

* $p < .05$; ** $p < .01$; *** $p < .001$

Note. Education is reported in number of years; Vocabulary = raw scores from Vocabulary subtest of Wechsler Adult Intelligence Scale-Revised.

Appendix G

Proportion Remembered Scores for Recognition Memory Battery Tests

RECOGNITION MEMORY TESTS:	Age Group (number of participants)										
	16-20 (n=28)	21-30 (n=30)	31-40 (n=30)	41-50 (n=48)	51-60 (n=44)	61-65 (n=38)	66-70 (n=30)	71-75 (n=36)	76-80 (n=30)	81-85 (n=28)	81+ (n=342)
Pictures											
Trial 2	47.68	47.50	47.73	47.96	47.98	47.95	47.97	47.89	47.87	47.64	47.84
Delay Memory (%Rem) ¹	47.75 (~100%)	47.80 (~100%)	47.90 (~100%)	47.79 (~100%)	47.89 (99.6%)	47.97 (~100%)	47.97 (100%)	47.83 (99.9%)	47.80 (99.9%)	47.86 (~100%)	47.86 (~100%)
Words											
Trial 2	45.96	44.70	46.10	46.19	46.23	45.32	45.56	45.08	43.43	44.81	45.41
Delay Memory (%Rem)	45.75 (99.5%)	45.63 (97.3%)	46.07 (99.9%)	45.67 (98.9%)	46.33 (~100%)	45.63 (~100%)	45.63 (~100%)	44.64 (99.0%)	43.47 (~100%)	45.30 (98.9%)	45.41 (100%)
Abstract Designs											
Trial 2	22.93	23.00	22.60	22.29	21.29	21.29	19.61	20.34	18.70	16.86	21.00
Delay Memory (%Rem)	22.82 (99.5%)	22.70 (98.7%)	22.53 (99.7%)	22.00 (98.7%)	20.88 (98.1%)	20.82 (98.8%)	19.45 (99.2%)	19.53 (95.6%)	18.07 (96.6%)	16.32 (96.8%)	20.62 (98.2%)

¹ %Rem is the percent correct on the delay trial when compared with the individual's score during the final learning trial (Trial 2); or, for each individual, (%Rem = Delay Trial Score ÷ Trial 2).

Appendix H

Rank Order of Item Difficulty for Recognition Memory Tests, Including % Wrong by Item Location

Appendix H.01 Recognition Memory for Pictures

ITEM	TRIAL 1		TRIAL 2		DELAY TRIAL	
	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION
baseball bat	1	0.6 (x/0.0/0.0/0.6)	14	0.3 (0.0/0.0/0.3/x)	5	0.0 (0.0/0.0/0.0/x)
lighter	2.5	0.9 (0.0/x/0.3/0.6)	22.5	0.6 (0.3/0.3/0.0/x)	48	3.5 (0.3/x/3.2/0.0)
muffin tin	2.5	0.9 (0.3/0.3/x/0.3)	14	0.3 (0.0/x/0.3/0.0)	5	0.0 (x/0.0/0.0/0.0)
acorn	5	1.2 (0.3/x/0.9/0.0)	14	0.3 (0.3/x/0.0/0.0)	17	0.3 (0.3/0.0/0.3/x)
spinning wheel	5	1.2 (0.3/x/0.9/0.0)	14	0.3 (0.0/0.0/x/0.3)	17	0.3 (x/0.3/0.0/0.0)
tin can	5	1.2 (0.3/0.9/x/0.0)	14	0.3 (0.0/x/0.3/0.0)	30.5	0.6 (0.0/x/0.0/0.6)
music staff	7	1.5 (0.3/x/0.6/0.6)	22.5	0.6 (x/0.0/0.0/0.6)	17	0.3 (x/0.3/0.0/0.0)
dart	8	1.8 (0.6/x/0.9/0.3)	14	0.3 (0.0/0.3/x/0.0)	30.5	0.6 (0.3/0.0/0.3/x)
tissue box	9	2.0 (0.0/1.5/x/0.5)	45.5	1.5 (x/0.3/0.0/1.2)	17	0.3 (x/0.0/0.0/0.3)
whistle	10	2.4 (1.2/0.9/0.3/x)	31	0.9 (0.3/x/0.3/0.3)	40	0.9 (0.6/0.3/0.0/x)
flashlight	12	2.7 (x/1.2/0.3/1.2)	39	1.2 (x/0.3/0.0/0.9)	17	0.3 (x/0.3/0.0/0.0)
tennis racquet	12	2.7 (0.9/x/1.8/0.0)	31	0.9 (x/0.0/0.0/0.9)	5	0.0 (x/0.0/0.0/0.0)

