1

PM-1 3 1/2"x4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

PRECISIONSM RESOLUTION TARGETS

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 :

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Geriatric Learning and Memory Battery: Test Development,
Psychometric Evidence, and Normative Data

by

Anita M. Hubley, M.A.

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

Doctor of Philosophy

Department of Psychology

Carleton University
Ottawa, Ontario
August, 1995

copyright
1995, Anita M. Hubley

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 0-612-08845-6

Canada

Name Bruce M. Hickey

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.

Tests and Measurements

0288

U·M·I

SUBJECT TERM

SUBJECT CODE

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
Finance 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 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
General 0289
Linguistics 0290
Modern 0291
Literature 0401
General 0294
Classical 0295
Comparative 0297
Medieval 0298
Modern 0316
African 0591
American 0305
Asian 0352
Canadian (English) 0355
Canadian (French) 0593
English 0311
Germanic 0312
Latin American 0315
Middle Eastern 0313
Romance 0314
Slavic and East European

PHILOSOPHY, RELIGION AND THEOLOGY

Philosophy 0422
Religion 0316
General 0321
Biblical Studies 0319
Clergy 0320
History of 0322
Philosophy of 0469
Theology

SOCIAL SCIENCES

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

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 Relations 0617
Public Administration 0814
Recreation 0452
Social Work 0626
Sociology 0627
General 0938
Criminology and Penology 0631
Demography 0628
Ethnic and Racial Studies 0629
Individual and Family Studies 0630
Industrial and Labor Relations 0700
Public and Social Welfare 0344
Social Structure and Development 0709
Theory and Methods 0999
Transportation 0453
Urban and Regional Planning 0453
Women's Studies

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES

Agriculture 0473
General 0285
Agronomy 0475
Animal Culture and Nutrition 0476
Animal Pathology 0359
Food Science and Technology 0478
Forestry and Wildlife 0479
Plant Culture 0480
Plant Pathology 0817
Plant Physiology 0777
Range Management 0746
Wood Technology

Biology 0306
General 0287
Anatomy 0308
Biostatistics 0309
Botany 0379
Cell 0329
Ecology 0353
Entomology 0369
Genetics 0793
Limnology 0410
Microbiology 0307
Molecular 0317
Neuroscience 0416
Oceanography 0433
Physiology 0821
Radiation 0778
Veterinary Science 0472
Zoology

Biophysics 0786
General 0760
Medical

EARTH SCIENCES

Biogeochemistry 0425
Geochemistry 0996

Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleocology 0426
Paleontology 0418
Paleozoology 0985
Palynology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES

Environmental Sciences 0768
Health Sciences 0566
General 0300
Audiology 0992
Chemotherapy 0567
Dentistry 0350
Education 0769
Hospital Management 0758
Human Development 0982
Immunology 0564
Medicine and Surgery 0347
Mental Health 0569
Nursing 0570
Nutrition 0380
Obstetrics and Gynecology 0354
Occupational Health and Therapy 0381
Ophthalmology 0571
Pathology 0419
Pharmacology 0572
Pharmacy 0382
Physical Therapy 0573
Public Health 0574
Radiology 0575
Recreation

Speech Pathology 0460
Toxicology 0383
Home Economics 0386

PHYSICAL SCIENCES

Pure Sciences 0485
Chemistry 0749
General 0486
Agricultural 0487
Analytical 0488
Biochemistry 0738
Inorganic 0490
Nuclear 0491
Organic 0494
Pharmaceutical 0495
Physical 0754
Polymer 0405
Radiation 0605
Mathematics 0986
Physics 0606
General 0608
Acoustics 0748
Astronomy and Astrophysics 0607
Elementary Particles and High Energy 0798
Fluid and Plasma 0759
Molecular 0609
Nuclear 0610
Optics 0752
Radiation 0756
Solid State 0611
Statistics 0463

Applied Sciences 0346
Applied Mechanics 0984
Computer Science

Engineering 0537
General 0538
Aerospace 0539
Automotive 0540
Agricultural 0541
Biomedical 0542
Chemical 0543
Civil 0544
Electronics and Electrical 0348
Heat and Thermodynamics 0545
Hydraulic 0546
Industrial 0547
Marine 0794
Materials Science 0548
Mechanical 0743
Metallurgy 0551
Mining 0552
Nuclear 0549
Packaging 0765
Petroleum 0554
Sanitary and Municipal 0790
System Science 0428
Geotechnology 0796
Operations Research 0795
Plastics Technology 0994
Textile Technology

PSYCHOLOGY

General 0621
Behavioral 0384
Clinical 0622
Developmental 0620
Experimental 0623
Industrial 0624
Personality 0625
Physiological 0989
Psychobiology 0349
Psychometrics 0632
Social 0451

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

GERIATRIC LEARNING AND MEMORY BATTERY:
TEST DEVELOPMENT, PSYCHOMETRIC EVIDENCE, AND NORMATIVE DATA

submitted by

Anita M. Hubley, B.A., M.A.

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

Chair, Department of Psychology

Thesis Supervisor

External Examiner

Carleton University
August 1995

Abstract

Two memory tests, the Diagnostic Test and the Tracking Test, are developed and together comprise the Geriatric Learning and Memory Battery. There are two main reasons for assessing memory performance in older adults. The first reason, addressed by the Diagnostic Test, is the diagnosis of memory impairment. The second reason, addressed by the Tracking Test, is to examine or track the learning and memory performance of individuals already diagnosed as suffering cognitive impairment. Problems with many of the currently available memory tests for use with older adults include materials that may be too difficult, tests that are too long, and procedures that do not take into account current knowledge about memory. Preliminary evidence for the reliability and validity of the test inferences and normative data based on a sample of community-dwelling older adults are presented for both the Diagnostic and Tracking Tests.

Acknowledgements

First, to Tom Tombaugh - my advisor, my mentor, my friend - I struggle to find the right words to express my appreciation for the support and guidance you have given me throughout my academic training. You have always had my best interests at heart and I have always been able to count on you to see the forest when I have gotten lost in the trees. The two words that I think best characterize our working relationship are "respect" and "trust". I have learned much from you over the years and my debt is great.

Secondly, to my husband and dearest friend Bruno Zumbo - I thank you for all of your love and patience throughout this process. I feel very fortunate to share a life with such a special person and to know that I can always count on your support and love for any endeavor that I undertake. I also wish to acknowledge the technical and statistical assistance that you provided.

There are a number of individuals who I would also like to acknowledge. First, I would like to thank my examining committee for their interest and encouragement: Dr. Jean Kozak (External Examiner), Dr. Roland Thomas (Internal Examiner), Dr. Bill Petrusic and Dr. Bruce Pappas (Committee Members). Second, there are a number of people who assisted me in these studies by recruiting volunteers, making bookings, and/or testing participants: Melissa Burgess, Dorothy Connor, Pat Cross, Lara Dickinson, Stephanie

Greenham, Dawn Hamilton, Karen Hicks, Karen Irving, Martha Jennings, Christine Kenney, Marilyn Keyes, Lisa McAvoy, Martha McDougall, Kate Oakley-McKeen, Laura Rees, Richard Simourd, Lianne Walker, and Marilee Zaharia.

I also wish to express my appreciation to those individuals from the community who volunteered their time to participate in this research. In addition, a number of community workers helped recruit participants: June Buck (CHEO), Lynn Bostrom (Queensway-Carleton), Grace Calvert (Grace General Hospital), Joan Cox (Ottawa Civic Hospital), Pat Coyne (RA Centre), Claire Guay (Ottawa General Hospital), Caroline Horgan (Elisabeth Bruyere Health Centre), and Laurel Pardy (Riverside Hospital).

I also appreciate the assistance of David Marso from Technical Support at SPSS. His provision of a computer program to calculate percentiles and cumulative percentiles made my task much easier.

A personal thank you goes to my friends Laura Rees and Debbie James for their support and friendship. I have fond memories of coffee and tea breaks, 3M sticky notes, fabric samples, Mike's Place and comprehensive exams, "wine lunches", long conversations, and the burbling coffeemaker. I also thank the many individuals at Carleton University and my colleagues at UNBC who supported and encouraged me.

And finally, I thank my parents, Jim and Marleen Hubley, for planting the dream that has come to pass...

Table of Contents

Acknowledgements.....	iii
Abstract.....	v
Table of Contents.....	vi
List of Tables.....	ix
Introduction.....	1
Purpose of Memory Assessment in Older Adults.....	5
Literature Review.....	8
Brief Review of Memory Stores and the Effect of Aging.....	8
Neuropsychology of Memory.....	13
Temporal Lobes.....	14
Diencephalon.....	16
Frontal Lobes.....	17
Tests of Memory Functioning.....	17
Specific Memory Tests.....	17
Comprehensive Memory Tests.....	20
Application of Theoretical and Clinical Knowledge.....	24
Use of Psychometric Criteria to Evaluate Tests of Memory.....	28
Geriatric Learning and Memory Battery (G-LAMB):	
Test Development.....	33
Diagnostic Test.....	33
Paragraph subtest.....	34
Geometric Figure subtest.....	42
Tracking Test.....	48
Paragraph subtest.....	49
Geometric Figure subtest.....	53
Method.....	56
Study 1.....	56
Goal.....	56
Subjects.....	56
Test Materials.....	57
Procedure.....	58
Study 2.....	61
Goal.....	61
Subjects.....	62
Test Materials.....	63
Procedure.....	64
Analyses.....	65
Results.....	72
Study 1.....	73
Demographic Information on Original Six Age Groups.....	73
Internal Consistency and Inter-rater	

Reliability.....	74
Factor Analyses.....	74
Regression Analyses.....	77
Demographic Information on Final	
Three Age Groups.....	79
Performance on Paragraph Subtest	
over Trials.....	82
Performance on Rectangle Figure	
Subtest over Trials.....	83
Normative Data.....	85
Study 2.....	88
Demographic Information on Original	
Six Age Groups.....	88
Internal Consistency and Inter-rater	
Reliability.....	90
Factor Analyses.....	90
Regression Analyses.....	93
Demographic Information on Final	
Three Age Groups.....	94
Performance on Paragraph Subtest	
over Trials.....	96
Performance on Triangle Figure	
Subtest over Trials.....	97
Normative Data.....	99
Discussion.....	101
1. Test Rationale.....	101
2. Scale Reliability.....	117
3. Validity.....	121
Age Effects.....	122
Test Procedures.....	125
Factor Analytic Findings.....	130
4. Normative Data.....	133
Summary.....	135
Limitations and Potential Difficulties.....	135
Future Research.....	140
Endnotes.....	143
References.....	145
Appendix A: Scoring System for the Rectangle	
Geometric Figure.....	233
Appendix B: Scoring System for the Triangle	
Geometric Figure.....	234
Appendix C: Brief Description of Previously	
Published Tests Used in Study 1	
and/or Study 2.....	235
Appendix D: Description of Unpublished Tests	
Used in Study 1 and/or Study 2.....	238
Appendix E: Ethics Approval.....	242
Appendix F: Study 1: Order of Test Administration....	243

Appendix G: Study 2: Order of Test Administration.....	244
Appendix H: Learning and Memory Battery (IAMB):	
Paragraph and Complex Figure.....	245

List of Tables

Table 1:	Demographic Information for Six Age Groups for Study 1.....	165
Table 2:	Correlations Among Demographic Variables in Study 1.....	166
Table 3:	Factor Analysis Pattern Matrix for Diagnostic Test.....	167
Table 4:	Factor Analysis Pattern and Correlation Matrix for Study 1.....	168
Table 5:	Summary of Stepwise Regressions in Study 1.....	169
Table 6:	Performance on Paragraph Over Six Age Groups for Study 1.....	170
Table 7:	Performance on Rectangle Over Six Age Groups for Study 1.....	171
Table 8:	Demographic Information for Three Age Groups in Study 1.....	172
Table 9:	Paragraph Performance for Three Age Groups in Study 1.....	173
Table 10:	Rectangle Figure Performance for Three Age Groups in Study 1.....	174
Table 11:	Standard Scores for the Paragraph and Rectangle Figure Subtest Total Acquisition Scores in Study 1.....	175
Table 12:	Standard Scores for the Paragraph and Rectangle Figure Subtest Delay Scores in Study 1.....	177
Table 13:	Cumulative Percentiles for Adults Aged 55-64 Years on the Paragraph Subtest Trial-by-Trial Scores in Study 1.....	179
Table 14:	Cumulative Percentiles for Adults Aged 65-74 Years on the Paragraph Subtest Trial-by-Trial Scores in Study 1.....	182
Table 15:	Cumulative Percentiles for Adults Aged 75-85 Years on the Paragraph Subtest Trial-by-Trial Scores in Study 1.....	185
Table 16:	Cumulative Percentiles for Adults Aged 55-64 Years on the Rectangle Figure Subtest Trial-by-Trial Scores in Study 1.....	188
Table 17:	Cumulative Percentiles for Adults Aged 65-74 Years on the Rectangle Figure Subtest Trial-by-Trial Scores in Study 1.....	191
Table 18:	Cumulative Percentiles for Adults Aged 75-85 Years on the Rectangle Figure Subtest Trial-by-Trial Scores in Study 1.....	194

Table 19:	Standard Error of Measurement (SEM) for Paragraph and Rectangle Figure Subtests in Study 1.....	197
Table 20:	Demographic Information for Six Age Groups for Study 2.....	198
Table 21:	Correlations Among Demographic Variables in Study 2.....	199
Table 22:	Factor Analysis Pattern Matrix for Tracking Test.....	200
Table 23:	Factor Analysis Pattern and Correlation Matrix for Study 2.....	201
Table 24:	Summary of Stepwise Regressions in Study 2.....	202
Table 25:	Performance on Paragraph Over Six Age Groups for Study 2.....	203
Table 26:	Performance on Triangle Over Six Age Groups for Study 2.....	204
Table 27:	Demographic Information for Three Age Groups in Study 2.....	205
Table 28:	Paragraph Performance for Three Age Groups in Study 2.....	206
Table 29:	Triangle Figure Performance for Three Age Groups in Study 2.....	207
Table 30:	Standard Scores for the Paragraph and Triangle Figure Subtest Total Acquisition Scores in Study 2.....	208
Table 31:	Standard Scores for the Paragraph and Triangle Figure Subtest Delay Scores in Study 2.....	210
Table 32:	Cumulative Percentiles for Adults Aged 55-64 Years on the Paragraph Subtest Trial-by-Trial Scores in Study 2.....	212
Table 33:	Cumulative Percentiles for Adults Aged 65-74 Years on the Paragraph Subtest Trial-by-Trial Scores in Study 2.....	215
Table 34:	Cumulative Percentiles for Adults Aged 75-85 Years on the Paragraph Subtest Trial-by-Trial Scores in Study 2.....	218
Table 35:	Cumulative Percentiles for Adults Aged 55-64 Years on the Triangle Figure Subtest Trial-by-Trial Scores in Study 2.....	221
Table 36:	Cumulative Percentiles for Adults Aged 65-74 Years on the Triangle Figure Subtest Trial-by-Trial Scores in Study 2.....	224
Table 37:	Cumulative Percentiles for Adults Aged 75-85 Years on the Triangle Figure Subtest Trial-by-Trial	

	Scores in Study 2.....	227
Table 38:	Standard Error of Measurement (SEM) for Paragraph and Triangle Figure Subtests in Study 2.....	230
Table 39:	Percent Recalled on the Diagnostic Test Paragraph and Geometric Figure by Final Three Age Groups.....	231
Table 40:	Percent Recalled on the Tracking Test Paragraph and Geometric Figure by Final Three Age Groups.....	232

Introduction

Ideally, a memory test should measure memory performance over the entire adult life span (i.e., 16 years through 80 years). It should distinguish normal memory performance from both superior and impaired performance, and it should be able to track an individual's performance over time, providing information about changes in mnemonic ability. Most memory tests currently used in clinical settings attempt to achieve these goals (e.g., Benton Visual Retention Test, California Auditory Verbal Learning Test, Fuld Object-Memory Evaluation, Learning and Memory Battery, Wechsler Memory Scale - Revised).

Individuals being assessed for memory problems usually are tested under standardized conditions. Everyone, regardless of age, is given the same instructions, the same study time, and the same time limits. Generally, this approach is desirable because it increases test reliability and validity. There are times, however, when such uniformity is less desirable and tends to decrease the validity of the test results. A prime example of this occurs when older adults are tested. Since older adults are slower in processing and learning new information (Salthouse, 1982), administering a standard memory test may seriously underestimate their memory abilities because of its failure to compensate for their slowed processing abilities. Failing to compensate for these normal age changes would be analogous to administering a test of

reading comprehension without allowing a "slow" reader enough time to read all of the text. If the person performs poorly on the test, you do not know whether the poor performance is due to impaired comprehension or reading speed. In a memory test, if an older adult is not given ample opportunity to learn as well as the younger adult, you do not know whether the older adult's poorer memory performance reflects impaired memory ability or simply less "opportunity" to acquire the material. Viewed in this way, currently available memory tests may contain an implicit bias against older adults.

Most memory tests are also designed so that younger individuals with above average to superior ability will find the test challenging. Psychometrically speaking, these tests are designed to have a high performance ceiling. One implication of this is that younger adults with average memory abilities will not readily master the material. For example, on the Learning and Memory Battery (LAMB; Tombaugh & Schmidt, 1992; Schmidt & Tombaugh, 1995) the average person in the 20-29 year age group recalls 46% of the paragraph on Trial 1, 80% on Trial 2, and 84% after a short delay. However, for the average older adult, performance is substantially lower. For example, the average person in the 70-74 year age group recalls only 21% of the paragraph on Trial 1, 52% on Trial 2, and 53% after a short delay. Although these percentages represent the "normal" levels of

performance for each age group, the average older adult clearly has less to feel optimistic about relative to the average younger adult.

The increased difficulty for older adults produces several problems. Individuals who obtain low scores or who do not feel as though they are performing well may become anxious or frustrated. This may have further detrimental effects by decreasing their motivation which, in turn, may cause them to expend minimal effort or become noncompliant. Further to this, if only a minimal amount of information is learned initially, then interpreting how well information has been retained over time (i.e., memory performance) becomes problematic. For example, one might conclude that if, after a 15 minute delay, a person only remembers two words out of a list of ten words, then the person has a poor memory. However, what if the person only learned two out of the ten words to begin with? Does this mean that the person is viewed as having a good memory because he or she demonstrated 100% retention? To the contrary, it is probably not even meaningful to talk about the person's memory ability when baseline levels of learning are so unreasonably low. Moreover, if performance is too low, the test cannot be used to monitor memory performance over time. This represents a serious psychometric shortcoming because the rate of cognitive decline is an important concern in the field of geriatrics and the ability to monitor changes in

the memory performance of impaired individuals represents an important dimension in the delivery of health care. Within this context, the most basic or common question concerns the rate of cognitive decline. However, there are a variety of other questions that also need to be addressed. For example, are some aspects of a person's memory declining whereas other aspects are unchanging or even "normal"? If you are dealing with someone with mild to moderate Alzheimer's disease, what are the deficits in their memory? Alternatively, what mnemonic strengths exist that can be used to help compensate or cope with the losses in memory? For example, does cuing in the form of asking questions about what the person needs to do next, or providing a short list, help?

As indicated above, many test developers attempt to develop the ideal memory test. However, it appears that too much may be expected from a single test or test battery. Many elderly would be better served with a test designed specifically for their age group. This would allow the focus to be directed on assessment and diagnostic issues that are specifically related to aging.

In older adults, the most common and perhaps obvious memory-related question is whether memory impairment exists, and if so, to what degree. This is contrasted to the type of assessment questions asked in younger age groups where more emphasis is directed towards identifying above average

to superior memory performance for purposes of academic or job placement. Practically-speaking, there is less interest in detecting superior memory performance in older adults and less need to discriminate among individuals who perform at the upper levels.

Purpose of Memory Assessment in Older Adults

Despite the large number of memory tests available today, one is hard-pressed to locate tests that specifically address the assessment needs of older adults. Overall, two goals need to be met in developing test materials that are appropriate for the assessment of memory performance in older adults. These two goals are: (a) diagnosis of memory impairment, and (b) evaluation or tracking of memory performance in individuals already diagnosed as cognitively impaired. Given these two distinct goals, it would seem reasonable that two tests, using different levels of test difficulty and/or test procedures, should be developed.

Diagnosis of Memory Impairment:

When the goal of a test is to diagnose or determine whether a memory impairment exists, procedures that best distinguish between normal and impaired memory performance are needed. Test materials should not be so difficult that they frustrate the individual being tested, but they should be difficult enough that the performance of cognitively intact and impaired individuals can be differentiated.

Tracking of Memory Performance:

When the goal of a test is to evaluate or track the memory performance of an impaired individual, procedures that provide individuals with increased support and the opportunity to utilize whatever memory strategies are available to them should be used. Test materials should be easier than those used for diagnosis so that a lower performance ceiling is obtained. Because one is attempting to assess a person's strengths as well as their deficits, it is important that feelings of frustration, discouragement, and anxiety be reduced so that maximum performance can be achieved.

The purpose of this dissertation is (a) to develop two sets of test materials for the assessment of memory in older adults, and (b) to provide normative data using a sample of community-dwelling adults aged 55 to 85 years. The first set of materials will be developed for the purpose of diagnosing memory impairment, whereas the second set of materials will be developed for the purpose of examining the performance of individuals with impaired memory. In each case, test materials and procedures appropriate to the goals will be used.

Before describing the development of the test materials, the relevant literature will be reviewed. This review will begin with a brief description of memory stores and processes as well as the age effects found. Next, the neuropsychology of memory will be briefly reviewed. This

will be followed by a description and critique of both specific and comprehensive tests of memory which represent the two main types of (or approaches to) memory assessment. Given that a major portion of this dissertation is devoted to the development of two tests of memory, several aspects of memory that are considered to be theoretically and clinically important to the assessment of memory functioning will be reviewed. Finally, four standard psychometric criteria that should be followed in any test development, and will ultimately be used to evaluate the tests developed here, will be described.

Review of the Literature

Brief Review of Memory Stores and the Effect of Aging

In this section, models or representations of learning and memory will be described as well as the normal changes that occur with age at each level. Many of the changes in memory that occur with age are related to the general slowing of both physical and cognitive performance found as people grow older (Syndulko & Tourtellotte, 1989; Van Gorp, Satz, & Mitrushina, 1990). For example, a progressive slowing in performance with age has been found on both physical and cognitive tests of reaction time (Birren, 1974; Gottsdanker, 1982; Wilkinson & Allison, 1989). In addition, both visual and auditory declines occur rapidly after age 50 (Claussen & Patel, 1990; Fozard, 1990; Hayes & Jerger, 1984; Owsley & Sloane, 1990).

Serial processing models typically identify three stages in the learning process: sensory memory, primary (or short-term) memory, and secondary (or long-term) memory. The term "sensory memory" was introduced by Atkinson and Shiffrin in 1968, and refers to a literal or exact sensory representation of an external stimulus. This representation only lasts about one second. Two types of sensory memory that have been proposed are iconic memory, which refers to a visual store (e.g., an icon or image; Sperling, 1960), and echoic memory, which refers to an auditory store (e.g., an echo; Massaro, 1970). Most or all of the information in

sensory memory is lost unless transferred to primary memory. Only small age differences appear to be present in sensory memory (Cerella, Poon, & Fozard, 1982; Davis & Bernstein, 1992). Older adults appear to need longer exposures to be able to register new stimuli (Craik, 1977; La Rue, 1992).

Both primary memory and secondary memory were first introduced by James (1890) and later described in a seminal paper by Waugh and Norman (1965). Primary memory is a temporary, limited-capacity store that is capable of holding approximately seven bits of information (i.e., 7 plus or minus 2 items). Such information remains in primary memory for about 10-30 seconds. If new information is introduced, the old information may be lost even more quickly. However, the contents of primary memory can be maintained through rehearsal. As in sensory memory, it has been suggested that primary memory also contains structures or systems related to visual and auditory information (i.e., the "visuo-spatial scratch pad" for visual information and the "articulatory loop" for speech; Baddeley, 1986). It appears that up to age 80, there are no age differences in the accuracy of recall across age groups (Craik, 1977; La Rue, 1992; Talland, 1968). In addition, there is little evidence to suggest that older adults are more susceptible to interference (in primary memory) than younger adults (Wickelgren, 1975). Older adults, however, do take longer to learn information (Anders, Fozard, & Lillyquist, 1972;

Craik, 1977; La Rue, 1992).

Secondary memory is assumed to be a relatively permanent store with unlimited capacity. There are three elements of memory to be considered: encoding, storage, and retrieval (Murdock, 1974). The first of these elements, *encoding*, refers to the modification or organization of items for storage and takes place during the process of consolidation (i.e., the transfer of information from primary to secondary memory). Information can be encoded by redefining and storing it as some type of meaningful representation. Thus, words that have some common element or relationship may be grouped together or "chunked". For example, in a word list the words "nurse", "lawyer", and "electrician" may be grouped together under the category of professions, or "green" and "spinach" may be linked together because spinach is green. Generally, the more chunking that can be accomplished, the greater the recall (Mandler, 1967).

Another way of encoding information involves elaboration or the processing of information at a deeper or more personal level. Thus, thinking of the word "nurse" as a profession or my niece's job will result in better memory storage than thinking of "nurse" as starting with "n" or as the third word in the list.

A third method of encoding involves imposing some structure on, or connection among, the information. For example, one might make a story out of the words in a list.

Thus, thinking of the "green tiger playing guitar in the basement for the nurse in the sweater" is easier than remembering the words "green", "tiger", "guitar", "basement", "nurse", and "sweater" individually.

Unlike younger adults, older adults are much less likely to spontaneously use verbal or visual associations or employ organizational strategies in learning material (Craik & Rabinowitz, 1984; Denney, 1974; Hulicka & Grossman, 1967). Moreover, it has been suggested that older adults engage in less extensive and less efficient encoding of information (Craik, 1977; Smith, 1980). As a result, older adults benefit more than younger adults when categorization and/or organization of materials is encouraged during the learning process (Earhard, 1977; Hultsch, 1971; Perlmutter & Mitchell, 1982; Schmitt, Murphy, & Sanders, 1981).

The second element of secondary memory, *storage*, is the process of retaining information that has been consolidated. Some decay or loss of information is expected in the first 24 hours following learning with very little forgetting after that. However, when material has been well-learned or overlearned, little or no forgetting occurs in the first 24 hours. This suggests that initial forgetting may be a result of poorer consolidation rather than a deterioration of the memory store itself. There is little evidence of age differences in storage (Craik, 1986).

The third element of secondary memory is *retrieval*,

which involves the process of recalling information from memory. One's ability to retrieve information can be measured using a retrieval continuum; that is, by examining an individual's free recall, cued recall, and recognition performance. Free recall involves recalling information on one's own without help. Cued recall involves the retrieval of information with the help of cues or hints. Recognition involves the retrieval of information by selecting the target item from among a number of other distractor items or "foils". Not surprisingly, being asked to recall the first word in a list (free recall) is more difficult than if one is reminded that the word is an animal (cued recall), which in turn is harder than if one is told that the word is either "camel", "tiger", "pig", or "lion".

Considerable research has found that older adults do not perform as well on free recall as younger adults, but were able to improve their performance when given cues (Petersen, Smith, Kokton, Ivnik, & Tangalos, 1992; Spinnler, Della Sala, Bandera, & Baddley, 1988). Craik (1986) and Craik and McDowd (1987) found that recognition memory exceeds free recall at all ages, although the magnitude of the difference is reliably greater for older adults than younger adults (Craik, 1991; Craik & McDowd, 1987; Kazniak, Poon, & Riege, 1986; West, 1986).

Finally, it must be noted that memory performance is affected by a number of factors. First, perceptual deficits

(e.g., hemi-neglect) will affect one's acquisition, and therefore one's recall, of information. Second, the ability to attend to information is important in the encoding process. Many individuals who complain of memory deficits actually have problems with attention or shifting their cognitive set rather than with memory. Third, motivation plays a role in both learning and memory. If a person is not interested in, or willing to try, a task, then his or her performance will be affected. It has been suggested that some older adults may lack the attention and motivation necessary to adequately encode and retrieve information (Camp & McKittrick, 1989; Zacks & Hasher, 1988). Finally, emotional factors such as anxiety or depression can also adversely affect one's ability to learn and concentrate.

Neuropsychology of Memory

More recently, attempts have been made to integrate cognitive models with our knowledge of brain structures, the result being an emergence of neuropsychological models of memory. These models may vary somewhat in their historical influences or terminology, but there is agreement that there are at least two broad classes of memory. I will refer to these as "explicit" and "implicit", although other terms have been used (e.g., declarative vs nondeclarative, memories vs habits). Consensus also exists that performance on explicit and implicit memory tests likely depends on different neural structures in the brain (Graf & Schacter,

1985; Moscovitch, 1984; Petri & Mishkin, 1994).

Because the tests of memory that are developed in this dissertation are tests of *explicit* memory, the focus here will be on the brain structures that are associated with this type of memory. Considerable emphasis has been placed on the role of the temporal lobes, the diencephalon, and the frontal lobes.

Temporal Lobes:

Within the temporal lobes, several areas have been suggested as the critical locus for damage and include the anterior temporal neocortex, the amygdala, the hippocampus, and the entorhinal cortex (Kolb & Whishaw, 1990). However, most evidence points to the hippocampus as the critical area.

Beginning in the 1950s, the importance of the hippocampus in memory was being clearly demonstrated in reports describing patients with bilateral hippocampal damage. One of these cases, H.M., has been the focus of hundreds of such studies (Kolb & Whishaw, 1990; Petri & Mishkin, 1994). In an attempt to prevent increasingly severe generalized epileptic seizures, H.M. agreed to an experimental procedure known as a bilateral medial temporal lobe resection. In this procedure, the hippocampus, amygdala, and underlying cortex were removed. Following the surgery, H.M. displayed a severe anterograde memory impairment. That is, H.M. was unable to form new memories

and showed impairment on verbal and nonverbal tests of explicit memory.

The case of H.M. has generated the most research examining the role of the medial temporal lobes and, in particular, the hippocampus in memory. However, it was the case of R.B. that demonstrated that it is possible to have severe anterograde amnesia resulting from bilateral damage to the hippocampus alone and that damage to additional medial temporal tissue would not have made his anterograde amnesia much worse (Squire, 1986; Kolb & Whishaw, 1990). R.B. suffered an ischemic episode during or shortly after a coronary bypass operation. As a result, he displayed severe anterograde amnesia as well as retrograde amnesia for the previous few years. Five years later, postmortem examination was unremarkable except for a bilateral loss of all cells in CA1 of the hippocampus.

Further research with individuals showing only unilateral hippocampal lesions has indicated that the left hippocampus can be functionally dissociated from the right (Bigler, 1988; Kolb & Whishaw, 1990). Work by Milner (1970) indicates that the left hippocampus appears to be more important for the memory of verbal material such as short stories and word lists whereas the right hippocampus seems to be important for the memory of visual and spatial material such as geometric drawings and faces.

Despite the importance of the hippocampus in memory,

its role is fairly circumscribed. That is, the role of the hippocampus seems to be to form new connections or associations between events (Jones-Gotman, 1986; Kesner, Hopkins, & Chiba, 1992; Squire, Shimamura, & Amaral, 1989). In addition, because individuals such as H.M. and R.B. exhibit severe anterograde amnesia and retrograde amnesia that is of limited duration, it has been suggested that the hippocampus is only transiently involved in the memory storage process and that other structures within the medial temporal lobes or elsewhere maintain the permanent memories (Kolb & Whishaw, 1990; Squire, 1992).

Diencephalon:

Three substructures of the diencephalon have been associated with amnesia; these are the dorsal medial nucleus, the group of anterior nuclei in the thalamus, and the mamillary body of the hypothalamus (Kolb & Whishaw, 1990). Evidence of diencephalic amnesia comes from two sources. The first source is patients with focal lesions of the medial thalamus, which are typically associated with vascular disease. The second, and most studied, source is Korsakoff's patients. Comparisons between diencephalic and temporal lobe patients have produced several distinctions between the two types of amnesia. For example, relative to patients with temporal lobe amnesia, diencephalic amnesics show normal rates of forgetting but extensive retrograde amnesia (Huppert & Piercy, 1978; Kolb & Whishaw, 1990;

Squire, 1982).

Frontal Lobes:

The role of the frontal lobes in memory appears to be one of organization and strategic processing (Moscovitch, 1992; Moscovitch & Umiltà, 1990). Thus, damage to the frontal lobes does not result in memory loss per se, but rather an impairment in organization at encoding (Incassa della Rochetta, 1986), search processes at retrieval (Delbecq-Derouesne, Beauvois, & Shallice, 1990; Mayes, 1988), and both contextual and temporal ordering (Moscovitch & Winocur, 1992). For example, frontal lobe patients tend to be susceptible to interference and show poor recall for memory involving temporal or spatial aspects (Kolb & Whishaw, 1990).

Tests of Memory Functioning

The study of memory and of memory dysfunction has led to the development of a number of memory tests. These memory tests basically can be divided into one of two types, either "specific" or "comprehensive". For each type of test, a general description will be provided and a representative test will be described.

Specific Memory Tests:

Specific memory tests focus on a specific type, or a single aspect, of memory such as verbal memory or visual memory. Specific tests of verbal memory include the California Verbal Learning Test (Delis, Kramer, Kaplan, &

Ober, 1987), the Fuld Object-Memory Evaluation (Fuld, 1981), the Rey Auditory Verbal Learning Test (Rey, 1964), and the Selective Reminding Test (Buschke & Fuld, 1974). Specific tests of visual memory include the Benton Visual Retention Test (Benton, 1974), the Biber Figure Learning Test (Glosser et al., 1989), the Rey-Osterrieth Complex Figure (Osterrieth, 1944; Rey, 1959), and the Taylor Complex Figure (Taylor, 1969).

The most commonly used specific memory test is the California Verbal Learning Test (CVLT). This test will be described here as a representative of this type of memory test. The CVLT consists of a shopping list of 16 words (List A) with four words taken from each of four semantic categories (fruits, tools, clothing, and spices/herbs). Immediate recall is tested over five acquisition trials. Next, a second shopping list (List B) is presented as an interference task. This list consists of 16 words taken from two of the same categories as List A (fruits and spices/herbs) and two different categories (utensils and fish). Immediate recall for List B is tested over a single acquisition trial. After this very short delay interval, the person first is asked to free recall List A and then is asked to recall the words in each of the four categories (cued recall). After a further delay of 20 minutes, the person is again asked to free recall List A and then recall the words in each category for cued recall. Finally, a

recognition trial is administered in which the person is read a list of 44 words and is asked to indicate whether or not the word was on List A. The foils on the recognition trial consist of either words from List B, words from neither list but from the same categories, words that were phonemically similar (e.g., "drums" rather than "plums"), or unrelated words.

Despite the valuable information that can be acquired from specific tests of memory, a major weakness of the available specific tests is the inconsistency in approaches and procedures used from one test to another. This is rather ironic given that an important advantage of these tests is that they permit an experimenter or clinician to choose the number of, and which, specific tests to use rather than be confined to those included in a comprehensive memory test. However, this ability to "mix and match" tests is problematic because it becomes very difficult to compare performance across various types of memory when tests use different procedures. For example, how can one compare one's performance on verbal versus visual memory when one test uses an intentional learning paradigm, in which the person knows memory is being tested and the other test uses an incidental learning paradigm, in which the person does not realize that memory will be tested? The same problem can occur with any number of test procedures including the use of single versus multiple trials, different combinations

of free recall, cued recall, and recognition, and different levels of task difficulty. In addition, it should be noted that difficulties in comparing test performance can also arise when the samples used to develop the normative data for each test differ (e.g., different age groupings, education levels, or gender ratio).

One way in which clinicians can perhaps avoid some of these problems is to use a battery of tests that have all been developed by the same test developer and have been normed on the same sample. Such batteries, which are referred to here as "comprehensive" memory tests, are described in the next section.

Comprehensive Memory Tests:

Comprehensive memory tests examine several different types or aspects of memory in one test battery (e.g., attention, verbal memory, visual memory). Such tests include the Guild Memory Test (Gilbert, Levee, & Catalano, 1981), the Learning and Memory Battery (Schmidt & Tombaugh, 1995; Tombaugh & Schmidt, 1992), the Memory Assessment Scales (Williams, 1991), the Randt Memory Test (Osborne, Brown, & Randt, 1982), and the Weschler Memory Scale - Revised (Weschler, 1987).

Of these tests, the Learning and Memory Battery (LAMB) will be described here as an example of a comprehensive memory test. Although this test is quite new, and certainly is not as widely used as the Weschler Memory Scale - Revised

(WMS-R), this test battery has been chosen as it is the forerunner of the test to be developed in this dissertation. The LAMB consists of seven subtests: Paragraph, Word List, Word Pairs, Digit Span, Supraspan, Simple Figures, and Complex Figure.

The Paragraph subtest, which consists of a 31-component story, is presented orally. Free recall, followed by cued recall for missing or incorrect information, is tested immediately on two acquisition trials and after a 20 minute delay. On the delay trial, any information that is still missed or incorrect is tested for using a four-choice recognition task.

Next, a list of 15 common words, each representing a different category (e.g., "dentist" - Profession) is administered orally. Five acquisition trials are used, although not all trials are given if the individual recalls all 15 words without help on two consecutive trials. Both free recall and cued recall (for any missing or incorrect information) are tested on each acquisition trial and on the 20 minute delayed recall trial. On the delay trial, any information that is still missing or incorrect is tested for using a four-choice recognition task. Buschke's (1973) selective reminding procedure, whereby only the words missed on both free recall and cued recall on the previous trial are presented on the next trial, is used during acquisition.

In the Word Pairs subtest, 14 pairs of words are read.

On each trial, when told the first word of the pair, the individual must recall the second word of the pair. Four acquisition trials are used, although not all trials are given if the individual recalls all 14 word pairs on two consecutive trials. On the 20 minute delayed recall trial, a four-choice recognition task is used for any word pairs that the person still cannot recall the second word of the pair. Buschke's (1973) selective reminding procedure, whereby only the word pairs missed on the previous trial are presented on the next trial, is used during acquisition.

The Digit Span subtest consists of a series of increasingly longer numbers (e.g., 3-7-2, 2-9-1-8, 4-2-8-6-7) which are presented orally to determine the longest series of numbers or digits that can be recalled with only one attempt. There are two trials or two series of numbers at each length (up to 8 or 9 digits) and each series is administered until the person fails two series at the same length. There are two parts to this subtest. In Digits Forward, when a number is presented (e.g., 2-9-1-8) the individual must recall it in the same order. In Digits Backward, when a number is presented (e.g., 3-9-5-8) the person must recall it in the opposite order (e.g., 8-5-9-3).

In the Supraspan subtest, a number is presented that is two digits longer than the longest number recalled correctly in Digits Forward. The person is given up to 12 trials to learn this new number, although not all trials are given if

the individual recalls the number correctly on three consecutive trials. A 20 minute delayed recall trial is administered, and up to 6 relearning trials are given if the person fails to recall the number correctly.

In the Simple Figures subtest, four simple figures displayed horizontally on a card are shown for 15 seconds. After removal of the card, the person is to draw the figures in the same order as they were shown. There are three acquisition trials, a 20 minute delayed recall trial, and a copy trial.

In the last subtest, the Taylor Complex Figure (Taylor, 1969) is presented for 30 seconds. After the figure is removed, the person is to draw it as accurately as possible. There are four acquisition trials, a 20 minute delayed recall trial, and a copy trial.

An advantage of comprehensive memory tests is that many different aspects of memory are examined in one test battery. However, some of the same weaknesses that plague specific tests of memory can be found in many comprehensive tests. That is, inconsistencies in the procedures (e.g., use of a retrieval continuum) and paradigms (e.g., intentional vs incidental learning) can be found even within subtests of a comprehensive test battery.

The present dissertation involves the development of two brief, but comprehensive, tests of memory for older adults. Both tests include measures of verbal and visuo-

spatial memory using an intentional learning paradigm. In the next section, testing procedures that have been shown in the theoretical, empirical, and clinical research to be critical in distinguishing normal and impaired memory performance are described.