Appendix H.01 (cont'd.)

ITEM	TRIAL 1		TRIAL 2		DELAY TRIAL	
	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION
toothbrush	12	2.7 (x/0.6/0.3/1.8)	22.5	0.6 (0.3/x/0.0/0.3)	17	0.3 (0.3/0.0/0.0/x)
feather	14	2.9 (0.3/0.3/2.3/x)	45.5	1.5 (0.3/x/1.2/0.0)	5	0.0 (0.0/x/0.0/0.0)
paper clip	15	3.0 (1.2/0.3/x/1.5)	14	0.3 (x/0.3/0.0/0.0)	17	0.3 (x/0.3/0.0/0.0)
saxophone	16	3.3 (0.0/1.8/1.5/x)	22.5	0.6 (0/0.3/x/0.3)	17	0.3 (0.0/0.3/x/0.3)
penant	17.5	3.9 (x/1.5/0.9/1.5)	39	1.2 (x/0.9/0.0/0.3)	30.5	0.6 (x/0.6/0.0/0.0)
toaster	17.5	3.9 (1.5/x/1.5/0.9)	14	0.3 (0.0/x/0.0/0.3)	17	0.3 (0.0/x/0.0/0.3)
sewing machine	19	4.1 (0.9/x/2.6/0.6)	1	0.0 (x/0.0/0.0/0.0)	40	0.9 (x/0.0/0.0/0.9)
rose	20	4.4 (x/3.2/0.9/0.3)	39	1.2 (0.0/1.2/x/0.0)	37	0.9 (0.3/0.3/x/0.3)
scissors	21.5	4.9 (2.3/x/0.6/2.0)	22.5	0.6 (0.6/0.0/0.0/x)	30.5	0.6 (0.6/0.0/x/0.0)
suitcase	21.5	4.9 (x/0.0/2.0/2.9)	39	1.2 (0.0/0.6/x/0.6)	45	2.0 (0.0/0.0/x/0.2)
fishing net	23.5	5.0 (0.9/1.2/2.0/x)	1	0.0 (x/0.0/0.0/0.0)	5	0.0 (0.0/x/0.0/0.0)
wrench	23.5	5.0 (0.6/2.6/x/1.8)	1	0.0 (0.0/x/0.0/0.0)	17	0.3 (0.0/x/0.3/0.0)