Application of Theoretical and Clinical Knowledge

To construct a memory test that would take into account *all* of the aspects of memory described in current theories and research would be impractical. A more realistic approach to developing test materials appropriate for the assessment of memory performance in older adults is to identify both the aspects of memory functioning and the testing procedures that are sensitive to distinguishing normal performance from impaired performance.

First, any memory test designed for use with an older population must be practical. For this reason, both the Diagnostic and Tracking memory tests to be developed in this dissertation will be comprised of only two subtests: (a) a paragraph measuring verbal memory and (b) a geometric figure measuring visuo-spatial memory. Deficits in the acquisition and retention of verbal and visuo-spatial information are a few of the earliest and most diagnostic symptoms of dementia (Fuld, 1983; Moss & Albert, 1988; Salmon, Shimamura, Butters, & Smith, 1988). There are several advantages to these two subtests. First, because there are only two subtests, this makes the two memory tests relatively short

tests and reduces the chances of unnecessarily fatiguing the individual being tested. This is particularly beneficial when testing older adults who may be frail, in poor health, or easily distracted. Each memory test takes approximately 30-45 minutes to administer. A second benefit is that both tests are easily portable because neither subtest requires the use of unwieldy materials. Third, both subtests are independent of each other in terms of instructions, tasks, and performance. That is, the subtests need not be administered together and neither subtest influences how the other subtest is administered.

Second, since it was noted in the earlier discussion of primary memory that measures of primary memory generally are insensitive to memory dysfunction, the materials used in the two tests to be developed here will focus primarily on secondary memory. These memory tests will allow the clinician to examine an individual's use of encoding and retrieval strategies. In particular, the ability of a person to use retrieval strategies should be examined through the use of a retrieval continuum. That is, free recall, cued recall, and recognition measures should be used so that one can differentiate between information that has actually been forgotten as opposed to information that is present but cannot be spontaneously retrieved. Age-related differences in memory performance can be substantially reduced or, in some cases, eliminated when cued recall and recognition

measures are used. Moreover, free recall, cued recall, and recognition measures have been shown to have differential utility in determining the presence and the severity of Alzheimer's disease (Branconnier, Cole, Spera, & DeVitt, 1982; Storandt, Botwinick, & Danziger, 1986; Tuokko & Crockett, 1989; Wilson & Kaszniak, 1986). Finally, such information can have significance in terms of rehabilitation. For example, the usefulness of cuing may identify the degree to which an individual would benefit from having cues built into his or her environment.

Third, it is important to include both immediate and delayed recall in any memory test that aims to provide an adequate representation of overall memory (Ardila & Rosselli, 1989; Bornstein & Chelune, 1989; Bornstein, Chelune, & Prifitera, 1989; Butters et al., 1988; Erickson & Scott, 1977; Larabee, Kane, Schuck, & Francis, 1985). Immediate recall, which is assessed immediately after the material has been presented, allows one to determine the amount of information learned during the acquisition phase. Delayed recall typically takes place 5-30 minutes after immediate recall. It is critical to employ a delay interval that is short enough for healthy individuals to retain most of the information, yet long enough to facilitate the loss of information among impaired individuals. Typically, a 15 to 20 minute delay is sufficient to reveal an abnormal level of forgetting. However, forgetting can be very rapid in

some disease processes. For example, with Alzheimer's patients, intervals from 5-10 minutes are typically long enough to show forgetting (Corkin, 1982; Hart, Kwentus, Harkins, & Taylor, 1988; Kopelman, 1985; Sahakian et al., 1988; Moss & Albert, 1988). Delay intervals on the Diagnostic and Tracking Tests are 10-15 minutes in length.

Use of immediate and delayed recall also permit the calculation of the proportion of information retained during the delay interval. Such a calculation has been shown to be a sensitive measure of memory deficit (Boinstein, Pakalnis, Drake, & Suga, 1988; Butters et al., 1988; Milner, 1975), and capable of differentiating Alzheimer's disease from healthy elderly, Korsakoff's disease, and Huntington's disease (Moss, Albert, Butters, & Payne, 1986). In addition, measures of the deterioration between immediate and delayed recall seem especially sensitive in differentiating genuine neurological conditions (e.g., early dementia) from those that might masquerade as neurological (e.g., depression, pseudodementia, malingering).

Finally, multiple trials should be used to assess an individual's ability to learn (Erikson & Scott, 1977; Schmidt & Tombaugh, 1995). Repeated presentation of the test stimuli serves several functions. First, multiple trials are needed in order to assess the ability to learn and to evaluate different types of learning patterns. The degree to which performance improves across trials reveals

the extent to which information is being learned or transferred into secondary memory. Second, multiple trials reduce the impact of other factors such as distractibility, inattention, or difficulty shifting cognitive sets on the overall test results. Finally, multiple learning trials also increase the likelihood that the material has been learned so that deficits found on retention trials are more interpretable.

The number of acquisition trials chosen for the Paragraph and Geometric Figure subtests was based on pilot testing with younger and older adults. For example, three trials were chosen for the Geometric Figure subtests because many intact individuals demonstrated a considerable increase in learning from the second to the third trial. Because the figure could prove to be frustrating for individuals who perform poorly on it, three trials seemed to provide an ideal balance between maximizing learning and minimizing frustration and length of testing.

Applying theoretical and clinical knowledge to test development is the first step in developing a practical and sensitive test. An important second step is to apply appropriate psychometric criteria to that test. The next section will discuss the psychometric criteria that are considered important for good test development.

Use of Psychometric Criteria to Evaluate Tests of Memory

Several writers have bemoaned the lack of adequate

measures of memory for clinical use and have proposed criteria for the "ideal" memory test (Erickson, Poon, & Walsh-Sweeney, 1980; Erickson & Scott, 1977; Mayes, 1988; Ritchie, 1988). The following standard psychometric criteria have been suggested by these authors and summarized by Schmidt & Tombaugh (1995).

First, there should be a clearly-stated rationale for test and item selection that should reflect an awareness of both the experimental literature and the current theoretical understanding of memory. The test rationale should also acknowledge practicality of the test in terms of clinical administration, and the strengths and weaknesses of other measures currently used in clinical settings.

Second, it is important that scale reliability be demonstrated. Clearly, a test cannot be of value unless it is able to measure the construct of interest consistently. One test of reliability that takes into account errors of measurement caused by different conditions or times of administration, is test-retest reliability. However, the measurement of test-retest reliability in memory tests is complicated by the impact on these tests of learning and practice effects. Thus, test-retest reliability is difficult to demonstrate unless a lengthy period of time has elapsed between test administrations. Another test of reliability is internal consistency. Internal consistency takes into account the effects of different samples of test

items on reliability. Different methods of calculating internal consistency include the split-half method, the Kuder-Richardson method, and Cronbach's alpha. It should be kept in mind that the reliability of a test is affected by (a) the length of the test, (b) the difficulty of the test items and (c) the heterogeneity of the group being tested. Finally, whenever two or more individuals are involved in scoring material, inter-rater reliability should be demonstrated either through a Pearson's correlation or an intraclass coefficient.

Third, test validity should be demonstrated. First, however, it should be noted that validation of an instrument is an ongoing process. An important aspect of validity is construct validity. One of the most popular lines of evidence for construct validity is the correlation of a new measure to an existing "gold standard". For example, performance on a new test of visual memory might be correlated with an already accepted test of visual memory. However, in the absence of a universally-accepted "gold standard" or when the test developer believes that other tests are substandard, a definition of the construct being examined must be developed and a "nomological net" needs to be demonstrated.

A nomological net is a set of predictions about what other variables should, or should not, relate to the construct (i.e., convergent and discriminant validity). For

example, a test should show strong correlations with related cognitive abilities, such as intelligence, and low correlations with variables that should not be strongly related to memory, such as confrontational naming or the ability to name objects. A nomological net would also predict the types of populations or groups that would differ from each other on the test. Thus, performance on a verbal memory test should be different for healthy individuals than for Alzheimer's patients.

A comment needs to be made about face validity, which is not a true form of validity at all, but considers whether a test *looks* like it is measuring what it claims to measure. Older adults tend to be particularly sensitive to the face validity of a test (Erickson & Scott, 1977). When a test seems to be a reasonable measure of the construct of interest older adults are more likely to be motivated and treat the testing situation seriously.

Finally, appropriate normative data should be developed. Test norms allow the examiner to understand the meaning of the attained scores. A test developer must base norms on groups that represent the population on which the test will be administered. Although representativeness stems from a multitude of variables, the minimal variables for a cognitive test include age, gender, educational level, and intelligence. A test developer's assurance that groups are representative is not, however, sufficient to provide

adequate norms. Many of these variables may interact with the memory scores. For example, older subjects will probably do less well on memory tests than younger subjects. If a demographic variable, (e.g., age) does interact with or affect test scores, then the developer must control it. A developer may achieve control by sample stratification. Sample stratification requires the establishment of separate norms for individuals at different levels of the to-be-controlled variable (e.g., age and/or education stratified norms). A second way a developer can control such variables is by using statistical procedures. This is typically done by developing a regression equation which allows one to adjust the obtained score. The new adjusted score is then applied to the existing normative tables. For the regression equation to be useful with memory tests, it must use nonlinear equations because the effects of age on memory are not linear. Additionally, because such methods tend to be statistically unstable, equations should be based on large samples. Because of such factors, use of equations is generally the less desirable alternative.

Geriatric Learning and Memory Battery (G-LAMB):

Test Development

Because the procedures and types of memory examined in this new memory test for older adults rely heavily upon evidence within the current literature regarding their theoretical value and clinical utility, the uniqueness of the G-LAMB does not lie in the development of entirely new memory assessment methodologies. Rather, its uniqueness lies in the application of existing methods in a standardized manner with adequate norms to enhance more systematic interscale comparison and clarity of interpretation.

The G-LAMB consists of two different memory tests, both of which test an individual's learning and memory for both verbal and visual stimuli. All subtests consist of original material; that is, none of the material used (e.g., paragraph content, geometric designs) is contained in other previously published tests.

Diagnostic Test

When the goal of a test is to diagnose whether a memory impairment exists, one needs to include procedures that best distinguish between normal and impaired memory performance. Thus, procedures such as using a delayed recall trial are employed in both the Paragraph and the Geometric Figure subtests. It is also important to use materials that are not so difficult that they frustrate the individual being

tested, but are still difficult enough that they can be used to distinguish between normal performance and mild impairment. Thus, the Rectangle Geometric Figure is simpler than the traditional complex figures (e.g., Rey-Osterrieth or Taylor complex figures), but more difficult than the very simple figures provided in some tests (e.g., WMS-R). And, although the number of details in the Paragraph subtest proves to be challenging, the use of a self-read procedure and two learning trials also provides individuals with the opportunity to employ their own learning strategies.

Next, a general description of the Paragraph and Rectangle Geometric Figure subtests will be given, along with the rationale for their use and the unique features or advantages of each subtest.

Paragraph Subtest:

The Paragraph subtest, which requires the individual to learn and remember a short story composed of 31 details or components, is presented below. Each of the 31 components is contained with slash marks with the key information within a phrase underlined.

Carol / Simpson, / aged 57, / is a bookkeeper / at
Murphy's / Garage. / She lives in a brown house / at 14 /
Willow Ave., / 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.

First, the story is read to the person at a normal reading pace. Next, he or she is given 30 seconds to read and study the story himself or herself. Then, the story is removed and the person is asked to recall as much of the story as he or she can. There are two learning trials on the paragraph. Performance is based on the recall of 31 components or details in the story. After a delay interval of approximately 10-15 minutes, free recall is tested again. For any components that are not remembered during free recall, the person is asked a (cued recall) question about that component (e.g., "What was the person's name?"). If the subject cannot recall the detail at all or correctly, he or she is immediately presented with a 4-item multiple choice recognition test (e.g., "What was the person's first name?...Was it Carol, Karen, Ann, or Cathy?"). The cued recall questions and recognition choices are presented

below.

1. What is the person's first name?
(Carol, Karen, Ann, Cathy)
2. What is the person's last name?
(Andrews, Murphy, Martin, Simpson)
3. What is the person's age?
(42, 53, 57, 69)
4. What is the person's occupation?
(secretary, mechanic, teacher, bookkeeper)
5. What type of place does she work?
(garage, bakery, department store, factory)
6. What is the name of her employer?
(Richmond, Murphy's, Young's, Mike's)
7. What is the colour of her house?
(green, white, grey, brown)
8. What is the number of her street address?
(3, 7, 14, 18)
9. What is the name of the street where she lived?
(Erindale, Queen, Willow, York)
10. What is the name of the city where she lived?
(Seattle, Vancouver, Boston, Toronto)
11. What time did she leave in the morning?
(6:15, 6:45, 7:15, 7:45)
12. What is the colour of her dog?
(grey, black, brown, white)
13. What type of dog did she have?
(terrier, poodle, collie, beagle)
14. What was the name of her dog?
(Rover, Champ, Peanut, Brandy)
15. Where did she walk the dog?
(park, woods, farm, field)
16. Where did she buy something?
(bakery, market, pharmacy, butcher)

17. What type or nationality was the 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. On 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. What type of transportation did she use?
(plane, car, bus, train)
24. What state did she go to?
(Kansas, Ohio, Missouri, Michigan)
25. Who did she go 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. What type of activity or sport did she do?
(gardening, swimming, fishing, or jogging)
29. What type of fish was she fishing for?
(sunfish, perch, trout, pickeral)
30. What was the first name of the magazine that she read?
(Country, Valley, Outdoor, Western)
31. What was the second name of the magazine that she read?
(Gazette, Times, Living, Digest)

Performance on each trial is based on the recall of 31 components or details in the story. Three scores, each out of a maximum 31 points, can be calculated for the Paragraph

subtest. The first score, free recall (FR), consists of the number of items correctly recalled without help. This score is calculated for each acquisition trial as well as the delay trial. The next two scores are calculated on the delay trial only. Free recall plus cued recall (FR+CR) consists of the FR score plus the number of correct responses on the cued recall portion. The final score, free recall plus cued recall plus recognition (FR+CR+Rec.), consists of the FR+CR score plus the number of correct responses on the recognition portion. Scoring is relatively straight-forward; one point is given for each detail or component recalled. The one caveat is that points are not given for information recalled in the wrong context. For example, if an individual recalls that "the woman had a brown poodle named Brandy that she takes for a walk", points are given for "poodle" and "Brandy", but not for "brown" which was the colour of the house and not the dog.

. Paragraph memory is assessed using a paragraph rather than a word list so that an individual's ability to use and benefit from the underlying structure of the material can also be examined. Often, verbatim recording of the individual's responses illuminates the reasons for poor performance on paragraph recall. Usually, those who perform poorly do so because they are unable to remember details, or they are unable to use the underlying structure as a cuing method. Paragraph learning and delayed recall have been

shown to be sensitive to the effects of aging on memory and in the assessment of general memory impairment (Corkin, Growdon, Sullivan, Nissen, & Huff, 1986; Danziger & Storandt, 1982; Osborne, et al., 1982; Storandt, et al., 1986).

The paragraph in the Diagnostic Test was developed with several features in mind. The first, and a feature unique to the G-LAMB, is that a self-read procedure is used. That is, individuals being tested are given the opportunity to read and study the paragraph themselves. It has been found that performance differences between younger and older adults are reduced when individuals are allowed to study information at their own pace (Campbell & Allen, 1988; Crosier, 1990). In the self-read procedure used in the G-LAMB, individuals are given a set period of time (30 seconds) to read and study the material. Pilot data indicated that 30 seconds was enough time for most older adults to comfortably read over the paragraph once on their own. This set period of time is important because it controls the length of exposure that individuals are given to the material.

There are several advantages to using a self-read procedure. First, allowing individuals to read the material themselves gives them the opportunity to use whatever skills or strategies they have at their disposal. For example, they can choose how many times they wish to read through the

material and to what parts of the material they wish to direct their attention. Second, a self-read procedure avoids the interference effects that can occur when the paragraph is only administered orally. That is, trying to listen to what the examiner is saying (e.g., "...is a bookkeeper at Murphy's Garage") interferes with the learning of the material already said (e.g., "Carol Simpson, age 57 is..."). Allowing the person to read the material enables him or her to overcome any initial feelings of being overwhelmed by the material. It should be noted, however, that in the self-read procedure, the examiner first reads the paragraph to the individual. This ensures that all individuals being tested are exposed to the paragraph in its entirety at least once. Finally, a self-read procedure not only adds to the face validity of the task, but it is eagerly accepted by individuals being tested. When the LAMB test was being normed in the late 1980s, many older adults indicated that they would remember more if they could only read the material.

Another feature of the Paragraph subtest is that, unlike most passages, the paragraph does not have a consistent story or theme. The presence of a theme has proven, for similar paragraphs, a major obstacle to reliable scoring and rational interpretation. In the present story, it is very difficult for an individual to correctly guess a detail based simply on the context. For example, the woman

in the story buys "three" cookies even though only herself and her dog are mentioned in the story. In addition, the dog in the story is "black", despite the fact that most people think of "poodles" as being white and a dog named "Brandy" as being brown. The lack of a consistent theme also makes scoring somewhat easier. For stories with themes, the dilemma has been that an individual who gives a complete rendition of the theme of the story while providing few details or alternate wording may get very few points whereas an individual who provides nothing other than a fragmented rendition of isolated bits of the story may get a large number of points. This calls into question the logic of concluding from the test scores that the first individual has a major memory deficit whereas the second individual does not.

A third feature is that cued recall is only used for items that are missed on free recall, and recognition is only used for items still missed on cued recall. Both cued recall and recognition are administered only in the delayed recall trial. There are two disadvantages in administering cued recall in the learning trials (as is done in the IAMB). The first is that it can be both frustrating and discouraging to individuals who do not recall very much of the paragraph, particularly on the first trial. As can be imagined, asking a question about every detail that is missed ignores what has been learned and highlights what has

not. Another disadvantage is that this procedure not only increases the length of the testing dramatically, but it can also lead to differences in the time interval between trials from one person to another.

In test batteries such as the LAMB, the cued recall portion of the test is administered in its entirety before the recognition portion is given. In the G-LAMB Paragraph subtest, recognition is administered immediately after cued recall for each item on the delay trial. That is, if the person does not recall an item with a cue, recognition for that item is given immediately before the next cued recall question is asked. The advantage of this procedure is that it reduces the length of time for testing since the cued recall question does not need to be repeated when giving the recognition choices.

Geometric Figure Subtest:

The Geometric Figure subtest requires the individual to learn and remember the following simple geometric figure.

The person is given 30 seconds to study the figure. The figure then is removed and the person is asked to draw as much of it as he or she can remember within a 2 minute time limit. Pilot testing suggested that most older adults were able to draw what they remembered of the figure well within two minutes. There are three learning trials on this subtest. After a 10-15 minute delay interval, the person is again asked to draw as much of the figure as he or she can remember within a 2 minute limit. Finally, with the geometric figure placed in front of him or her, the person is asked to copy the figure without using their memory within a 3 minute time limit. A copy trial is necessary to determine whether the person can actually copy the figure accurately. If performance on immediate and/or delayed

recall is poor and yet the copy trial of the figure is acceptable, then one has more confidence in attributing the poor performance to learning or memory difficulties. If, however, the copy trial of the figure is also poor (i.e., missing details or extremely sloppy drawing), then one must consider whether other factors (e.g., hemi-neglect, inattention, lack of motivation, or perceptual difficulties) are at fault. Pilot testing indicated that most older adults were able to copy the figure well within a three minute time limit.

Performance is based on the recall of the 13 components and their placement in the figure. A scoring system for the Rectangle Geometric Figure was devised, using the scoring systems for the Rey-Osterrieth and Taylor complex figures provided in Lezak (1995) as models, and is provided in Appendix A.

Visual memory is assessed using a geometric figure. Geometric figures of varying degrees of complexity traditionally have been used in neuropsychological assessment because of their sensitivity to visuo-spatial memory deficits (Lezak, 1995) and to dementing processes (Becker, Boller, Saxton, & McGonigle-Gibson, 1987; Bigler, Rosa, Schultz, Hall, & Harris, 1989). In addition, complex figures often can provide additional qualitative information that is useful in identifying visuo-spatial problems. For example, learning strategies can be examined when a record

is kept of the evolution of the drawings from trial to trial. In addition, an examination of the types of errors made (e.g., repetition or perseveration of details, missing details, distortions of details) can be very useful in distinguishing among different disorders (Binder, 1982; Gainotti & Tiacci, 1970; Martin, 1987). For example, if details or portions of the geometric figure consistently are missing from one side or the other of the figure, hemi-neglect or an inattention to one side of the visual space may be suspected. The geometric figure in the Diagnostic Test was designed to assist in identifying hemi-neglect by placing several figural components to the far left (i.e., a circle with an equal sign in it) and the far right (i.e., a square with an arrow in it) of the figure.

The geometric figure in the Diagnostic Test was developed with several features in mind. First, the geometric figure used here is of an intermediate level of difficulty. Very simple figures consist of only a few simple components. A disadvantage of these figures is that they can be verbally encoded relatively easily. For example, the simple figures used in the Learning and Memory Battery (Schmidt & Tombaugh, 1995; Tombaugh & Schmidt, 1992) consist of four horizontally-presented simple drawings that basically look like a squarish-looking backwards "5", a square with a curve inside it, an upside-down triangle with a short line inside it, and a circle that looks like a pie

chart. Typically, such figures are used to assess visual memory functioning in individuals who find complex figures virtually impossible to learn. Complex figures, such as the Rey-Osterrieth and Taylor figures, contain as many as 18 components. More complex figures have the advantage of being much more difficult to encode verbally which makes them "purer" tests of visual memory.

There are difficulties though in using the currently available geometric figures with older adults. Several researchers have reported that complex figures may be too difficult and frustrating for many older adults (Read, 1987; Tombaugh, Faulkner, & Schmidt, 1989). Yet, simple figures may not be difficult enough to be sensitive to minor or early visual memory impairments. The geometric figure presented in the Diagnostic Test consists of 13 components. The fewer number of components makes the figure less overwhelming to the person trying to remember it while still making the figure challenging and difficult to encode verbally.

A second feature of the geometric figure has to do with the use of an intentional learning procedure (Tombaugh, et al., 1992) similar to that used with the Paragraph subtest. Generally, there are two standard procedures in administering visual memory tests using geometric figures. These procedures seem to be related to the complexity of the design. Memory tests for simpler designs require the

subject to observe a design for a brief period (e.g., 10 seconds) and then reproduce it from memory after it is removed. Examples of this are the Benton Visual Retention Test (Benton, 1974) and the Visual Reproduction subtest of the WMS-R (Wechsler, 1987). In some instances, an additional delayed recall trial is employed. Memory tests for complex geometric designs, such as the Rey-Osterrieth and Taylor complex figures, typically use an incidental learning procedure. First the subject is asked to copy the figure while it is present. Then the figure is removed and the subject must reproduce it from memory. In this case, the subject is not informed of the retention test prior to its administration. Again, a delayed recall trial is sometimes used.

Performance on many of the visual tests using complex figures are difficult to interpret, however, because of the use of an incidental learning procedure. There are three problems with using an incidental procedure. First, the single acquisition trial (copy trial, in this case) does not provide sufficient information concerning rate of learning. Because of the difficulty level of complex designs, it is important to ensure that the material has been adequately learned before testing for retention. Second, since most verbal tests use intentional learning, this makes it difficult to compare an individual's performance on tests of verbal and visual memory. In addition, use of an incidental

procedure confounds visual memory with such unrelated variables as lack of attention, concentration, and motivation to remember. Third, the procedure of allowing unlimited time to copy the figure leads to differential amounts of exposure time. This makes it difficult to interpret differences in memory scores for a person who has taken 3 minutes to copy the design versus one who has taken only 1 minute since the longer an individual is exposed to the figure, the more accurate the person's recall (Tombaugh & Hubley, 1991).

Tracking Test

When the goal of a test is to monitor an impaired individual's memory performance over time, one needs to use procedures that give the person ample opportunity to compensate for any memory difficulties. It is for this reason that procedures such as self-reading of the paragraph are included. One also requires materials that are easier than even those suggested in the Diagnostic Test. As a result, the Paragraph subtest consists of only 21 rather than 31 components. Not only is it important that the individual remain motivated, but one is now interested in knowing what the person's strengths are, in addition to his or her weaknesses. One needs test materials that provide more support (e.g., are less overwhelming or confusing) so that one can examine when such support helps the person and when it does not. Thus, the Triangle Geometric Figure

subtest in the Tracking Test is more balanced than the one in the previous test and the Paragraph subtest in Tracking Test has a more consistent storyline.

Paragraph Subtest:

The Paragraph subtest, which requires the individual to learn and remember a short story that is composed of 21 details or components, is presented below.

On Sunday, / Jane / and her five-year old / brother / went to a circus / in Paris. / At 4:15, / when they were buying / their tickets, / it started to rain. / As they were running / for the animal / tent, / the boy tripped / and cut / his leg. / A clown / gave him two / balloons, / and phoned / his mother.

First, the story is read to the person at a normal reading pace. Next, the person is given 30 seconds to read and study the story to himself or herself. Then, the story is removed and the person is asked to recall as much of the story as he or she can. There are two learning trials on the paragraph. Performance is based on the recall of 21 components or details in the story. After a 10-15 minute delay, free recall is tested again. For any components that are not remembered during free recall, the person is asked a question about that component and given four possible answers from which to choose (e.g., "How old was Jane's

brother?...Was he five, seven, ten, or thirteen?"). The recognition questions used in the Tracking Test are presented below.

1. On what day did the story take place?
(Tuesday, Thursday, Friday, Sunday)
2. Was the woman's name...
(Jennifer, Betty, Jane, Sarah)
3. Who did Jane take with her?
(her brother, her sister, her nephew, her niece)
4. How old was Jane's brother?
(five, seven, ten, thirteen)
5. Did Jane and her brother go to a...
(museum, zoo, circus, concert)
6. Was the circus in...
(New York, Paris, Toronto, London)
7. When Jane and her brother were at the circus, did it...
(snow, cloud over, hail, rain)
8. Did it start to rain at...
(1:15, 1:45, 4:15, 4:45)
9. When it started to rain, Jane and her brother were...
(looking for something, eating something, picking up something, buying something)
10. When it started to rain, Jane and her brother were buying...
(tickets, peanuts, programmes, hotdogs)
11. When it started to rain, did Jane and her brother...
(put up an umbrella, run, put on their raincoats, go home)
12. Did Jane and her brother run toward the...
(gate, tent, washrooms, bus shelter)
13. Did Jane and her brother run toward the...
(red tent, striped tent, animal tent, open tent)
14. Did Jane's brother...
(win a prize, trip and fall, get lost, lose his money)

15. When Jane's brother tripped, did he...
(cut himself, drop his ice-cream, faint, break his arm)
16. Did the boy cut his...
(head, hand, leg, foot)
17. Who helped the boy?
(a stranger, Jane, a midget, a clown)
18. Did the clown...
(bandage the wound, carry the boy home, sell more tickets, phone someone)
19. Did the clown phone...
(his supervisor, the boy's mother, an ambulance, a nurse)
20. Did the clown give the boy...
(balloons, a big smile, an elephant ride, popcorn)
21. How many balloons did the clown give to the boy?
(two, three, four, five)

Performance on each trial of the Paragraph subtest is based on the recall of 21 components or details in the story. Two scores, each out of a maximum 21 points, can be calculated for the Paragraph subtest. The first score, free recall (FR), consists of the number of items correctly recalled without help. This score is calculated for each acquisition trial as well as the delay trial. Free recall plus recognition (FR+Rec.) consists of the FR score plus the number of correct responses on the recognition portion. Scoring is relatively straight-forward; one point is given for each detail or component recalled. As with the Paragraph subtest in the Diagnostic Test, points are not given for information recalled in the wrong context. For example, if an individual recalls that "a clown phoned

Jane", points are given for "clown" and "phoned", but not for "Jane" because even though Jane is a character in the story, the clown actually called the boy's mother.

Verbal memory is assessed using a paragraph in the Tracking Test for the same reasons as in the Diagnostic Test, and also to parallel this test. Many of the same features of this subtest appear again. For example, a self-read procedure is used and details from the paragraph cannot be easily guessed from the context (although the storyline is perhaps more consistent than in the Diagnostic Test paragraph).

There are two major differences between the Paragraph subtests in the Diagnostic and Tracking tests. The first difference is that the paragraph used in the Tracking Test is comprised of only 21 components. Fewer details, combined with the same period of exposure (30 seconds), makes this paragraph considerably less difficult to learn and remember. The second difference is that in the Tracking Test, only free recall and recognition procedures are used. The main reason for removing cued recall is that, in the Tracking Test, the primary interest at retention is in determining: (a) what can be spontaneously recalled, and (b) what material is still there (i.e., has not been entirely forgotten). The latter information can be obtained through recognition procedures without the need for cued recall. In addition, removing cued recall shortens the length of the

test considerably.

Geometric Figure Subtest:

The Geometric Figure subtest requires the individual to learn and remember the following simple geometric figure.

The person is given 30 seconds to study the figure. Then, the figure is removed and the person is asked to draw as much of it as he or she can remember within a 2 minute time limit. There are three learning trials on this subtest.

After a 10-15 minute delay interval, the person is again asked to draw as much of the figure as he or she can remember within a 2 minute limit. Finally, with the geometric figure placed in front of him or her, the person is asked to copy the figure without using their memory within a 3 minute limit.

Performance is based on the recall of the 13 components and their placement in the figure. A scoring system for the Triangle Geometric Figure was devised, using the scoring systems for the Rey-Osterrieth and Taylor complex figures provided in Lezak (1995) as models, and is provided in Appendix B.

Visual memory is assessed using an intermediate level geometric figure in the Tracking Test for the same reasons as in the Diagnostic Test, and also to parallel this test. The main difference between the two figures has to do with their design. With a triangle as the main shape, the figure is neatly divided into three sections. The components of the figure are drawn more closely together. There is no attempt to diagnose disorders, such as hemi-neglect, with this geometric figure. Rather, the interest is in maintaining a balance in the figure that individuals may be able to use to help them organize, learn, and recall the information.

Now that the materials and procedures for the G-LAMB Diagnostic and Tracking tests have been described, the next

step is to collect normative data for these tests and to accumulate evidence of reliability and validity.

Method

Study 1

Goal

The Diagnostic Test was designed to diagnose memory impairment in older adults by using test materials and procedures that have been shown in the literature to best distinguish between normal and impaired performance. Study 1 involves the collection of both normative data and some preliminary evidence of the reliability and validity of the Diagnostic Test'.

Subjects

The normative sample in Study 1 consists of 184 community-dwelling volunteers aged 55-85 years from the Ottawa-Carleton region. Participants were recruited through hospital volunteer groups, community organizations, seniors buildings, volunteer booths at shopping centres, and personal contacts. They did not receive any financial remuneration for taking part in the study.

A brief self-reported medical history and a list of all currently prescribed medication was obtained from each participant. In order to reflect the average older adult more accurately, only individuals with a history of multiple cognitive risk factors (e.g., head injury or organic diseases such as heart attack, stroke, or diabetes) and/or unusually low performance in the study that might be related to a history of head injury or organic disease, were

excluded from the study. In addition, participants who presented as cognitively impaired (score ≤ 23) on the Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) and/or moderately depressed (score ≤ 13) on the Geriatric Depression Scale (GDS; Yesavage, et al., 1983) were also excluded from the study. Such exclusions reduced the sample from 215 to 184 participants.

The participants were divided into the following six age groups: 55-59, 60-64, 65-69, 70-74, 75-79, 80-85. Each group contained approximately 30 subjects. For each of these age groups, approximately two-thirds of the participants have an educational level less than or equal to a high school education and approximately one-third of the subjects have an educational level greater than high school.

Test Materials

A variety of test materials were administered in Study 1. The Diagnostic Test materials were described in the previous chapter and were administered as described. This next section includes a brief description of (a) the tests used to screen the participants, and (b) the "filler" tasks used in the delay intervals of the Diagnostic Test.

Screening Tests

Three tests are given in this study that are used to describe and/or exclude subjects from the normative sample. These tests are the WAIS-R Vocabulary subtest, the Mini-Mental Status Examination (MMSE), and the Geriatric

Depression Scale (GDS). These tests are commonly used in memory studies with older adults, and standard administration procedures will be used. However, for those who are less familiar with these tests, a brief description is provided in Appendix C.

Filler Tasks

A number of "filler" tasks are used in the intervals between the acquisition and delay trials on the Paragraph and Geometric Figure subtests. Some of these tasks are previously published tests for which the standard administration procedures were used. These tasks included: Trails A & B, Word Fluency (F, A, S, and Animals), a subjective age questionnaire, and the Boston Naming test. A brief description of each of these tests is provided in Appendix C. Other tasks included recently-developed tests that have not been previously published. These tasks included: a 15-item word list, a 24-item recognition memory task for abstract drawings, a chronological age questionnaire, a second depression scale, and a memory change questionnaire. These tasks are described in more detail in Appendix D. Use of all test materials was approved by Carleton University's Ethics Committee (see Appendix E).

Procedure

Before testing began, informed consent was obtained and subjects were administered the MMSE. Subjects were informed

that they would be given several opportunities to learn and remember different types of information. The order in which the tests were administered in Study 1 is provided in Appendix F. Next, the instructions for the Paragraph and Geometric Figure subtests of the Diagnostic Test will be provided.

Paragraph:

Instructions for the Paragraph subtest are as follows:

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."

Once the tester has read the story and the subject has had 30 seconds to read and study it, the subject is told, "Now, tell me all you can remember about the story". Immediately afterwards, Trial 2 is administered as follows:

Trial 2: "I will read the story to you again. Then I will give the story to you for 30 sec. to read and study yourself. After that, I will remove the story and ask you to tell me all you can remember about it."

Once again, after the tester has read the story and the subject has had 30 seconds to read and study it, the subject is told, "Now, tell me all you can remember about the story".

Approximately 10-15 minutes later, the delay trial is administered as follows:

Delay: "Remember the story that I read and you studied? Tell me all you can remember about it."

Once the subject has recalled all they can remember, cued recall and recognition are administered as follows:

"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."

Geometric Figure:

On each trial of the Geometric Figure subtest, and before the instructions are given, a blank sheet of paper is placed horizontally in front of the subject and the person's initials and the trial number are printed in the top right corner. Instructions are then as follows:

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 give you 30 seconds to study the design. Then I'll remove the design, and ask you to draw it from memory."

Once the subject has studied the geometric figure for 30 seconds, he or she is told, "Now, draw the geometric design I showed you as accurately as you can. Make it approximately the same size". The subject has a maximum of two minutes in which to draw what they can remember for each acquisition trial. Immediately afterwards, Trials 2 and 3 are administered as follows:

Trial 2 & 3: "I am going to show you the same design again for 30 seconds. Then I'll remove the design, and ask you to draw it from memory."

Once again, after the subject has had 30 seconds to study the design, he or she is told, "Now, draw the geometric design I showed you as accurately as you can. Make it

approximately the same size".

Approximately 10-15 minutes later, the delay trial is administered as follows:

Delay: "Remember the figure I showed you earlier? Please draw it as accurately as you can. Make it approximately the same size."

The subject has a maximum of two minutes in which to draw all he or she can remember. Immediately afterwards, a copy trial is administered as follows:

Copy: "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."

The subject is given a maximum of three minutes in which to copy the geometric design.

Method: Study 2

Goal

The Tracking Test was designed to examine or track memory performance in individuals who have already been diagnosed as suffering from a memory impairment. Somewhat easier materials are used than those described in the Diagnostic Test in order to reduce anxiety, frustration, and discouragement in the person being tested. In addition, the test includes procedures that provide individuals with the opportunity to utilize all of their memory strategies. Study 2 involves the development of normative data for the Tracking Test using a community-dwelling sample of older adults as well as provide some initial evidence for the

reliability and validity of this test.

Subjects

The normative sample in Study 2 consists of 180 community-dwelling volunteers aged 55-85 years from the Ottawa-Carleton region. Although participants generally were recruited through hospital volunteer groups, community organizations, seniors buildings, volunteer booths at shopping centres, and personal contacts, 156 were participants from Study 1 who volunteered to return for Study 2. A minimum delay of three months was required for subjects who participated in both studies in order to avoid interference from the materials in one study to the other. Participants did not receive any financial remuneration for taking part in the study.

A brief self-reported medical history and a list of all currently prescribed medication was obtained from each participant. In order to reflect the average older adult more accurately, only individuals with a history of multiple cognitive risk factors (e.g., head injury or organic diseases such as heart attack, stroke, or diabetes) and/or unusually low performance in the study that might be related to a history of head injury or organic disease, were excluded from the study. In addition, participants who presented as cognitively impaired (score ≤ 23) on the Mini-Mental State Examination (MMSE; Folstein, et al., 1975) and/or moderately depressed (score ≤ 13) on the Geriatric

Depression Scale (GDS; Yesavage, et al., 1983) were also excluded from the study. Such exclusions reduced the sample from 248 to 130 participants.

As with Study 1, the participants were divided into the following six age groups: 55-59, 60-64, 65-69, 70-74, 75-79, 80-85. Each group contained approximately 30 subjects. For each of these age groups, approximately two-thirds of the participants have an educational level less than or equal to a high school education and approximately one-third of the subjects have an educational level greater than high school.

Test Materials

In Study 2, a number of different tests were administered. The Tracking Test materials were described in the previous chapter and were administered as described. The next section includes a brief description of (a) the tests used to screen the participants, and (b) the "filler" tasks used in the delay intervals of the Tracking Test.

Screening Tests

As with Study 1, the WAIS-R Vocabulary subtest, the MMSE, and the GDS were administered to provide descriptive information about the subjects and/or to exclude subjects from the normative sample. Standard administration procedures were used. Readers are referred to Appendix C for a brief description of these tests.

Filler Tasks

A number of "filler" tasks are also used in the

intervals between the acquisition and delay trials on the Paragraph and Triangle Geometric Figure subtests in the Tracking Test. The only previously-published task used here was the subjective age questionnaire (see Appendix C). The rest of the tasks were comprised of recently-developed tests that have not been previously published. These tasks included: a chronological age questionnaire, a 48-item recognition memory task for common line drawings, and a 48-item recognition memory task for common words. These tasks are described in more detail in Appendix D. Use of all test materials has been approved by Carleton University's Ethics Committee (see Appendix E).

Procedure

Before testing began, informed consent was obtained and subjects were administered the MMSE. Subjects were informed that they would be given several opportunities to learn and remember different types of information. The order in which the tests were administered in Study 2 is provided in Appendix G.

The general instructions for the Paragraph and Triangle Geometric Figure subtests of the Tracking Test are essentially the same as those for the subtests of the Diagnostic Test. The only difference is that there is no cued recall portion for the Paragraph subtest.

In the next section, a description of the analyses used in Studies 1 and 2 will be provided.

Analyses

Demographic Information:

Demographic information is reported for each of the six age groups and for the total sample. This demographic information includes the percentage of males and females, as well as the means and standard deviations for age, education, vocabulary, and depression scores. The six age groups were examined for statistically significant differences in education level as well as vocabulary, mental status, and depression scores.

Reliability:

Three types of reliability evidence for the G-LAMB are reported. First, internal consistency was examined for the Paragraph and Geometric Figure subtests. Internal consistency is perhaps the most common form of reliability data. However, it must be noted that there are several difficulties in using such a measure of reliability with learning and memory tests such as the G-LAMB (Aiken, 1994; Schmidt & Tombaugh, 1995). Generally, items on a test are assumed to be of equal importance or weight to the measurement of a construct. In terms of a memory test, recall of items is assumed to be independent; that is, all items are assumed to have an equal likelihood of being recalled. However, the existence of phenomena such as serial position effects (i.e., primacy and recency effects), highly context-dependent items (e.g., "three

ginger cookies", as well as the recall of items imbedded in a story or context generally), and the use of multiple trials all affect the observed reliability coefficients. Thus, the internal consistency coefficients obtained on the G-LAMB must be interpreted with caution. Internal consistency was assessed using Cronbach's alpha (or, more accurately, using KR-20 since dichotomous data were used).