ITEM	TRIAL 1		TRIAL 2		DELAY TRIAL	
	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION
paintbrush	25	5.3 (1.8/2.9/x/0.6)	36	1.2 (0.03/x/0.0/0.9)	30.5	0.6 (0.3/0.0/0.3/x)
life preserver	26.5	5.5 (x/2.9/0.3/2.3)	48	2.7 (0.3/x/1.8/0.6)	40	0.9 (0.0/x/0.9/0.0)
pail	26.5	5.5 (x/2.6/2.9/0.0)	47	1.9 (0.0/0.0/1.5/x)	5	0.9 (0.0/0.0/0.0/x)
wheelbarrow	28	5.6 (x/4.4/0.9/0.3)	22.5	0.3 (x/0.0/0.0/0.6)	17	0.3 (x/0.3/0.0/0.0)
mouse	29	5.9 (1.5/1.2/3.2/x)	22.5	0.3 (0.0/x/0.0/0.6)	30.5	0.6 (0.0/x/0.3/0.3)
lightbulb	30	6.2 (0.3/4.7/1.2/x)	1	0.0 (0.0/0.0/0.0/x)	48	3.5 (0.3/x/3.2/0.0)
birthday cake	31.5	6.7 (6.4/0.0/0.3/x)	31	0.9 (0.3/0.3/0.3/x)	30.5	0.6 (0.0/0.0/x/0.6)
pencil	31.5	6.7 (x/3.5/1.2/2.0)	39	1.2 (0.6/0.0/0.6/x)	30.5	0.6 (0.3/0.0/0.3/x)
stool	33	7.0 (3.8/1.2/2.0/x)	39	1.2 (x/0.9/0.0/0.3)	5	0.0 (0.0/0.0/x/0.0)
candle	34	7.1 (1.8/0.6/x/4.7)	39	1.2 (x/0.3/0.3/0.6)	17	0.3 (0.0/x/0.3/0.0)
umbrella	35	7.3 (3.2/1.5/x/2.6)	39	1.2 (0.0/0.6/x/0.6)	40	0.9 (0.3/0.0/x/0.6)
football	36	7.6 (3.2/3.2/1.2/x)	31	0.9 (0.0/0.9/0.0/x)	44	1.5 (0.6/0.6/x/0.3)

ITEM	TRIAL 1		TRIAL 2		DELAY TRIAL	
	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION
carrot	37	8.0 (1.2 / 5.3 / 1.5 / x)	22.5	0.6 (x / 0.0 / 0.3 / 0.3)	30.5	0.6 (x / 0.0 / 0.3 / 0.3)
stepstool	38.5	8.5 (1.5 / x / 1.2 / 5.8)	14	0.3 (0.0 / x / 0.3 / 0.0)	30.5	0.6 (0.6 / x / 0.0 / 0.0)
motorcycle	38.5	8.5 (x / 2.9 / 5.0 / 0.6)	1	0.0 (0.0 / 0.0 / 0.0 / x)	5	0.0 (0.0 / 0.0 / 0.0 / x)
leaf	40	9.5 (3.8 / 4.1 / x / 2.0)	1	0.0 (0.0 / 0.0 / x / 0.0)	17	0.3 (0.0 / 0.3 / x / 0.0)
ice cream cone	41	10.5 (3.2 / x / 3.5 / 3.8)	1	0.0 (0.0 / 0.0 / x / 0.0)	17	0.3 (0.0 / 0.3 / x / 0.0)
iron	42	10.6 (8.5 / 1.8 / x / 0.3)	14	0.3 (0.0 / 0.0 / x / 0.3)	30.5	0.6 (0.0 / 0.3 / x / 0.3)
jack-in-the-box	43.5	11.1 (5.3 / 2.3 / x / 3.5)	31	0.9 (0.6 / 0.0 / x / 0.3)	17	0.3 (0.0 / 0.3 / x / 0.0)
stapler	43.5	11.1 (1.2 / 5.8 / x / 4.1)	14	0.3 (0.0 / 0.0 / 0.3 / x)	5	0.0 (0.0 / 0.0 / 0.0 / x)
roller skates	45	11.7 (2.0 / 8.2 / 1.5 / x)	22.5	0.6 (0.0 / 0.0 / x / 0.6)	30.5	0.6 (0.0 / x / 0.3 / 0.3)
shopping cart	46	12.0 (4.4 / 6.4 / x / 1.2)	22.5	0.6 (0.0 / 0.0 / 0.6 / x)	40	0.9 (0.0 / 0.6 / 0.3 / x)
birdhouse	47	12.6 (x / 8.2 / 2.6 / 1.8)	31	0.9 (0.3 / 0.3 / 0.3 / x)	46	2.1 (x / 1.8 / 0.3 / 0.0)
duck	48	17.5 (14.6 / 2.3 / 0.6 / x)	31	0.9 (0.0 / 0.0 / 0.9 / x)	40	0.9 (0.0 / 0.0 / 0.9 / x)