Second, inter-scorer reliability estimates were computed for the two Geometric Figure scoring systems. Reliability estimates were calculated for each of the three learning trials. Because retention scores tend to be highly correlated with performance on the last acquisition trial and the range of scores on the copy trial typically is highly restricted, interscorer reliability on the delay and copy trials was not calculated.

Third, the standard error of measurement (SEM) was calculated for each of the age groups reported in the normative tables as well as for the total sample using the obtained reliability estimates (Murphy & Davidshofer, 1994). Although reliable tests are more accurate than unreliable tests, reliability coefficients only provide a *relative* measure of test score accuracy (i.e., a test with a reliability of .93 is more reliable than a test with a reliability of .86). They do not provide absolute test score accuracy; that is, reliability coefficients do not tell us how much consistency (or alternatively, variability)

we can expect in the test scores. To determine the variability in test scores that can be expected as a result of measurement error, one needs to compute the SEM, which is a function of test score variability and test reliability. As variability increases, so does the SEM; as reliability increases, the SEM decreases.

The advantage of providing the SEM is that it can later be used to calculate the confidence interval or "percentile band" for an individual score (Aiken, 1994; Murphy & Davidshofer, 1994). Because 95% of all scores in a normal distribution fall within 1.96 standard deviations of the mean, a confidence interval around a mean can be computed by multiplying the SEM by 1.96. In the case of an individual score (e.g., 25 out of 31 on Trial 1 of the Paragraph subtest), the SEM can be used to calculate a confidence interval or the expected range of scores that would be obtained 95% of the time if repeated testing was used.

A percentile band is often useful when comparing an individual's performance on two subtests. When a person's scores on several subtests are plotted on a profile sheet, a percentile band can also be plotted. This band consists of one or two SEM around the subtest scores. Generally, the difference between the scores obtained by the same person on two different subtests should be greater than the larger SEM before that difference can be viewed as significant (Aiken, 1994).

Finally, it should be noted that test-retest reliability evidence was not be gathered in this dissertation. Administration of the battery of tests given in each of the two studies takes 1-1/2 to 2 hours per subject. The decision was made not to administer the same test battery twice to each subject for the following two reasons: (a) time constraints given that both Studies 1 & 2 are conducted, and (b) what may be termed "subject burnout" since the same subjects are proposed to be used for both Studies 1 and 2 whenever possible.

It must be pointed out that there are several difficulties in assessing the test-retest reliability of learning and memory tests. One problem is that such reliability estimates may be inflated because of practice effects and the fact that the information is well-learned. It would be difficult to suggest a test-retest interval that would be long enough to counteract this learning. For example, some subjects who participated in a memory study conducted at Carleton University five or six years ago can still recount a fair portion of a story they had to learn and remember. An opposing problem is that the longer one makes the test-retest interval, the greater the chances that an event (e.g., head injury, stroke) can occur in the interim to disrupt normal performance.

Validity:

Three types of validity evidence concerning the G-LAMB

were examined. First, patterns of performance on the Paragraph and Geometric Figure that would be expected according to the literature and with testing experience were examined as evidence of construct validity. That is, it was expected that, on each subtest, performance would (a) improve over trials, (b) decrease with increasing age, and (c) improve with increased retrieval support (e.g., performance would be lowest when a free recall procedure was employed and highest when recognition was used).

Second, correlations were conducted between the Paragraph and Geometric Figure subtests and a number of other tests administered in the test session. For example, convergent validity was examined in Study 1 by correlating the Paragraph subtest with a Word List and the Geometric Figure subtest with an Abstract (Figure) Recognition Memory test. Discriminant validity was examined by correlating the two subtests with tests that are not expected to be highly related to memory performance (e.g., Trails A & B and Word Fluency).

Third, exploratory factor analyses were conducted to examine (a) the factor structure of the Diagnostic and Tracking Tests alone, and (b) the factor structure of the subtests of the Diagnostic and Tracking Tests when combined with several other cognitive and neuropsychological tasks. Based on the factor analytic findings on the LAMB (Schmidt & Tombaugh, 1995) and on the WES-R (Bornstein & Chelune,

1988), it is predicted that the Paragraph and Geometric Figure subtests would load on two factors: verbal and visual learning and memory. Based on the factor analytic findings of Schmidt and Tombaugh (1995), it was also expected that when the Diagnostic and Tracking Tests were factor analyzed with other tests, verbal tasks (e.g., Paragraph, Word List, Word Recognition Memory) would load together and visual tasks (e.g., Geometric Figure, Picture Recognition Memory, Abstract Recognition Memory) would load together on a separate factor. The WAIS-R Vocabulary subtest should load on a separate factor but be strongly correlated with the verbal factor. Other neuropsychological tests should load separately from the memory tasks.

Normative Data:

Before normative data could be reported, it was important to examine the potential effects of variables such as age, education, and gender on the G-LAMB subtest scores. This was accomplished by entering age, education, and gender into a stepwise regression analysis. Norms would then be reported on the basis of the relationship of each of these variables to the subtest scores. For example, if age and education were found to be related to the Paragraph subtest, then norms would be provided for different age and education groups, regardless of gender.

Once the appropriate manner of describing normative data was determined, two sets of norms were provided.

First, total acquisition and delay scores were reported using standard scores with a mean of 100 and a standard deviation of 15. Standard scores were calculated to facilitate comparisons between subscales. Second, information about rates of learning (i.e., trial-by-trial performance) was provided using cumulative percentiles.

Results

Throughout the results section, an alpha level of .05 was used. Post-hoc comparisons were conducted using Tukey's honestly significant difference (HSD) procedure ($p < .05$). For each significance test, a sample-based measure of effect size, eta-squared (η^2), is reported. Whenever the sample size is large, there is an increased chance that small differences will become statistically significant. Effect size is a way of operationalizing the notion of the size or magnitude of an effect and, unlike the p-value, is independent of the sample size. Eta-squared was chosen as the measure of effect size because of its flexibility; that is, unlike other measures of effect size (e.g., omega-squared, Cohen's d), it can be obtained easily from many different types of analyses such as designs involving repeated measures (Tabachnik & Fidell, 1989).

Cohen (1992) has provided a rough guideline for interpreting eta-squared based on a review of empirical studies in abnormal and social psychology. He indicates that a medium effect size (η^2) is equal to approximately .059 to .168) is "likely to be visible to the naked eye of a careful observer" (p. 159). A small eta-squared is equal to approximately .010 to .059, and a large eta-squared is greater than approximately .168. An examination of the effect sizes obtained in Studies 1 and 2 indicate that the results obtained in the analyses were not trivial; that is,

they are not simply an artifact of large sample sizes.

Study 1

Demographic Information on Original Six Age Groups

The demographic data are shown in Table 1. In the overall sample, 46.2% of the participants were male and 53.8% were female. All age groups contained a comparable number of male and female subjects. No significant differences were found among the six different age groups in the number of years of education ($F(5,178) = 0.231$, $p = .949$, $\eta^2 = .006$), on WAIS-R vocabulary scores ($F(5,178) = 1.171$, $p = .325$, $\eta^2 = .032$), or on the Geriatric Depression Scale ($F(5,178) = 1.652$, $p = .149$, $\eta^2 = .044$). There was a statistically significant difference among the scores obtained on the MMSE ($F(5,177) = 3.255$, $p = .01$, $\eta^2 = .084$). Post-hoc analyses showed that MMSE scores differed only between the 60-64 year old group (Mean = 29.143, s.d. = 0.848) and the 80-85 year old group (Mean = 27.968, s.d. = 1.224). However, these differences are not clinically significant because the scores are well within the "normal" range (i.e., MMSE scores between 24 and 30).

Correlations among the demographic variables are presented in Table 2. A low negative correlation was found between age and mental status scores, indicating a tendency for mental status scores to decrease as age increased. Education and vocabulary scores were moderately positively correlated and both of these variables showed a low positive

correlation with mental status scores. That is, high levels of education and high vocabulary scores were associated with each other and with higher scores on the MMSE.

Internal Consistency and Inter-scorer Reliability

Separate internal consistency estimates were calculated for Trial 1, last acquisition trial, and the delay trials for the Paragraph and Rectangle figure subtests. The resulting Cronbach alphas of .84, .85, and .86 for the Paragraph subtest and .75, .84, and .85 for the Rectangle figure subtest showed moderately high levels of internal consistency for these subtests.

Two scorers (the author and a Masters-level graduate student, M.J.) independently scored each subject's geometric figure reproductions. Inter-scorer reliability estimates were obtained for each learning trial. The retention and copy scores were not analyzed because the former scores are very highly correlated with the last acquisition trial scores and the copy scores produce too small a range of scores to be analyzed appropriately. Inter-scorer reliability estimates were as follows: Trial 1 = .99, Trial 2 = .99, Trial 3 = .98 (all with $p < .001$). These estimates provide strong evidence for the inter-scorer reliability of the scoring system for the Rectangle Geometric Figure.

Factor Analyses

Two sets of factor analyses were conducted on the Diagnostic Test. The first factor analysis examined the

factor structure of the Diagnostic Test alone. The second factor analysis allowed for an examination of the convergent and discriminant validity of the Diagnostic Test by factor analyzing it when combined with several other cognitive or neuropsychological tasks.

A principal axis factor analysis with oblique (oblimin) rotation was conducted on the total acquisition, delay, and copy scores for the Paragraph and Rectangle figure subtests. Following Cliff's (1987) suggestion, principal axis factor analysis (PAF) was used rather than other factor analytic techniques or principal components analysis because (a) the goal of conducting the factor analysis is to understand the relationship among the variables, (b) the factors are well-defined and (c) improper communality estimates greater than 1 (i.e., the Heywood case) were likely given that there would not be three substantial loadings for each common factor. Oblique rotation was used instead of orthogonal rotation because it, more realistically, permits correlations among the factors (Hakstian & Abel, 1974).

The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy involves comparing the magnitudes of the observed correlation coefficients to the magnitudes of the partial correlation coefficients (Norusis, 1986). A small KMO value (i.e., < 0.50) indicates that the correlations between pairs of variables cannot be explained by the other variables and thus a factor analysis is not recommended (Kaiser, 1974). A

KMO value of .79 was obtained, indicating that use of a factor analysis was appropriate. Bartlett's test of sphericity was examined to determine whether the correlation matrix differed from an identity matrix (Norusis, 1986). That is, in order to factor analyze the variables, it is necessary that the variables be correlated to some degree with each other. Bartlett's test indicated a value of 1232.362, $p < .001$, which also supports the appropriateness of the factor analysis.

As shown in Table 3, a two-factor solution was obtained. The first factor, which can be labelled verbal memory, accounted for 67.5% of the total variance¹. This factor included both the acquisition and delay trials for the Paragraph subtest. The second factor, visual memory, was responsible for 19.3% of the total variance. This factor was comprised of the acquisition and delay trials for the Geometric Figure subtest. Given the strong memory component to both subtests, it is not surprising that the two factors show a moderate correlation ($r = .48$).

A second factor analysis was conducted to examine the convergent and discriminant validity of the Diagnostic Test. A principal axis factor analysis with oblique (oblimin) rotation was conducted using the Paragraph and Rectangle figure subtests as well as Word List, Trail-Making (Trails A and B), Abstract Recognition Memory, Verbal Fluency (F, A, S, and Animals), and WAIS-R Vocabulary. A KMO value of .83

was obtained, indicating that use of a factor analysis was appropriate. Bartlett's test of sphericity indicated a value of 1909.115, $p < .001$, which also supports the appropriateness of the factor analysis.

As shown in Table 4, a three-factor solution was obtained. This solution was accepted on the basis of three criteria: (a) the eigenvalue-greater-than-one rule, (b) Scree test, and (c) the interpretability of the obtained factor solution (Tabachnick & Fidell, 1989). The first factor, which may be labelled visual processing, accounted for 41.6% of the total variance. This factor included the acquisition and delay trials for the Geometric Figure subtest and the Abstract Recognition Memory test as well as Trails A and B. The second factor, expressive language, accounted for 10.7% of the total variance. It included each of the verbal fluency categories and the WAIS-R Vocabulary subtest. Finally, the third factor, verbal memory, was responsible for 7.1% of the total variance¹. This factor was comprised of the acquisition and delay trials for the Paragraph subtest and the Word List. The three factors showed low to moderate correlations with each other. The first and third factors are most highly correlated ($r = .60$) which perhaps reflects a memory component to most of these tasks.

Regression Analyses

Stepwise regression analyses were performed on the

total acquisition scores and on the delay scores for both the Paragraph and Rectangle figure subtests in order to examine the impact of a number of demographic variables and to help clarify on which variables the normative data should be stratified. The variables of interest included age, years of education, and gender. Vocabulary level was not included because it would not serve as a practical variable on which to stratify the normative data as it would necessitate the administration of another test (i.e., WAIS-R Vocabulary) for the norms to be used. Notably however, vocabulary is moderately correlated with years of education ($r = .48$, $p < .001$) which is included in the analysis. Stepwise regression was chosen over other regression methods because the interest was simply in determining (a) which variables would statistically significantly predict subtest performance and (b) the amount of variance that would be explained by each variable. Hierarchical regression, in particular, was not used as there was no theoretical reason to enter the variables in a specific order.

The results of the stepwise regression analyses are shown in Table 5. Performance on both the total acquisition and the delay scores on the Paragraph subtest was predicted by, in order of explained variance, age, education, and gender. This finding was also supported by the semi-partial correlation (r_{sp}) results reported in Table 5. 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_p for a variable, the greater the relative importance of that variable. In terms of the Rectangle figure subtest, performance on both the total acquisition and the delay scores was predicted by, in order of explained variance, age and education. Again, this finding was supported by the semi-partial correlation results. Performance on the Rectangle figure subtest was not statistically significantly predicted by gender.

The results of the stepwise regression analyses suggest that normative data should be stratified on the basis of age, years of education, and gender for the Paragraph subtest, and on the basis of age and education for the Rectangle figure subtest. However, age alone will be used to stratify the norms because splitting the sample into too many groups makes the number of subjects within each group unacceptably low. In addition, age was chosen because it enters first for each subtest and the contributions of education and gender were relatively small.

Demographic Information on Final Three Age Groups

The next step in the analyses was to determine whether norms are best reported for each of the six age groups or if the data could be represented using fewer age groups. Several criteria were considered in making this decision and are presented in an approximate order of importance. First, the same age stratification should be used for both

subtests. This reduces the possibility that a person (e.g., clinician) will use the wrong norms when referring from one subtest to the other. Second, age effects within the selected age groups should be minimized. That is, the majority of subjects within an age group should be well-represented by the normative data reported for that group. Third, there should be no significant (i.e., meaningful) differences between the age groups on the demographic variables. This reduces the possibility that differences in performance on the subtests could be due to factors other than age. Fourth, the sample sizes within the age groups should consist of at least 50 subjects to ensure an adequate representation of that age group. Finally, differences in performance on the subtests 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.

Visual inspection of the data presented in Tables 6 & 7 suggested that collapsing the data into fewer age groups would be beneficial because performance on the subtests was highly similar for some of the original six age groups. Several ways of collapsing the age groups were suggested by natural splits that occurred in the performance levels of these six groups. Three different types of age groupings were considered: (a) two 15-year age groups (i.e., 55-69 yrs and 70-85 yrs), (b) one 20-year group and one 10-year group

(i.e., 55-74 yrs and 75-85 yrs), and (c) three 10-year groups (i.e., 55-64 yrs, 65-74 yrs, and 75-85 yrs). A comparison of these proposed age groupings found that each grouping met the criteria for sample size and for no meaningful differences on the demographic variables. Although the first two age groupings showed more significant between-group differences in subtest performance, the age grouping with the three 10-year groups showed the fewest within-group age effects and hence was most representative of the data.

The demographic data for the three age groups are shown in Table 8. No significant differences were found among the three groups in the number of years of education ($F(2,181) = 0.069$, $p = .933$, $\eta^2 = .001$), WAIS-R vocabulary scores ($F(2,181) = 0.017$, $p = .984$, $\eta^2 = .0002$), or on the Geriatric Depression Scale ($F(2,181) = 0.586$, $p = .558$, $\eta^2 = .006$). There was a statistically significant difference among the scores obtained on the MMSE ($F(2,180) = 6.161$, $p < .01$, $\eta^2 = .064$). However, follow-up analyses indicated that the only statistically significant difference was between the 55-64 year old group (Mean = 28.883, s.d. = 1.121) and the 75-85 year old group (Mean = 28.098, s.d. = 1.338). Clearly, this difference is not very meaningful from a practical point of view since both scores are well within the "normal" range (i.e., MMSE scores between 24 and 30).

Performance on the Paragraph Subtest over Trials

Table 9 shows the mean trial-by-trial scores on the Paragraph subtest. An Age x Trials repeated measures analysis of variance (ANOVA) was conducted and revealed a significant within-subjects effect for Trial ($F(1,181) = 760.14, p < .001, \eta^2 = .808$), and a significant between-groups effect for Age ($F(2,181) = 13.00, p < .001, \eta^2 = .126$). The Age x Trials interaction was significant ($F(2,181) = 4.75, p < .01, \eta^2 = .105$), indicating that the rate of learning over the trials differed for the three age groups.

Post-hoc analysis indicated that on both trials, the 55-64 year old group and the 65-74 year old group performed at statistically significantly higher levels than the 75-85 year old group. An examination of the data in Table 9 indicates that the significant interaction appears to be due to the greater improvement in performance by the oldest age group relative to their (lower) Trial 1 score. This greater level of improvement by the oldest group also probably reflects a ceiling effect of performance on Trial 2 for the two younger age groups.

An examination of the free recall delay scores on the Paragraph showed there was a statistically significant difference between the three age groups ($F(2,181) = 10.655, p < .001, \eta^2 = .105$) with the 55-64 year old group (Mean = 26.197, s.d. = 3.999) and the 65-74 year old group (Mean =

25.113, s.d. = 4.549) performing at significantly higher levels than the 75-85 year old group (Mean = 22.049, s.d. = 6.553). Similar results were found for the free recall + cued recall delay scores ($F(2,181) = 7.890, p < .001, \eta^2 = .080$) and for free recall + cued recall + recognition delay scores ($F(2,181) = 7.973, p < .001, \eta^2 = .077$). In order to control for differences in performance on delayed recall that might result from differences in the amount of information learned, delay scores were converted to a "proportion-remembered" (free recall delay/trial 2) score and reanalyzed. As with the analysis conducted on the raw scores, there was a statistically significant difference between the three age groups ($F(2,181) = 6.129, p < .01, \eta^2 = .063$) with the 55-64 year old group (Mean = .923, s.d. = .101) and the 65-74 year old group (Mean = .918, s.d. = .083) recalling a greater proportion of learned material than did the 75-85 year old group (Mean = .862, s.d. = .133).

Performance of the Rectangle Figure Subtest over Trials

Table 10 shows the mean trial-by-trial scores on the Rectangle Figure subtest. An Age x Trials repeated measures ANOVA was conducted and revealed a significant within-subjects effect for Trial ($F(6, 108) = 69.23, p < .001, \eta^2 = .784$), and a significant between-groups effect for Age ($F(2,181) = 21.03, p < .001, \eta^2 = .119$). The Age x Trials interaction was significant ($F(12, 362) = 5.15, p < .001$).

2

PM-1 3½"x4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 10:10a ANSI/ISO #2 EQUIVALENT

PRECISIONSM RESOLUTION TARGETS

.001, $\eta^2 = .054$), indicating that the rate of learning over the trials differed for the three age groups.

Post-hoc analysis indicated that, on Trial 1, all three groups performed at significantly different levels from each other, whereas on Trials 2 and 3, the 75-85 year old group performed at a significantly lower level than both the 55-64 year old and 65-74 year old groups. This pattern of results can be seen clearly in Table 10 and appears to be due to the greater improvement in performance by the oldest group relative to their (much lower) Trial 1 score. However, this finding also seems to reflect the ceiling effect evident by Trial 3 for the 55-64 year old and 65-74 year old groups.

In order to determine the degree to which the poorer performance of older subjects might reflect a perceptual or constructional effect rather than a memory effect, all trials were converted to "proportion of copy" scores and reanalyzed. The results did not change (Age: $F(2,181) = 19.46$, $p < .001$, $\eta^2 = .177$; Trials: $F(2,362) = 660.18$, $p < .001$, $\eta^2 = .785$; Age $\times$ Trials: $F(4,362) = 6.13$, $p < .001$, $\eta^2 = .064$). Thus, the lower performance by the older group cannot be attributed to a decrease in constructional abilities. In addition, the same pattern of results were found in the post-hoc analysis. That is, on Trial 1, the three age groups differed significantly from one another, whereas on Trials 2 and 3, only the 75-85 year old group differed significantly from the other two groups.

An examination of the delay scores on the Rectangle figure showed that there was a statistically significant difference among the three age groups ($F(2,181) = 13.291$, $p < .001$, $\eta^2 = .128$). Post-hoc analysis showed that the 55-64 year old group (Mean = 23.812, s.d. = 3.405) and the 65-74 year old group (Mean = 23.057, s.d. = 3.401) performed at a significantly higher level than the 75-85 year old group (Mean = 20.254, s.d. = 5.034).

In order to control for differences in performance on delayed recall that might result from differences in the amount of information learned, delay scores were converted to a "proportion-remembered" (free recall delay/trial 3) score and reanalyzed. This analysis indicated that there was no statistically significant difference in the proportion of learned material that was recalled by the three age groups ($F(2,181) = 2.337$, $p = .100$, $\eta^2 = .025$). Mean proportions recalled ranged from .965 to .989.

A statistically significant difference was found among the three age groups on their copy scores ($F(2,181) = 5.837$, $p < .01$, $\eta^2 = .061$), with the 55-64 year old group (Mean = 25.672, s.d. = 0.870) and the 65-74 year old group (Mean = 25.532, s.d. = 0.882) performing at slightly higher levels than the 75-85 year old group (Mean = 25.016, s.d. = 1.489).

Normative Data

Normative data for the Paragraph and Rectangle figure subtests are reported separately for each of the three age

groups: 55-64 years, 65-74 years, and 75-85 years. As is the convention for tests of learning and memory, standard scores (Mean = 100, s.d. = 15) are reported for the total acquisition (see Table 11) and delay (see Table 12) scores. Because the distribution of scores in Study 1 is not normally distributed (i.e., it is negatively skewed), a nonlinear transformation was used to calculate the standard scores.² That is, standard scores were obtained by first computing percentiles for each raw score and then converting the percentiles into standard scores. Such a procedure is analogous to that described in the calculation of "normalized" standard scores (see Aiken, 1994, p. 85).

Normative data for the trial-by-trial scores are reported separately for each age group using cumulative percentiles (see Tables 13-18). A cumulative percentile is interpreted as the percentage of scores that fall at or below a given raw score. For example, if a raw score of 21 on the Rectangle figure subtest falls at the 90th cumulative percentile, then one can say that 90% of the normative sample scored 21 or less on that subtest. Cumulative percentiles are used rather than percentiles (which refer to the percentage of scores that fall below a given raw score) because they represent the data better when the score distribution is skewed and/or truncated. A key factor in a learning and memory test is the difficulty level of the task. When the difficulty level is too low or too high, the

score distribution can become both skewed and truncated. This becomes an issue in the present study because performance increases (i.e., task difficulty decreases for the subject) as the number of learning trials in the subtests increases. This is also the case in a relatively easy task such as the copy trial of the Rectangle figure subtest. To demonstrate the difference between cumulative percentiles and percentiles, let us consider the normative performance of the 65-74 year old group on the copy trial of the Rectangle figure subtest (maximum score = 26) as shown below:

<u>raw score</u>	<u>cumulative percentile</u>	<u>percentile</u>
26.0	100	65
25.5	66	43
25.0	31	20
24.5	21	13
24.0	10	6
23.5	8	5
23.0	7	4
22.5	5	3
22.0	3	2

When the score distribution is normally distributed with considerable variability among the scores, the cumulative percentiles and percentiles can be almost identical. However, as can be seen in the present case, a perfect score of 26 places a person at the 100th *cumulative percentile*, but only at the 65th percentile!

An estimate of how widely scores might vary for an individual tested an infinite number of times is the standard error of measurement (s.e.m.). The s.e.m. is

reported for each trial of both the Paragraph and the Triangle figure subtests for each of the three age groups as well as the total sample (see Table 19). It is calculated using the following formula,

$$\text{s.e.m.} = s \sqrt{1 - r_{11}} ,$$

where s is the standard deviation of the raw scores and r_{11} is the measure of reliability (e.g., test-retest reliability coefficient, internal consistency coefficient) (Aiken, 1994, p. 92).

To illustrate the utility of the s.e.m., let us consider the example of a 60 year old person who scores 23 out of 26 on Trial 1 of the Rectangle Figure subtest. The s.e.m. is assumed to be normally distributed, meaning that approximately 68 times out of 100 the true score falls between ± 1 s.e.m., and approximately 95 times out of 100 it falls between ± 2 s.e.m. Thus, using the s.e.m. of 2.34 provided in Table 19, it can be concluded with 68% confidence that with an observed score of 23, this person's true score falls between 20.66 and 25.34.

Study 2

Demographic Information on Original Six Age Groups

The demographic data are shown in Table 20. In the overall sample, 46.1% of the participants were male and 53.9% were female. All age groups, except the 60-64 year old group, contained comparable numbers of male and female

subjects. No significant differences were found between the six different age groups in the number of years of education ($F(5,174) = 0.288$, $p = .920$, $\eta^2 = .008$) or WAIS-R vocabulary scores ($F(5,174) = 0.309$, $p = .907$, $\eta^2 = .009$). There was a statistically significant difference among the scores obtained on the MMSE ($F(5,174) = 2.789$, $p < .05$, $\eta^2 = .074$). Post-hoc analyses showed that the 60-64 year old group (Mean = 29.100, s.d. = 0.885) differed statistically significantly from the 80-85 year old group (Mean = 27.966, s.d. = 1.401). A statistically significant difference was also found on the Geriatric Depression Scale ($F(5,174) = 2.920$, $p < .05$, $\eta^2 = .077$). Post-hoc analyses showed that the 60-64 year old group (Mean = 2.567, s.d. = 2.402) differed statistically significantly from both the 65-69 year old group (Mean = 4.750, s.d. = 2.874) and the 80-85 year old group (Mean = 5.276, s.d. = 2.776). However, the differences found on both the MMSE and the GDS are not clinically significant because scores on both tests are well within the "normal" ranges (i.e., MMSE scores between 24 and 30; GDS scores between 0 and 10).

Correlations among the demographic variables are presented in Table 21. The results were similar to those found in Study 1. That is, a low negative correlation was found between age and mental status scores, indicating a tendency for mental status scores to decrease as age increased. Education and vocabulary scores were moderately

positively correlated and both of these variables showed a low positive correlation with mental status scores.

Internal Consistency and Inter-scorer Reliability

Separate internal consistency estimates were calculated for Trial 1, last acquisition trial, and the delay trials for each of the Paragraph and Triangle figure subtests. The resulting Cronbach alphas of .70, .72, and .72 for the Paragraph subtest and .69, .71, and .76 for the Triangle figure subtest showed moderate levels of internal consistency for these subtests.

Two scorers (the author and a Masters-level graduate student, M.M.) independently scored each subject's geometric figure reproductions. Inter-scorer reliability estimates were obtained for each learning trial. The retention and copy scores were not analyzed because the former scores are very highly correlated with the last acquisition trial scores and the copy scores produce too small a range of scores to be analyzed appropriately. Inter-scorer reliability estimates were as follows: Trial 1 = .98, Trial 2 = .98, Trial 3 = .95 (all with $p < .001$). These estimates provide strong evidence for the inter-rater reliability of the scoring system for the Triangle Geometric Figure.

Factor Analyses

Two sets of factor analyses were conducted with the Tracking Test. The first factor analysis examined the factor structure of the Tracking Test alone. The second

factor analysis allowed for an examination of the convergent and discriminant validity of the Tracking Test by factor analyzing it along with several other cognitive tasks.

Table 22 shows the results of a principal axis factor analysis with oblique (oblimin) rotation conducted on the total acquisition and delay scores from both the Paragraph and the Triangle Geometric Figure subtests, as well as the copy score for the figure. The principal axis factor analysis and oblique rotation were used in this study for the same reasons as in Study 1. A KMO value of .69 was obtained, indicating that use of a factor analysis was appropriate. An examination of Bartlett's test of sphericity showed a value of 521.257, $p < .001$ was obtained and also supports the appropriateness of using factor analysis.

As shown in Table 22, a two-factor solution was obtained. The first factor, which can be labelled verbal memory, accounted for 53.8% of the total variance. It was comprised of the acquisition and delay trials for the Paragraph subtest. The second factor, visual memory, was responsible for 21.8% of the total variance¹. This factor included the acquisition and delay trials for the Geometric Figure subtest. The finding that the two factors were moderately correlated ($r = .41$) probably reflected the strong memory component to both of these subtests.

A second factor analysis was conducted to examine the

convergent and discriminant validity of the Tracking Test. A principal axis factor analysis with oblique (oblimin) rotation was conducted using the Paragraph and Triangle figure subtests as well as Word Recognition Memory, Picture (Common Objects) Recognition Memory, and WAIS-R Vocabulary. A KMO value of .72 was obtained, indicating that use of a factor analysis was appropriate. Bartlett's test of sphericity indicated a value of 841.787, $p < .001$, which also supports the appropriateness of the factor analysis.

A four-factor solution was obtained and is shown in Table 23. The criteria used to accept the factor solution are described in the factor analysis section of Study 1. The first factor, which can be labelled contextual verbal memory, accounted for 36.0% of the total variance. Both the acquisition and delay trials for the Paragraph subtest as well as WAIS-R Vocabulary loaded on this factor. However, it is recommended that only variables with loadings of .30 and above are interpreted (Tabachnik & Fidell, 1989). Thus, although vocabulary loaded on this first factor, it is not included in the interpretation of the factor. The second factor, labelled abstract visual memory, accounted for 12.8% of the total variance. It was comprised of the acquisition and delay trials for the Geometric Figure subtest. The third factor, concrete verbal recognition, was responsible for 9.6% of the total variance and consisted of the acquisition and delay trials for the Word Recognition Memory

test. Finally, a fourth factor, concrete visual recognition, accounted for 6.4% of the total variance'. It consisted of the acquisition and delay trials for the Picture Recognition Memory test. Correlations among the factors ranged from .22 to .43.

Regression Analyses

Stepwise regression analyses were performed on the total acquisition scores and on the delay scores for both the Paragraph and Triangle figure subtests in order to examine the impact of a number of demographic variables and to help clarify on which variables the normative data should be stratified. The variables of interest included age, years of education, and gender. As in Study 1, vocabulary level was not included in this analysis because it is not a practical variable upon which to stratify the norms. Stepwise regression was chosen for the same reasons as in Study 1.

The results of the stepwise regression analyses are shown in Table 24. Performance on the total acquisition scores on the Paragraph subtest was predicted by, in order of explained variance, age, gender, and education. Delay scores on the Paragraph subtest were predicted by, in order of explained variance, education and age. These findings were also supported by the semi-partial correlation (r_p) results reported in Table 24. In terms of the Triangle figure subtest, performance on the total acquisition scores

was predicted by, in order of explained variance, age and education, whereas the delay scores were predicted by, in order of explained variance, age and gender. These findings were also supported by the semi-partial correlation results. Overall, the results of the stepwise regression analyses show that age contributes significantly to performance on both subtests, albeit more so on the Triangle figure subtest. The contributions of both education and gender tend to be small and inconsistent. Thus, only age will be used to stratify the sample.

Demographic Information on Final Three Age Groups

The next step in the analyses was to determine whether norms are best reported using the six original age groups or if the data would be better represented by using fewer age groups. As in Study 1, several criteria were considered: (a) the same age stratification should be used for both subtests, (b) age effects within the selected age groups should be minimized, (c) there should be no significant (i.e., meaningful) differences between the age groups on the demographic variables, (d) the sample sizes within the age groups should consist of at least 50 subjects, and (e) differences in performance on the subtests by the selected age groups would be desirable.

Visual inspection of the data (see Tables 25 & 26) suggested that collapsing the data into fewer age groups would be beneficial because performance on the subtests was

highly similar for some of the original six age groups. Inspection of the data also suggested two different ways of collapsing the age groups: (a) two 15-year age groups (i.e., 55-69 yrs and 70-85 yrs), and (b) three 10-year groups (i.e., 55-64 yrs, 65-74 yrs, and 75-85 yrs). A comparison of these proposed age groupings found that each grouping met the criteria for sample size and for no meaningful differences on the demographic variables. Although the first age grouping showed more significant between-group differences in subtest performance, the age grouping with the three 10-year groups showed the fewest within-group age effects. In addition, using the same age groupings as in the Diagnostic Test was appealing as it simplifies use of the normative data.

The demographic data for the three age groups are shown in Table 27. No significant differences were found among the three groups in number of years of education ($F(2,177) = 0.031$, $p = .970$, $\eta^2 = .0003$), or WAIS-R vocabulary ($F(2,177) = 0.351$, $p = .705$, $\eta^2 = .004$). There was a statistically significant difference among the scores obtained on the MMSE ($F(2,177) = 5.598$, $p < .01$, $\eta^2 = .060$), with the 55-64 year old group obtaining slightly higher scores (Mean = 28.895, s.d. = 1.205) than the 75-85 year old group (Mean = 28.119, s.d. = 1.391). Clearly, this difference is not very meaningful from a practical point of view since both scores are well within the "normal" range

(i.e., MMSE scores between 24 and 30). A statistically significant difference was also found among the scores obtained on the GDS ($F(2,177) = 4.326$, $p < .05$, $\eta^2 = .047$), with the 55-64 year old group obtaining a lower mean score (Mean = 3.386, s.d. = 3.195) than the 75-85 year old group (Mean = 4.966, s.d. = 2.853). This difference is also not clinically significant because both scores are well within the "normal" range (i.e., GDS scores between 0 and 10).

Performance on the Paragraph Subtest over Trials

Table 28 shows the mean trial-by-trial scores on the Paragraph subtest. An Age $\times$ Trials repeated measures ANOVA was conducted and revealed a significant within-subjects effect for Trial ($F(1,177) = 237.84$, $p < .001$, $\eta^2 = .573$). An examination of Table 28 shows that performance increased from Trial 1 to Trial 2. There was also a significant between-groups effect for Age ($F(2,177) = 6.41$, $p < .001$, $\eta^2 = .068$). Post-hoc analysis indicated that acquisition on the Paragraph subtest was significantly higher for the 55-64 year old group (Mean = 37.895, s.d. = 3.144) than for both the 65-74 year old group (Mean = 35.484, s.d. = 5.108) and the 75-85 year old group (Mean = 35.068, s.d. = 5.156). The Age $\times$ Trials interaction was not statistically significant ($F(2,177) = 2.55$, $p = .081$, $\eta^2 = .028$).

An examination of the free recall delay scores on the

Paragraph showed that there was a statistically significant difference on performance among the three age groups ($F(2,177) = 4.149$, $p < .05$, $\eta^2 = .045$). Post-hoc analyses indicated that the 55-64 year old group (Mean = 19.000, s.d. = 2.018) performed at a significantly higher level than the 75-85 year old group (Mean = 17.678, s.d. = 2.921). Similar results were found with the recognition delay scores ($F(2,177) = 5.704$, $p < .01$, $\eta^2 = .061$). A statistically significant difference was found between the 55-64 year old group (Mean = 20.877, s.d. = 0.331) and the 75-85 year old group (Mean = 20.458, s.d. = 0.837) on these scores; however, such a small difference would not seem to be clinically significant.

In order to control for differences in performance on delayed recall that might result from differences in the amount of information learned, delay scores were converted to a "proportion-remembered" (free recall delay/trial 2) score and reanalyzed. This analysis indicated that there were no statistically significant differences in the proportion of learned material that was recalled by the three age groups ($F(2,177) = 1.111$, $p = .331$, $\eta^2 = .012$). Mean proportions recalled ranged from .923 to .950.

Performance on the Triangle Figure Subtest over Trials

Table 29 shows the mean trial-by-trial scores on the Triangle Figure subtest. An Age $\times$ Trials repeated measures ANOVA was conducted and revealed a significant within-

subjects effect for Trial ($F(2,354) = 367.42, p < .001, \eta^2 = .675$), and a significant between-groups effect for Age ($F(2,177) = 19.28, p < .001, \eta^2 = .179$). The Age $\times$ Trials interaction was significant ($F(4,354) = 3.09, p < .05, \eta^2 = .034$), indicating that the rate of learning over the trials differed for the three age groups.

Post-hoc analysis indicated that on Trial 1, all three age groups performed at significantly different levels from each other, whereas on Trials 2 and 3, the 75-85 year old group performed at a significantly lower level than both the 55-64 year old and the 65-74 year old groups. This pattern can be seen clearly in Table 29, and appears to be due to the somewhat greater improvement in performance for the oldest group relative to their (lower) Trial 1 score. However, this finding also reflects a clear ceiling effect on Trial 3 for the two younger groups.

As with Study 1, all trials were converted to "percentage of copy" scores and reanalyzed in order to determine the degree to which the poorer performance of older subjects might reflect a perceptual or constructional effect rather than a memory effect. The overall results did not change (Age: $F(2,177) = 18.27, p < .001, \eta^2 = .171$; Trials: $F(2,354) = 365.64, p < .001, \eta^2 = .674$; Age $\times$ Trials: $F(4,354) = 3.18, p < .05, \eta^2 = .037$). Thus, the lower performance by the older group cannot be attributed to a decrease in constructional abilities. However, in the

follow-up analysis, the 75-85 year old group performed at a lower level than the other two age groups on all three trials.

An examination of the delay scores on the Triangle figure showed that there was a statistically significant difference among the three age groups ($F(2,177) = 13.300$, $p < .001$, $\eta^2 = .131$). Post-hoc analyses indicated that the 55-64 year old group (Mean = 25.202, s.d. = 1.706) and the 65-74 year old group (Mean = 24.086, s.d. = 3.234) performed at a significantly higher level than the 75-85 year old group (Mean = 22.492, s.d. = 3.254). In order to control for differences in performance on delayed recall that might result from differences in the amount of information learned, delay scores were converted to a "proportion-remembered" (free recall delay/trial 3) score and reanalyzed. This analysis indicated that there was no statistically significant difference in the proportion of learned material that was recalled by the three age groups ($F(2,177) = 0.255$, $p = .775$, $\eta^2 = .003$). Mean proportions recalled ranged from .988 to .998. There was no statistically significant difference among the three age groups on the copy scores ($F(2,177) = 2.784$, $p = .065$, $\eta^2 = .031$).

Normative Data

Normative data for the Paragraph and Triangle figure subtests are reported separately for each of the three age

groups: 55-64 years, 65-74 years, and 75-85 years. As described in Study 1, standard scores (Mean = 100, s.d. = 15) are reported for the total acquisition (see Table 30) and delay (see Table 31) scores.

Normative data for the trial-by-trial scores are reported using cumulative percentiles (see Tables 32-37). Cumulative percentiles were used for the same reasons as in Study 1.

Finally, the standard error of measurement (s.e.m.), as described in Study 1, is reported for each trial of both the Paragraph and Triangle figure subtests for each of the three age groups and the total sample (see Table 38).

Discussion

The purpose of this dissertation was to develop two tests for the assessment of memory in older adults and to provide normative data using a sample of community-dwelling adults aged 55-85 years. The following four psychometric criteria are used to evaluate and discuss the development of the G-LAMB tests: (1) test rationale, (2) scale reliability, (3) validity, and (4) normative data.

1. Test Rationale

There are two main reasons for assessing memory in older adults. The first, and most common, reason is to determine whether a memory impairment exists. The second reason is to examine or track the learning and memory performance when an individual has already been diagnosed with a memory impairment. The G-LAMB Diagnostic and Tracking Tests were developed to address each of these reasons.

a) Diagnostic Test

The purpose of the Diagnostic Test is to determine whether an older adult is suffering from a memory impairment, and if so, to what degree. The test is comprised of (a) a Paragraph subtest to measure verbal learning and memory, and (b) a Geometric Figure subtest to measure visuo-spatial learning and memory.