Appendix H.02 Recognition Memory for Words

ITEM	TRIAL 1		TRIAL 2		DELAY TRIAL	
	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION
calculator	1	9.4 (4.7/x/4.1/0.6)	4	1.2 (0.3/x/0.9/0.0)	3.5	1.8 (0.6/x/0.0/1.2)
brain	2	9.5 (5.3/2.1/x/2.1)	11	2.4 (x/0.0/1.2/1.2)	11.5	3.0 (x/0.3/0.3/2.4)
fossil	3	11.2 (0.6/3.2/7.4/x)	19.5	4.1 (0.6/2.6/x/0.9)	19.5	4.2 (0.6/1.2/2.4/x)
doughnut	4	11.4 (3.5/4.1/x/3.8)	13.5	3.0 (0.9/x/0.9/1.2)	11.5	3.0 (1.8/x/0.3/0.9)
harmonica	5	11.5 (3.2/2.1/6.2/x)	28	5.6 (0.0/x/2.4/3.2)	29.5	6.2 (5.9/x/0.3/0.0)
pizza	6	11.7 (5.0/3.2/3.5/x)	5.5	1.5 (0.0/1.2/0.3/x)	1.5	1.2 (0.3/0.3/0.6/x)
raft	7	13.8 (x/2.6/1.8/9.4)	35.5	7.4 (6.2/0.9/x/0.3)	33	6.8 (4.4/0.3/x/2.1)
flea	8	15.0 (2.1/4.1/8.8/x)	32	6.5 (x/2.4/3.8/0.3)	34	7.0 (1.5/2.9/x/2.6)
camera	9	15.2 (2.9/3.8/x/8.5)	37	7.6 (2.6/0.6/x/4.4)	19.5	4.2 (0.9/2.4/0.9/x)
blouse	10.5	15.3 (x/8.8/1.2/5.3)	47.5	16.8 (0.9/10.3/x/5.6)	6	2.1 (x/0.3/0.3/1.5)
valley	10.5	15.3 (x/4.4/5.9/5.0)	7.5	1.8 (0.9/0.3/x/0.6)	19.5	4.2 (1.2/0.9/x/2.1)
alligator	12	15.9 (7.4/5.3/x/3.2)	5.5	1.5 (0.0/x/0.9/0.6)	11.5	3.0 (0.9/x/1.8/0.3)

Appendix H.02 (cont'd.)

ITEM	TRIAL 1		TRIAL 2		DELAY TRIAL	
	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION
trapeze	13	16.1 (8.2 / 5.3 / x / 2.6)	24	5.0 (0.6 / x / 0.9 / 3.5)	25.5	5.1 (2.1 / x / 0.6 / 2.4)
lobster	14	16.2 (5.9 / x / 4.7 / 5.6)	11	2.4 (x / 0.9 / 0.9 / 0.6)	8	2.4 (x / 1.5 / 0.0 / 0.9)
walnut	15	16.4 (5.3 / 8.2 / x / 2.9)	13.5	3.0 (0.9 / 1.2 / x / 0.9)	29.5	6.2 (0.3 / 3.5 / 2.4 / x)
mattress	16	17.4 (x / 2.1 / 3.2 / 12.1)	18	3.9 (1.2 / 0.6 / 2.1 / x)	14.5	3.3 (1.5 / 0.6 / 1.2 / x)
elevator	17	17.6 (0.6 / 7.6 / x / 9.4)	17	3.5 (0.6 / 0.3 / x / 2.6)	6	2.1 (0.3 / 1.5 / 7.4 / x)
bench	18	17.7 (11.5 / 4.7 / 1.5 / x)	35.5	7.4 (4.4 / x / 1.2 / 1.8)	24	5.0 (0.6 / X / 2.6 / 1.8)
dancer	19.5	18.9 (x / 5.9 / 6.2 / 6.8)	42	9.7 (4.1 / x / 0.3 / 5.3)	17	4.1 (2.9 / 0.9 / 0.3 / x)
saddle	19.5	18.9 (5.3 / x / 7.0 / 6.5)	25.5	5.1 (2.1 / 0.9 / 2.1 / x)	16	3.6 (2.1 / 1.2 / x / 0.3)
garage	21	19.3 (8.5 / 2.6 / x / 8.5)	2.5	0.9 (0.3 / 0.3 / x / 0.3)	1.5	1.2 (0.0 / 0.3 / x / 0.9)
helmet	23	20.0 (x / 0.6 / 5.0 / 14.4)	15.5	3.3 (1.5 / 0.6 / 1.2 / x)	31	6.6 (2.4 / 2.1 / x / 2.1)
magnet	23	20.0 (x / 1.5 / 10.9 / 7.6)	33.5	7.3 (x / 0.3 / 4.1 / 2.9)	27	5.9 (2.4 / 0.0 / 3.5 / x)
village	23	20.0 (7.6 / x / 7.1 / 5.3)	19.5	4.1 (0.9 / 0.6 / x / 2.6)	22	4.5 (x / 2.4 / 0.9 / 1.2)