A critical factor in developing any memory test is to ensure that the difficulty level permits measurement over a

broad range of ability. To create a test that is difficult enough to demonstrate superior performance, one must have a high performance ceiling or individuals with superior abilities will obtain scores that reflect the maximum possible score rather than show their maximum ability. Within the context of test development, optimally cognitively intact individuals should recall approximately 50-60% of the material on the first trial and approximately 80-90% by the last acquisition trial. Overall learning (i.e., total acquisition) generally should fall somewhere in the 70-80% range. Achieving these optimal values becomes increasingly more difficult as the age range of the test increases. This occurs because performance on memory tests typically decreases with age, thereby producing an overall suppression of the scores. Thus, a test designed to obtain optimal scores with younger adults tends to be too difficult for older adults.

An example of this is seen in the performance of younger and older adults on the LAMB (Tombaugh & Schmidt, 1992; Schmidt & Tombaugh, 1995). Here, the percent of information correctly recalled by 20-29 year old adults on the LAMB Paragraph subtest was 46% on Trial 1 and 80% on the last acquisition trial (Trial 2). Performance averaged over all trials (total acquisition) was 63%. Likewise, on the LAMB Complex Figure subtest younger adults correctly recalled 58% of the figure on Trial 1 and 91% by the last

acquisition trial (Trial 4). Overall learning was 77%.

Much lower performance is found with older adults.

Performance by 70-74 year old adults on the LAMB Paragraph subtest was only 21% on Trial 1 and 52% by the last acquisition trial (Trial 2). Overall learning was 37%. On the LAMB Complex Figure subtest, older adults recalled 32% of the figure on Trial 1 and 60% by the last acquisition trial (Trial 4). Overall learning was 47%.

Although the difficulty level of the LAMB provides ample opportunity to examine superior performance by some older adults, there is very little opportunity to detect memory impairment. Because performance was almost entirely restricted to the lower half of the range of possible scores, floor effects became a problem. That is, the performance by cognitively *intact* older adults is so low that it is difficult to assess the performance of memory *impaired* adults.

One solution to this problem is to create a test on which cognitively intact older adults correctly recall approximately 50-60% of the task material on the first trial and approximately 80-90% by the last acquisition trial. This allows performance to fall along the entire range of test scores.

Two different strategies were used in the Diagnostic Test to achieve the appropriate level of test difficulty and obtain the desired level of performance by cognitively

intact older adults. First, a self-read procedure allowing for review of the test materials was used with the Paragraph subtest. Second, a relatively simple geometric design (i.e., one with fewer components than is typical) was used in the Geometric Figure subtest.

Diagnostic Paragraph:

An examination of the performance by older adults on the Diagnostic Test Paragraph showed that the use of a self-read procedure achieved the goal of obtaining appropriate levels of recall. For example, 70-74 year old adults recalled 64% of the paragraph on Trial 1 and 88% on the last acquisition trial (Trial 2). Overall learning was 76%. Although the percent correctly recalled was slightly higher than the goal of 50-60% on the first trial, the last acquisition trial and total acquisition performances were well within the targets of 80-90% and 70-80%, respectively.

It was also expected that the use of the self-read procedure would produce two other effects. First, the performance of the older adults on the Diagnostic Test should exceed that of the older adults on the LAMB. Second, the performance of the older adults on the Diagnostic Test should be similar to the performance of the younger adults on the LAMB. The following table contains comparative data for younger and older subjects on the LAMB Paragraph and for older subjects on the Diagnostic Paragraph subtest.

	LAMB Paragraph		Diagnostic Paragraph
	20-29 yrs	70-74 yrs	70-74 yrs
Trial 1	46%	21%	64%
Trial 2	80%	52%	88%
Tot. Acq.	63%	37%	76%

The finding that performance by the 70-74 year old adults on the Diagnostic Paragraph greatly exceeds that of the 70-74 year old adults on the LAMB Paragraph on Trial 1, Trial 2 (last acquisition trial), and on overall learning clearly shows that the goal of increased performance by older adults on the Diagnostic Test was achieved.

A comparison of the performance by older adults on the Diagnostic Paragraph and younger adults on the LAMB Paragraph yielded variable results. On Trial 1, the performance of older adults on the Diagnostic Paragraph was considerably higher than that of younger adults on the LAMB Paragraph (64% vs 46%). However, by the last acquisition trial, the performance of the older adults was only slightly higher than that of the younger adults (88% vs 80%). Thus, a similarity in performance between the younger and older adults on these tests was achieved by the last acquisition trial, but not on the initial learning trial.

The increased performance of the older adults on the Diagnostic Test Paragraph relative to that of the younger adults on the LAMB Paragraph appears to be attributable to some aspect of the self-read procedure. The length and organization of the two paragraphs should not be a factor as

both paragraphs contain the same number of components (31) and are similar in organizational complexity.

There are several reasons why the self-read procedure might result in improved performance by older adults. Although these reasons probably influence performance through their interactions with each other, they will be described separately.

First, a self-read procedure likely permits the use of a relevant strategy for verbal recall (i.e., the ability to view the materials), thereby increasing the face validity of the test and improving motivation. Older adults are much less willing than younger adults to take part in a task simply because it is presented before them (Erickson & Scott, 1977). This makes the face validity, or the appearance that a task is measuring what it claims to measure, particularly important when testing older adults. A memory test using paragraph recall reminds people of trying to remember information from a newspaper or book. However, because most people read such materials *themselves*, they may view a paragraph test that is only presented auditorally as preventing them from using a critical strategy (i.e., the opportunity to read the material themselves). In fact, during the collection of the normative data on the LAMB, a number of older adults argued that they would remember more if they could read the material themselves rather than simply listen to the tester

read it. Certainly the results found here support this argument.

Second, a self-read procedure may produce a feeling of greater control over the testing experience by the individual being tested. Any testing situation can be anxiety provoking, particularly if the person feels he or she has a lot at stake. A self-read procedure makes the task seem more manageable because the person is given the opportunity to use the exposure time to review or study the material. This, combined with the increased face validity of the task may result in higher levels of self-confidence and/or lower levels of anxiety during testing.

Third, the success of the self-read procedure for the older adults might simply be due to increased exposure time to the paragraph. In both the LAMB and the Diagnostic Test, the paragraph is read outloud by the tester. However in the latter test, the *addition* of a 30 second self-read procedure allows further exposure to the test materials. This additional time also permits an individual to use whatever skills or strategies are available to him or her (e.g., increased rehearsal, greater organization of material) and likely decreases anxiety. This argument is consistent with other research that has found that a self-read procedure that permits the individual to review the materials results in significantly better recall than simply having an individual read the materials themselves once (Taub, 1976;

Taub & Kline, 1978; Verhaeghen, Marcoen, & Goossens, 1993).

Finally, another advantage of the self-read procedure is that it removes the divided attention component that occurs when the paragraph is only presented auditorally (Corgiat, Templer, & Newell, 1989; Craik & Simon, 1980; Hartley, Harker, & Walsh, 1980). During an auditory presentation, the individual must listen, process and organize information simultaneously. A self-read procedure permits the to-be-remembered information to be present during acquisition so that the organization of information need not occur simultaneously.

Diagnostic Geometric Figure:

An examination of the performance by older adults on the Geometric Figure also showed that appropriate levels of recall were obtained for this group. The percent correctly recalled by 70-74 year old adults was 60% on Trial 1 and 88% on the last acquisition trial (overall learning = 75%). These results are all well within the goal of 50-60% correctly recalled task material on the first trial, 80-90% by the last acquisition trial, and between 70-80% on total acquisition.

It was also expected that, because of the use of fewer components on the Geometric Figure subtest relative to the LAMB Complex Figure, the performance of the older adults on the Diagnostic Test would both exceed that of the older adults on the LAMB and would be similar to the performance

of the younger adults on the LAMB. The following table permits a comparison of the percent correctly recalled by younger and older adults on the LAMB Complex Figure to that of the older adults on the Diagnostic Geometric Figure subtest.

	LAMB Complex Figure ¹		Diagnostic Rectangle
	20-29 yrs	70-74 yrs	70-74 yrs
Trial 1	58%	32%	60%
Trial 3	91%	60%	88%
Tot. Acq.	77%	47%	75%

It is clear that the performance of the 70-74 year old adults on the Diagnostic Geometric Figure greatly exceeds that of the 70-74 year old adults on the LAMB Complex Figure on Trial 1, Trial 3 (last acquisition trial), and on overall learning (total acquisition). Thus, the Diagnostic Geometric Figure achieved its goal of increasing performance for older adults. In addition, a comparison of the performance by older adults on the Diagnostic Geometric Figure and younger adults on the LAMB Complex Figure showed that the performance of the two groups was remarkably similar on Trial 1, Trial 3 (last acquisition trial for the Diagnostic figure), and overall learning (total acquisition).

The increased performance of the older adults on the Diagnostic Test Geometric Figure, compared to that of the younger adults on the LAMB Complex Figure, appears to be

attributable to the fewer number of components in that figure. All other aspects of this test are the same, including the use of geometric designs and an intentional learning paradigm. The LAMB Complex Figure is composed of 18 components whereas the Diagnostic Geometric Figure has only 13 components. Fewer components would not only provide less information to encode, but given the same amount of exposure time (30 seconds), would allow proportionally more time to rehearse each component and its location. In addition, because the fewer components would make the figure less overwhelming upon initial exposure, older adults might feel less anxiety and frustration and greater motivation during the task.

b) Tracking Test

The purpose of the Tracking Test is distinctly different from that of the Diagnostic Test. Whereas the Diagnostic Test was developed to determine whether an older adult has a memory impairment, the Tracking Test was designed to assess the degree of memory impairment in an older adult already diagnosed as cognitively impaired. Like the Diagnostic Test, the Tracking Test is also comprised of (a) a Paragraph subtest to measure verbal learning and memory, and (b) a Geometric Figure subtest to measure visuo-spatial learning and memory.

With the Diagnostic Test, it was important to develop a test that was difficult enough that it could be used to

assess the range of memory abilities from intact levels to moderate impairment. Thus, care was taken to ensure that an appropriate performance ceiling occurred for cognitively intact older adults. The Tracking Test, however, is designed to be used with older adults who had been previously diagnosed as cognitively impaired. Because the purpose of this test is to differentiate performance at various levels of impairment (e.g., mild, moderate, severe), it is important that this is accomplished without producing floor effects. Thus, an optimal test should have both a low ceiling and low floor. When used with cognitively impaired individuals, the difficulty level should be such that performance will be spread across the full range of scores. In order to achieve this, cognitively intact individuals should perform at a relatively high level; for example, approximately 70-80% of the material should be recalled on the first trial and approximately 90-100% on the last acquisition trial. Overall learning generally should fall somewhere in the 80-90% range. Three other effects should also be evident; performance by older adults on the Tracking Test should be higher than the performance of (a) older adults on the LAMB, (b) younger adults on the LAMB, and (c) older adults on the Diagnostic Test.

Two different strategies were used in the Tracking Test to achieve the appropriate level of test difficulty. First, both a self-read procedure allowing for review of the test

materials and a shorter paragraph (i.e., one with fewer components) were used in the Paragraph subtest. Second, a simple geometric design with the same number of components as the Diagnostic Geometric Figure but organized in a simpler fashion was used in the Geometric Figure subtest.

Tracking Paragraph:

An examination of the performance by older adults on the Paragraph showed that the goal of obtaining appropriate levels of recall was achieved. The percent correctly recalled by 70-74 year old adults was 77% on Trial 1 and 92% on the last acquisition trial (overall learning = 84%). These results are all well within the goal of 70-80% correctly recalled task material on the first trial, 90-100% by the last acquisition trial, and between 80-90% on total acquisition.

It was also expected that the use of the self-read procedure and a shorter paragraph would result in better performance by the older adults on the Tracking Paragraph than the performance of younger and older adults on the LAMB. The following table permits a comparison of the percent correctly recalled by younger and older subjects on the LAMB Paragraph with that obtained by older subjects on the Tracking Paragraph subtest.

	LAMB Paragraph		Tracking Paragraph
	20-29 yrs	70-74 yrs	70-74 yrs
Trial 1	46%	21%	77%
Trial 2	80%	52%	92%
Tot. Acq.	63%	37%	84%

It is clear that the performance of the 70-74 year old adults on the Tracking Paragraph greatly exceeded that of both the younger and older adults on the LAMB Paragraph on Trial 1, Trial 2 (last acquisition trial), and on overall learning (total acquisition). Thus, the Tracking Paragraph achieved its goal of greatly increasing performance by older adults.

Finally, the effect of using a shorter paragraph, over and above the use of a self-read procedure, can be examined by comparing the performance of 70-74 year old adults on the Diagnostic and Tracking Test Paragraphs. Reducing the number of components by approximately one-third (from 31 to 21 components) resulted in an increase in performance by older adults. That is, performance on Trial 1 increased from 64% recalled on the Diagnostic Test to 77% recalled on the Tracking Test. On the last acquisition trial, performance increased from 88% on the Diagnostic Test to 92% on the Tracking Test, whereas overall learning increased from 76% recalled on the Diagnostic Test to 84% recalled on the Tracking Test.

The increased performance of older adults on the

Tracking Test, relative to the performance of younger and older adults on the LAMB Paragraph and older adults on the Diagnostic Paragraph, appears to be attributable to both the self-read procedure and the use of a shorter paragraph. The reasons why the self-read procedure might result in improved performance by older adults were described earlier (see discussion on the Diagnostic Paragraph). The additional improvement in scores seen when a shorter paragraph was used is likely the result of requiring less information to be encoded within the same exposure period as the longer Diagnostic Paragraph (30 seconds). Thus, the task of learning and remembering the paragraph would appear less overwhelming and anxiety-provoking upon initial exposure and might result in greater motivation and confidence. In addition, because there is no change in the exposure period, the individual would have more time to rehearse the material.

Tracking Geometric Figure:

An examination of the performance by older adults on the Tracking Test Geometric Figure also showed that appropriate levels of recall were obtained for this group. The percent correctly recalled by 70-74 year old adults was 67% on Trial 1 and 90% on the last acquisition trial (overall learning = 80%). Although the percent correctly recalled was slightly lower than the goal of 70-80% on the first trial, the last acquisition trial and total

acquisition performances achieved the goals of 90-100% and 80-90%, respectively.

It was also expected that, because the Geometric Figure contained fewer components than other figures (e.g., complex figures such as the LAMB Complex Figure) and because the organization of components within the figure was easier than in the Diagnostic figure, the performance of the older adults on the Tracking Test would exceed that of both the younger and older adults on the LAMB. The following table permits a comparison of the percent correctly recalled by younger and older subjects on the LAMB Complex Figure with that obtained by older subjects on the Tracking Test Geometric Figure.

	LAMB Complex Figure ¹ 20-29 yrs	70-74 yrs	Tracking Triangle 70-74 yrs
Trial 1	58%	32%	67%
Trial 3	91%	60%	90%
Tot. Acq.	77%	47%	80%

The goal of obtaining higher performance on the Tracking Test Geometric Figure by 70-74 year old adults relative to younger and older adults on the LAMB Complex Figure was achieved. For the older adults, the percent recalled on the Tracking Test greatly exceeded the percent recalled on the LAMB on all learning trials. An examination of the performance by older adults on the Tracking Test and younger adults on the LAMB showed that the percent recalled

was considerably higher for the older adults on Trial 1 and for overall learning, but was at essentially the same level by the last acquisition trial.

The increased performance of the older adults on the Tracking Test Geometric Figure, compared to that of the younger and older adults on the LAMB Complex Figure, appears to be attributable to the fewer number of components in that figure, as well as to the organization of components within the figure. The impact that fewer components would have on learning the figure was described earlier (see discussion on Diagnostic Geometric Figure). A more subtle difference between the two figures has to do with their organization. The components of the Diagnostic Test figure tend to be more spread-out (e.g., the circle to the far left of the figure and the square to the far right of the figure) and less balanced (e.g., only the top half of the main rectangle is bisected, arrow is in upper right corner of square). The components of the Tracking Test figure, however, tend to be closer together and more centrally located within the tripartite structure of the main triangle (e.g., circle centred in top portion of main triangle, horizontal line centred within circle).

Despite the rather subtle difference in the difficulty level of the two figures, it was expected that the percent recalled by older adults on the latter figure would exceed that of older adults on the Diagnostic Test figure. In

fact, older adults did tend to perform better on the Tracking Test than on the Diagnostic Test. Overall learning by 70-74 year olds increased from 75% recalled on the Diagnostic Test figure to 80% recalled on the Tracking Test figure. Most of this difference came from the increase in performance on Trial 1 from 60% recalled on the Diagnostic Test figure to 67% recalled on the Tracking Test figure. However, by Trial 3, the percent recalled was almost identical for the two figures (88% vs 90%).

One reason why the organization of the Tracking Test Geometric Figure might result in increased performance is that the placement of the components may be less difficult to encode and remember. By locating many of the components in central locations, the individual has one less distinct feature to recall. Perhaps some individuals notice this feature of the figure and are able to incorporate it into their encoding strategy. Alternatively, people may have a natural tendency to seek a sense of balance by placing components centrally when they lack information to the contrary. It may also be that the greater balance in this figure simply makes it look easier and more esthetically appealing to subjects and they respond with a higher level of motivation that translates into better performance.

2. Scale Reliability

Reliability of the test scores obtained on the Diagnostic and Tracking Tests was assessed by examining (a)

internal consistency on both subtests, and (b) inter-scorer reliability on the Geometric Figure subtests.

(a) Internal Consistency

Internal consistency values were good on the Diagnostic Test, ranging from .84 to .86 on the Paragraph subtest and from .75 to .85 on the Rectangle Figure subtest. Internal consistencies were satisfactory on the Tracking Test, ranging from .70 to .72 on the Paragraph subtest and from .69 to .76 on the Triangle subtest.

As noted earlier in the description of the analyses, internal consistency estimates conducted on tests of learning and memory must be interpreted with caution. Two assumptions of psychometric measures such as Cronbach's alpha and KR-20 are not met with such tests (Aiken, 1994; Anastasi, 1988). First, the "items" in these tests are not sampled from a content domain as would be the case with, for example, a test of metamemory or depression. Second, due to the characteristics of learning and memory tests, each "item" does not contribute equally to the measurement of the construct at hand (e.g., memory). That is, because of variables such as multiple trials and context-dependent or imbedded items (e.g., "three ginger cookies" or dots imbedded within a square), not all items have an equal likelihood of being recalled. In test development, attempts to increase internal consistency typically involve deleting those items that correlate poorly with the total score. In

the present instance, this approach is clearly inappropriate. That is, if an item performs poorly because of its placement or order in a test (e.g., due to serial position effects), replacing it will not improve reliability estimates.

Interpretation of the obtained internal consistencies must also take into account that these results were obtained with a homogeneous sample (i.e., only cognitively intact individuals) using relatively easy task materials. Both sample homogeneity and easy tasks' tend to produce a restricted range of scores, which in turn, tends to result in lower estimates of internal consistency. These variables may have impacted upon the reported internal consistency estimates for the G-LAMB subtests, and are the probable explanation for the particularly low internal consistency estimates on the Tracking Test Paragraph and Geometric Figure.

Another variable that affects internal consistency is the number of items comprising a test. Generally, the greater the number of items in a test, the higher the internal consistency (Aiken, 1994). Because the Paragraph subtest of the Diagnostic Test contains more items than any of the other G-LAMB subtests, it is not surprising that the highest internal consistency estimates were obtained with this subtest.

In summary, the internal consistency estimates obtained

on the Diagnostic and Tracking Tests were in the satisfactory to good range and would be expected to increase if administered to a more heterogenous sample.

(b) Inter-scorer Reliability

For any test or subtest in which there is a subjective component to the scoring, an estimate of the inter-scorer reliability (i.e., the degree of agreement between the scorers) should be provided. The scoring of the Paragraph subtests is relatively straightforward and involves no subjective component. Thus, inter-scorer reliability is not assessed for these subtests. However, there is the potential for substantial subjective judgment in scoring the Geometric Figure subtests. Here, the scorer must make a decision regarding the correctness of each component and its placement within the figure.

Inter-scorer reliability estimates were calculated for each of the three learning trials. These estimates ranged from .98 to .99 on the Diagnostic Test Geometric Figure and from .95 to .98 on the Tracking Test Geometric Figure, providing strong evidence for the inter-scorer reliability of these subtests.