Visual Recognition Memory . . . 290

Appendix H.02 (cont'd.)

ITEM	TRIAL 1		TRIAL 2		DELAY TRIAL	
	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION
ambulance	25	20.3 (2.6 / 11.5 / x / 6.2)	38	7.7 (2.4 / 0.3 / x / 5.0)	6	2.1 (0.9 / 0.3 / x / 0.9)
fort	27	20.9 (0.9 / 12.6 / x / 7.4)	47.5	16.8 (0.9 / 10.3 / x / 5.6)	35.5	7.1 (0.9 / 5.3 / x / 0.9)
hose	27	20.9 (x / 7.6 / 12.1 / 1.2)	27	5.4 (0.9 / 2.4 / 2.1 / x)	35.5	7.1 (x / 1.8 / 4.1 / 1.2)
orchard	27	20.9 (x / 5.9 / 2.1 / 12.9)	41	8.6 (x / 6.2 / 2.4 / 0.0)	14.5	3.3 (x / 2.1 / 1.2 / 0.0)
factory	29	21.0 (2.6 / x / 2.9 / 15.6)	11	2.4 (x / 0.3 / 1.5 / 0.6)	39	8.5 (x / 3.2 / 3.8 / 1.5)
calendar	30	22.9 (8.2 / 6.2 / 8.5 / x)	2.5	0.9 (0.3 / 0.0 / 0.6 / x)	3.5	1.8 (0.3 / 0.3 / 1.2 / x)
pancake	31	23.6 (2.4 / 11.5 / 9.7 / x)	22.5	4.8 (x / 0.9 / 2.4 / 1.5)	11.5	3.0 (0.3 / x / 1.5 / 1.2)
burlap	32	24.0 (x / 8.5 / 7.6 / 7.9)	22.5	4.8 (1.5 / x / 1.8 / 1.5)	19.5	4.2 (1.5 / x / 1.8 / 0.9)
shield	33	26.2 (9.1 / x / 2.4 / 14.7)	31	6.2 (x / 0.3 / 5.9 / 0.0)	45	11.8 (X / 5.9 / 2.4 / 3.5)
rock	34	26.4 (8.2 / 5.0 / 13.2 / x)	15.5	3.3 (1.5 / 1.8 / x / 0.0)	40	8.5 (0.9 / 5.0 / x / 2.6)
flour	35	26.7 (x / 13.8 / 3.5 / 9.4)	40	8.2 (0.3 / 4.1 / 3.8 / x)	38	7.7 (x / 4.4 / 1.5 / 1.8)
chimney	36	27.4 (12.1 / 1.8 / x / 13.5)	1	0.0 (0.0 / x / 0.0 / 0.0)	28	6.1 (0.6 / x / 2.9 / 2.6)

Visual Recognition Memory . . . 291

Appendix H.02 (cont'd.)

ITEM	TRIAL 1		TRIAL 2		DELAY TRIAL	
	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION
forest	37	29.5 (10.9 / 1.8 / 16.8 / x)	7.5	1.8 (0.0 / x / 0.6 / 1.2)	32	6.7 (4.1 / x / 2.6 / 0.0)
bacon	38	30.0 (4.7 / 17.9 / 7.4 / x)	9	2.1 (x / 1.5 / 0.3 / 0.3)	9	2.7 (x / 0.0 / 0.9 / 1.8)
monkey	39.5	31.5 (1.2 / x / 13.5 / 16.8)	30	5.9 (2.9 / x / 0.9 / 2.1)	47	13.8 (2.9 / x / 0.9 / 10.0)
swan	39.5	31.5 (x / 5.6 / 15.3 / 10.6)	21	4.4 (0.9 / 2.9 / 0.6 / x)	40	8.5 (2.4 / 3.5 / 2.6 / x)
graph	41	32.1 (13.2 / 1.8 / x / 17.1)	44	11.7 (x / 2.6 / 2.9 / 6.2)	43.5	10.0 (x / 4.7 / 0.9 / 4.4)
tire	42	33.8 (10.0 / x / 16.2 / 7.9)	43	10.6 (1.5 / x / 1.2 / 7.9)	42	9.1 (3.8 / x / 1.8 / 3.5)
ribbon	43	34.0 (7.6 / x / 13.8 / 12.6)	45	13.3 (1.5 / 1.8 / 10.0 / x)	48	19.8 (2.1 / 1.5 / x / 16.2)
grass	44	37.1 (18.8 / 2.4 / 15.9 / x)	25.5	5.1 (0.9 / 1.8 / x / 2.4)	25.5	5.1 (1.8 / 1.5 / x / 1.8)
spider	45	38.4 (6.5 / 10.0 / x / 20.9)	33.5	7.3 (0.3 / 2.6 / 4.4 / x)	43.5	10.0 (0.6 / 4.4 / 5.0 / x)
dollar	46	39.1 (12.6 / 2.4 / 24.1 / x)	29	5.7 (2.1 / 1.2 / 2.4 / x)	23	4.7 (0.6 / 2.6 / 1.5 / x)
muzzle	47	41.2 (21.5 / x / 7.1 / 12.6)	39	7.9 (x / 0.9 / 4.1 / 2.9)	37	7.4 (x / 5.0 / 1.2 / 1.2)
tooth	48	51.8 (15.6 / x / 5.6 / 30.6)	46	16.1 (4.4 / 3.8 / 7.9 / x)	46	13.6 (1.2 / 5.0 / 7.4 / x)

Appendix II.03 Recognition Memory for Words

ITEM	TRIAL 1 RANK ORDER TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION	TRIAL 2 RANK ORDER TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION	DELAY TRIAL RANK ORDER TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION
1 10.6 (2.4 / 4.1 / 4.1 / x)	10.5 8.9 (3.0 / 5.0 / 0.9 / x)	15 12.7 (x / 2.1 / 4.7 / 5.9)
2 12.4 (x 2.4 / 5.3 / 4.7)	5 6.6 (1.5 / 1.8 / 3.3 / x)	5 7.4 (x / 3.3 / 0.3 / 3.8)
3 13.6 (6.8 / 6.5 / x / 0.3)	8 8.6 (2.4 / 2.4 / 3.8 / x)	8 8.4 (x / 2.1 / 3.0 / 3.3)
4 14.9 (3.0 / x / 1.5 / 10.4)	16 14 (3.0 / 7.1 / x / 4.4)	11 9.5 (3.0 / 1.2 / 5.3 / x)
5 18.7 (0.9 / 5.3 / x / 12.4)	2 5.4 (x / 2.1 / 0.3 / 3.0)	7 8.3 (4.4 / x / 1.2 / 2.7)
6 17.5 (3.0 / x / 9.8 / 4.7)	13 9.8 (x / 7.4 / 0.9 / 1.5)	1 2.1 (0.6 / 0.3 / 1.2 / x)

Appendix II.03 (cont'd.)

ITEM	TRIAL 1		TRIAL 2		DELAY TRIAL	
	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION
7	20.0 (4.1 / 11.5 / x / 4.4)	7	8.3 (0.0 / 5.0 / 3.3 / x)	2	3.6 (0.6 / 0.6 / x / 2.4)
8	21.9 (3.3 / 14.8 / x / 3.8)	4	6.8 (x / 0.6 / 4.4 / 1.8)	9.5	9.2 (5.3 / 3.0 / x / 0.9)
9	24.5 (x / 3.8 / 14.8 / 5.9)	15	10.3 (5.9 / x / 0.3 / 4.1)	6	8.0 (5.3 / 0.6 / x / 2.1)
10	24.8 (15.1 / 5.0 / 4.7 / x)	10	8.9 (2.1 / 2.4 / x / 4.4)	14	11.3 (1.5 / 6.5 / x / 3.3)
11	25.5 (15.1 / x / 7.4 / 3.0)	1	4.5 (0.9 / 0.6 / x / 3.0)	3	4.2 (x / 0.6 / 3.3 / 0.3)
12	28.1 (11.2 / 10.1 / x / 6.8)	5	7.4 (0.6 / 5.3 / x / 1.5)	12.5	10.4 (0.3 / 8.9 / x / 1.2)

Appendix H.03 (cont'd.)

ITEM	TRIAL 1		TRIAL 2		DELAY TRIAL	
	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION
13	29.9 (4.1 / 22.2 / x / 3.6)	24	27.2 (0.6 / x / 4.4 / 22.2)	23	28.8 (9.8 / x / 16.9 / 2.1)
14	30.5 (5.0 / x / 21.9 / 3.6)	14	10.1 (6.5 / 2.1 / 1.5 / x)	16	13.3 (1.5 / x / 7.4 / 4.4)
15	34.6 (13.6 / 15.1 / 5.9 / x)	17	14.4 (6.2 / 3.8 / x / 4.4)	18	16.3 (3.6 / 11.8 / 0.9 / x)
16	34.9 (10.9 / x / 7.7 / 16.3)	18	16.0 (9.5 / x / 2.1 / 4.4)	17	15.3 (x / 6.2 / 3.8 / 5.3)
17	35.5 (7.7 / x / 5.6 / 22.2)	14	16.6 (3.0 / x / 11.5 / 2.1)	22	27.5 (8.3 / 13.3 / x / 5.9)
18	38.2 (x / 5.3 / 31.1 / 1.8)	23	24.0 (6.2 / 1.5 / x / 16.3)	9.5	9.2 (5.0 / 0.9 / 3.3 / x)

Appendix II.03 (cont'd.)

ITEM	TRIAL 1		TRIAL 2		DELAY TRIAL	
	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION	RANK ORDER	TOTAL PERCENT WRONG AND PERCENT WRONG BY LOCATION
19	39.9 (26.6 / 8.9 / 4.4 / x)	21	19.2 (4.7 / x / 15.5 / 13.0)	4	4.5 (0.3 / x / 2.4 / 1.8)
20	40.8 (x / 1.8 / 11.8 / 27.2)	6	8.1 (x / 2.4 / 2.1 / 3.6)	12.5	10.4 (4.7 / x / 3.3 / 2.4)
21	47.0 (8.9 / 19.2 / 18.9 / x)	23	20.4 (x / 11.2 / 6.5 / 2.7)	20	21.9 (13.0 / x / 7.1 / 1.8)
22	50.0 (x / 13.3 / 17.5 / 19.2)	9	8.7 (x / 3.3 / 2.1 / 3.3)	21	23.7 (9.5 / 6.8 / 7.4 / x)
23	55.3 (x / 15.4 / 14.5 / 25.4)	20	18.6 / 0.9 / x / 11.5 / 6.2)	19	21.0 (6.2 / 8.0 / 6.8 / x)
24	56.5 (32.0 / 13.6 / x / 10.9)	12	9.5 (5.6 / 1.8 / 2.1 / x)	24	35.6 (x / 2.1 / 1.8 / 31.7)

Visual Recognition Memory 29

Appendix I

Data for 'High' and 'Low' Depression Subjects

Appendix 1.01 Demographic Information for Paired Sample of 'High' and 'Low' Depression Subjects (n=48)

Demographic Information	Scores on Geriatric Depression Scale	
	'Low' Depression	'High' Depression
Age		
Mean	28.1	28.6
(s.d.)	(11.5)	(11.1)
Education		
Mean	12.6	12.5
(s.d.)	(1.7)	(1.9)
Vocab		
Mean	46.9	46.8
(s.d.)	(13.8)	(11.1)
MMSE		
Mean	28.8	28.8
(s.d.)	(0.9)	(1.4)
GDS		
Mean	5.8	18.3
(s.d.)	(2.9)	(4.0)
Gender		
Males	10 (41.7%)	10 (41.7%)
Females	14 (58.3%)	14 (58.3%)

Note. Education is reported in number of years; Vocab = raw score on Vocabulary subtest from Wechsler Adult Intelligence Scale-Revised (maximum score = 70); MMSE = Mini-Mental State Examination (maximum score = 30); GDS = Geriatric Depression Scale (maximum score = 30); Gender is reported by the number of participants as well as the percent. The cut-off point between high and low depression was between 13 and 14 on the GDS.

Appendix I.02 Differences on G-LAMB Scores Between 'High' and 'Low' Depression Groups

MEANS AND STANDARD DEVIATIONS				
G-LAMB MEASURES	'Low' Depression (GDS Score below 14)	'High' Depression (GDS Score 14 and above)	<i>F</i> ratio (<i>& probability</i>)	
PARAGRAPH				
Trial 1	21.04 ^a (5.27) ²	22.58 (5.59)	0.97	(0.33)
Last Learning	28.38 (2.60)	29.25 (1.82)	1.82	(0.18)
Trial	49.42 (7.44)	51.83 (6.73)	1.39	(0.24)
Total Acquisition	27.58 (2.78)	27.58 (2.73)	0.00	(1.00)
Delay Memory				
WORD LIST				
Trial 1	6.79 (2.30)	7.00 (1.56)	0.13	(0.72)
Last Learning	14.08 (1.14)	13.83 (1.27)	0.51	(0.48)
Trial	30.21 (21.84)	36.83 (23.99)	1.00	(0.32)
Total Acquisition	14.58 (0.78)	14.29 (1.23)	0.96	(0.33)
Delay Memory				
RECTANGLE FIGURE				
Trial 1	18.96 (5.12)	19.42 (2.84)	0.15	(0.70)
Last Learning	25.00 (1.35)	24.52 (1.77)	1.09	(0.30)
Trial	66.63 (8.39)	67.94 (6.07)	0.38	(0.54)
Total Acquisition	25.07 (1.13)	24.38 (1.74)	2.58	(0.12)
Delay Memory				

* $p < .05$; ** $p < .01$; *** $p < .001$

^a First number represents mean of scores from that group

^b Number in parenthesis represents standard deviation of scores for that measures from that group

Note. Highest possible scores are: Paragraph - 31; Word List - 15; Rectangle Figure - 26.

Appendix I.03 Means and Standard Deviations on Neuropsychological Measures for Paired Sample of 'High' and 'Low' Depression Subjects

Measure:	Geriatric Depression Scale Score	
	'Low' Depression	'High' Depression
Word Fluency		
<i>Mean</i>	60.33	65.71
<i>(s.d.)</i>	(14.0)	(12.4)
Trails A		
<i>Mean</i>	29.85	25.65
<i>(s.d.)</i>	(11.3)	(7.3)
Trails B		
<i>Mean</i>	58.54	53.57
<i>(s.d.)</i>	(17.6)	(17.3)
Boston Naming Test		
<i>Mean</i>	55.63	52.13
<i>(s.d.)</i>	(3.6)	(6.8)

Note. Word Fluency is the total number of words produced in 1 minute each for the F, A, S, and Animals categories; Trails A and Trails B are presented as the number of seconds taken to complete each task; scores on the Boston Naming Test represent the number of correct answers provided without any cues.