Two factors may have contributed to the achievement of high inter-scorer reliabilities on these subtests. First, although the figures were scored independently by the two scorers, practice sessions and informal discussions of scoring rules did take place prior to scoring the figures,

as recommended by Smith (1986). Such a procedure tends to improve inter-scorer reliability. Second, a restricted range consisting only of high scores (due to an easy test) should result in high inter-scorer reliability because the accuracy of subjects' construction, and placing, of the components leaves little room for disagreement between the scorers⁶.

Given that a restricted range of scores due to an easy test should result in increased inter-scorer reliability, it is somewhat surprising that estimates were lower on the easier Tracking Test figure than on the Diagnostic Test figure. It could be argued that since only one scorer (the author) was involved in the scoring of both the Diagnostic and Tracking Test figures, the difference in inter-scorer reliabilities between the figures may be attributable to characteristics associated with the scorers rather than the figures. However, it would also seem that there is very little difference between average inter-scorer reliabilities of approximately .99 for the Diagnostic Test figure and .97 for the Tracking Test figure.

3. Validity

Validation of inferences made from test scores is an ongoing process that is not restricted to any single study. In the present dissertation, evidence for validity is documented by examining (a) the effect of age on the Paragraph and Geometric Figure scores, (b) the pattern of

results found with testing procedures such as immediate and delayed recall, a retrieval continuum, and multiple trials, and (c) the results of factor analyses examining the pattern of relationships among several tests administered in Studies 1 and 2.

(a) Age Effects

Performance on learning and memory tests consistently has been shown to decrease with increasing age (Davis & Bernstein, 1992; Howieson, et al., 1993; Petersen et al., 1992; Schaie, 1994; Spinnler et al., 1988). Such age effects are evident even within elderly samples, and thus would be expected in the present study as well. An examination of either the mean raw scores (see Tables 9, 10, 28 & 29) or the mean percent recalled (see Tables 39 & 40) by each of three age groups (55-64 yrs, 65-74 yrs, and 75-85 yrs) on the Diagnostic and Tracking Tests clearly shows that performance decreases as age increases across the three groups.

Age contributed significantly to the scores obtained on acquisition and delayed recall for both the Paragraph and Geometric Figure subtests. The proportion of variance explained by age ranged from .05 to .20, and was greater for acquisition than delayed recall (see Tables 5 and 24). The impact of age on delayed recall scores generally seems to reflect differences seen on acquisition since, with the exception of the Diagnostic Paragraph, there were no age

differences observed on the proportion of learned material that was remembered.

On the Paragraph subtests, acquisition performance was generally found to increase across trials and decrease with increasing age. On the Diagnostic Test, these age effects varied across trials with the fastest rate of acquisition occurring for the oldest age group. Although the performance of the oldest group was significantly lower than that of the two younger groups on both acquisition trials, this group's greater increase in performance across trials was probably related to its low level of initial scores which, in turn, allowed performance to increase at a faster rate. A similar pattern was observed on the Tracking Test, although the interaction was not statistically significant.

Similar relationships were observed with the Geometric Figure subtests. Acquisition performance decreased with age and interacted with trials so that all three age groups performed significantly differently from one another on Trial 1, but the performance of the oldest group was significantly lower than that of the other two groups on Trials 2 and 3. This finding seemed to reflect the greater opportunity of the 75-85 year old adults to improve their performance across trials given their lower initial scores.

Significant age differences were found for performance on the copy trial of the Diagnostic Test Geometric Figure. The constructional abilities of the 75-85 year old group

were found to be significantly poorer than those of the two younger age groups, despite differences between the groups of less than one point (25.0 vs 25.5 and 25.7). This trivial, but statistically significant, difference was undoubtedly due to the small within-group variance. No significant age differences were found on the copy trial of the Tracking Test Geometric Figure. Findings from the literature on complex figures are inconsistent, with some researchers finding decreased constructional ability with increasing age (Boone, Lesser, Hill-Gutierrez, Berman, & D'Elia, 1993; Rosselli & Ardila, 1991; Ska & Nespoulous, 1988; Tombaugh & Schmidt, 1992), and other researchers reporting no significant age differences on copy performance (Chiulli, Yeo, Haaland, & Garry, 1989; Delbecq-Derousesne & Beauvois, 1989; Spreen & Strauss, 1991). In the present case, the small age differences on the Diagnostic Test figure and the lack of age differences on the Tracking Test figure may be attributable, at least in part, to the easier (i.e., less complex) test materials used in this task.

It should also be noted that the proportion of variance explained by age also tended to be greater for the Geometric Figures than for the Paragraphs (see Tables 5 and 24). This finding is consistent with the literature which indicates that tests of visuospatial memory are affected by age more than tests of verbal memory (Benton, Eslinger, & Damasio, 1981; Janowsky & Thomas-Thrapp, 1993; Tombaugh & Schmidt,

1992).

In summary, age was found to have a significant impact upon the acquisition of verbal and visuospatial stimuli. This finding, which was particularly evident with the Geometric Figure subtests, is consistent with the literature. Age also had a significant impact upon copy scores for the Diagnostic figure, but not the Tracking figure; this finding parallels the inconsistency in the literature regarding age effects on constructional abilities. Although age effects were found on delayed recall scores, this finding seems to reflect the age differences seen on acquisition since, with one exception, there were no age differences observed on the proportion of learned material that was remembered.

(b) Test Procedures

There are a number of test procedures used in the G-LAMB that most researchers agree can help distinguish between normal and impaired memory performance. These procedures include (a) a retrieval continuum, (b) immediate and delayed recall, and (c) multiple trials.

A retrieval continuum refers to the use of free recall, cued recall, and recognition procedures to examine an individual's ability to retrieve information. These procedures represent a hierarchy of retrieval support with free recall providing the least support and requiring the most cognitive resources (Crain, 1986). If material has

been properly stored and consolidated, but is not easily retrieved through free recall, it may be retrieved through the use of appropriate cues (i.e., cued recall). Finally, further retrieval of information may be accomplished through the use of a recognition format, which requires the least cognitive resources. Any other information not retrieved is assumed to have been forgotten. There is considerable evidence in the literature, with both cognitively intact and impaired individuals, that shows recognition performance to be higher than cued recall performance which, in turn, is higher than free recall performance (Butters et al., 1985; Butters & Cermak, 1980; Craik, 1991; Kazniak et al., 1986; Moscovitch & Winocur, 1992; Smith, 1977).

Given that a retrieval continuum has been shown to be useful, there are still a number of considerations to be made. First, it must be decided whether a retrieval continuum is best used on both verbal and visuospatial tests. Second, it must be decided whether the entire retrieval continuum should be used, or only part of the continuum. And third, it must be decided exactly where such a continuum should be used within a task; that is, should a retrieval continuum be used on both acquisition and delay trials or only on delay?

It was decided that a retrieval continuum would only be used with the Paragraph subtests and that the Geometric Figure subtests would be tested only through free recall.

Use of a retrieval continuum with complex figure tasks is still in the experimental stages. Thus, given the difficulty in developing appropriate cues and recognition items with this type of task and the added time for administration, the need to keep the G-LAMB tests short and practical outweighed the potential benefits of such a procedure.

Next, all three components of a retrieval continuum are used only with the Diagnostic Test Paragraph. Because the Diagnostic Test is intended for diagnostic use, it was decided that important clinical distinctions could best be made using all three levels of retrieval. The Tracking Test Paragraph, however, uses only free recall and recognition procedures. Because the Tracking Test was developed to examine the memory strengths and weaknesses of impaired individuals who were already diagnosed, it was decided that only free recall and recognition procedures were essential to determine the amount of information that a person could recall (a) without help and, and (b) with maximal support. The exclusion of cued recall also reflected an attempt to decrease the amount of administration time.

Finally, a retrieval continuum was only used on the delay trial of both the Diagnostic and Tracking Paragraph subtests. Although cued recall and recognition trials can be used effectively during acquisition (e.g., LAMB Paragraph), it was decided that with an elderly and/or

impaired population any information gained by such a procedure would be counteracted by either fatigue due to the added time component or frustration if acquisition performance is poor.

As expected from the literature, performance on the Diagnostic and Tracking Paragraph subtests increased as retrieval support increased. That is, performance was (a) higher on recognition than on either free recall or cued recall, and (b) higher on cued recall than free recall (see Tables 6, 9, 25, and 28).

The use of both immediate and delayed recall in tests of memory to distinguish learning from memory has been strongly recommended in the literature (Ardila & Rosselli, 1989; Bornstein & Chelune, 1989; Erickson & Scott, 1977; Larrabee & Crook, 1995; Schmidt & Tombaugh, 1995). Immediate recall alone allows one to determine the amount of information learned during the acquisition phase whereas delayed recall permits one to determine the amount of information either retained or forgotten. Delayed recall has been shown to be more sensitive to the effects of various types of memory deficit than is immediate recall (Ardila & Rosselli, 1989; Lezak, 1995; Welsh, Butters, Hughes, Mohs, & Heyman, 1991). In addition, immediate and delayed recall can be used to calculate the proportion of encoded information that is retained after delay (i.e., proportion remembered or "savings" score). Very little

difference is found between immediate and delayed recall among cognitively intact older adults on verbal or visual tasks (Tombaugh, Faulkner, & Schmidt, 1992). As a result, any decline in performance (i.e., forgetting) that occurs between the last acquisition trial and delay trial tends to be clinically significant (Bornstein et al., 1988; Butters et al., 1988; Moss et al., 1986).

For the most part, these findings seem to be supported by the data available on the G-LAMB. That is, very little information is forgotten (less than 10%) from the last acquisition trial to the delayed recall trial using the the Diagnostic Test Rectangle figure or the Tracking Test Paragraph or Triangle figure. This was not the case with the Paragraph subtest from the Diagnostic Test. Here, the 75-85 year old group forgot significantly more information (approximately 14% versus 8%) than did the two younger groups. This probably occurred because the Diagnostic Test Paragraph is relatively difficult compared to the other subtests and so differences between the groups became more apparent.

Finally, the use of multiple trials has been recommended to permit the examination of learning curves and to reduce the impact of factors such as anxiety or distractibility on performance (Erikson & Scott, 1977; Lezak, 1995; Schmidt & Tombaugh, 1995). Moreover, an individual's performance across acquisition trials can also

be used to distinguish among different disorders (Lezak, 1995; Luria, 1981). In both of the G-LAMB tests, the Paragraph and the Geometric Figure subtests consist of two and three acquisition trials, respectively. Normative data is also provided on a trial-by-trial basis so that learning curves can be examined.

(c) Factor Analytic Findings

Two sets of principal axis factor analyses were conducted in each study. The first factor analysis examined separately the factor structure of the Diagnostic and Tracking Tests. Two factors were found in each case: (1) a verbal factor consisting of the acquisition and delay trials of the Paragraph subtest, and (2) a visual factor consisting of the acquisition and delay trials of the Geometric Figure subtest. This finding is consistent with the literature. When factor analyses were conducted with the LAMB, the Paragraph and Complex Figure subtests loaded on verbal and visual factors, respectively (Schmidt & Tombaugh, 1995). Similar results were found with the WMS-R when both acquisition and delayed recall scores were used in the analysis (Bornstein & Chelune, 1982). For both the Diagnostic and Tracking Tests, the two factors were moderately correlated, which might reflect a common learning and memory component, a component of verbal encoding common to these tasks, or both.

The second factor analysis allowed for an examination

of the convergent and discriminant validity of the Diagnostic and Tracking Tests by factor analyzing them with several other cognitive tasks. In Study 1, the factor analysis included the Paragraph and Geometric Figure subtests as well as a Word List, Trail-Making (Trails A and B), Abstract Recognition Memory, Verbal Fluency (F, A, S, and Animals), and WAIS-R Vocabulary. Based on the results of factor analyses conducted with the LAMB (Schmidt & Tombaugh, 1995), it was expected that the verbal tasks (Paragraph and Word List) would load on one factor, visual tasks (Geometric Figure and Abstract Recognition Memory) would load on a second factor, and the WAIS-R Vocabulary would load on a third factor. It was not clear whether Trail-Making and Verbal Fluency would load on separate factors or on some of the other factors. The result, a three-factor solution, was close to the expected results. The visual tasks did load together on one factor, but the inclusion of Trails A and B with the Geometric Figure and Abstract Recognition Memory test changed the interpretation of the factor from one of memory to one of visual processing. The second factor consisted of the WAIS-R Vocabulary subtest and the Verbal Fluency measures and so was labelled expressive language. Finally, as predicted, the Paragraph and Word List loaded on the same factor, verbal memory. The verbal memory and visual processing factors were the most highly correlated ($r = .60$), which may

reflect either a verbal and/or memory component to these tasks. The next highest correlation was between the expressive language and verbal memory factors and probably reflects the strong verbal component to both of these factors.

In Study 2, the factor analysis included the Paragraph and Geometric Figure subtests as well as Word Recognition Memory, Picture Recognition Memory, and WAIS-R Vocabulary. Again, based on the factor analysis conducted with the LAMB (Schmidt & Tombaugh, 1995), it was expected that three factors would be found with the Paragraph and Word Recognition Memory test comprising one factor, the Geometric Figure and Picture Recognition Memory test comprising a second factor, and WAIS-R Vocabulary loading on the third factor. Instead, a four-factor solution was obtained with each of the four learning and memory tests loading on separate factors. WAIS-R Vocabulary loaded on the same factor as the Paragraph subtest, but was not interpreted because the factor loading was less than .30. The resulting factors seem to be based on both the stimuli (i.e., verbal versus visual) and task demands (i.e., free recall versus recognition). However, since the Geometric Figure subtest and Abstract Recognition Memory loaded on the same factor in Study 1, perhaps more weight needs to be given to the specific type of stimuli used. For example, both the Geometric Figure subtest and the Abstract Recognition Memory

test used abstract visual stimuli, whereas in Study 2, the Geometric Figure involved abstract visual stimuli but the Picture Recognition Memory used more concrete visual stimuli. Likewise, the Paragraph in Study 2 may have used more contextual, complex information than the Word List. Correlations among the factors ranged from low to moderate levels, with the highest correlation occurring between the two verbally-loaded factors (i.e., contextual verbal memory and concrete verbal recognition).

4. Normative Data

The Diagnostic and Tracking Tests were normed on relatively large samples (184 and 180 subjects, respectively) of community-dwelling men and women aged 55-85 years. Individuals were screened for cognitive impairment using the MMSE, and for depression using the GDS. In addition, individuals with a history of multiple cognitive risk factors (e.g., heart attack, thyroid problems) and/or unusually low performance that might be related to a history of head injury or organic disease, were also excluded from the study.

The prominence of age in the regression analyses dictated that the norms were stratified by age. Determination of the specific age groupings was based on the following criteria: (a) the same age groups should be used for both subtests, (b) age effects within the selected age groups should be minimized, (c) the selected age groups

should not show differences on the demographic variables, (d) relatively large sample sizes for each age group should be employed, and (e) the selected age groups should show differences in performance on the subtests. As a result, norms are reported for three age groups: 55-64 years, 65-74 years, and 75-85 years.

The development of norms for the G-LAMB is important because it provides an age-related baseline against which to compare an individual's performance. As is the convention, the norms for total acquisition, delayed recall, and (in the case of the Geometric Figure) the copy trial are reported in terms of standard scores. A mean of 100 and standard deviation of 15 was selected to facilitate comparison with several other tests (e.g., WAIS-R, WMS-R) that use the same standard scores. The use of standard scores assumes that the test scores are normally distributed in the population. Because scores in a learning and memory test become quite negatively skewed in later trials, reflecting increased performance over trials, cumulative percentiles (rather than standard scores or percentiles) are used in reporting the norms for the trial-by-trial data. Cumulative percentiles indicate the percentage of scores that fall at or below a given raw score. In using the normative tables, keep in mind that Trial 1 performance gives an indication of how quickly an individual learns initially, although a low score might also reflect distractability or fatigue. The total

acquisition score indicates overall learning, and the delayed recall score shows retention or memory for the learned material. The copy trial of the Geometric Figure provides an indication of constructional and visuo-perceptual processes. Learning and memory performance on the Geometric Figure subtest should not be interpreted unless these processes appear to be intact.

Summary

The Diagnostic and Tracking Tests are short, practical tests of memory designed specifically for use with older adults. Several features have been included in the development of these tests that are important for examining memory within this age group. These features include the use of easier task materials, a self-read procedure with the Paragraph subtests, recommended test procedures (such as a retrieval continuum, immediate and delayed recall, and multiple trials) and the development of normative data stratified by age for adults aged 55-85 years.

Limitations and Potential Difficulties

There are a number of limitations or potential difficulties in using the G-LAMB tests that must be considered. These limitations are associated with use of (a) a self-read procedure, (b) a Paragraph subtest to assess verbal memory, and (c) relatively easier test materials.

(a) Self-Read Procedure

The interest in using a self-read procedure stemmed

from the unsolicited comments of older adults during the norming of the LAMB test that they would remember more if only they could read the task materials themselves. This general belief appears to be supported by the substantial increase in performance by older adults on the Diagnostic Test Paragraph (relative to older adults' performance reported on the LAMB Paragraph in Tombaugh and Schmidt, 1992). One reason that a self-read procedure may help older adults is that the additional (30 second) exposure to the test materials provides individuals with the opportunity to encode and rehearse whatever portions of the material they deem necessary.

Unfortunately, the use of a self-read procedure may not be beneficial for all older adults. Individuals who have reading difficulties (e.g., they are illiterate, dyslexic, slow readers, or have difficulty with reading comprehension), display severe attentional problems, are in an acute confusional state, or are demented are likely to be disadvantaged by a self-read procedure. Clearly, a person who is illiterate will be unable to use the self-read procedure. For other individuals with reading difficulties, the self-read procedure may actually interfere with encoding processes as the person spends more time trying to understand what they are reading than trying to learn or remember the information. Because research suggests that it is the opportunity to review (rather than simply read) the

material that is the key to the success of a self-read procedure (Taub, 1976; Taub & Kline, 1978; Verhaeghen, et al., 1993), it would seem that individuals who cannot use the 30 second self-read time efficiently (e.g., due to difficulties with reading, attention, or following instructions) may be differentially penalized on their learning and memory scores for verbal material. Therefore, it is recommended that the self-read procedure of the Diagnostic and Tracking Test Paragraphs not be used with older adults who would appear to be disadvantaged by such a procedure.

(b) Use of a Paragraph to Assess Verbal Memory

Both the Diagnostic and Tracking Tests use a paragraph rather than a word list to examine learning and memory for verbal materials. There are two main reasons for this decision. First, the underlying context of a story tends to provide more support in learning the material (Delis, 1989). Second, a narrative (such as the paragraphs used in the G-LAMB) is considered to be a good match in complexity to a (complex) geometric figure (Freides & Avery, 1991).

There may be several drawbacks, however, to using paragraphs to assess verbal learning and memory. First, the advantages of using a paragraph task for older adults (e.g., story context) may be counteracted by the sheer increase in the number of words that need to be learned in a paragraph. For example, the G-LAMB Paragraphs are 21- and 31-items

long, whereas most word lists tend to be 10 to 16 items in length.

Second, whereas some individuals may find that the context of a story makes it easier to recall or guess other parts of the story, it is also possible that this context may interfere with learning and recall. For example, the use of a paragraph may result in less flexibility in encoding so that the individual is less able to make their own associations among the information to be recalled. In addition, the fact that the information is connected in a story may mean that if a person misses or forgets part of the story, it becomes more difficult to recall other parts of the story.

Third, primacy and recency effects, which have been shown to be important in differentiating certain diseases, are not examined as easily with a paragraph task. And finally, paragraphs are harder to score than other verbal tasks (e.g., a word list) because decisions need to be made with regard to (a) the words that are acceptable alternate responses, and (b) how to score information that is recalled out of sequence.

(c) Use of Relatively Easy Test Materials

One of the problems with currently available memory tests, such as the LAMB, is that many of the test materials are too difficult for the average older adult. In an attempt to avoid the problems associated with overly

difficult tests, the task materials in the G-LAMB tests were made easier by decreasing the number of components to be remembered and/or using a self-read procedure. This resulted in test performance by cognitively intact older adults that resembles that of younger adults on the LAMB. By increasing the performance of intact older adults, a greater proportion of the range of scores are available to describe the performance of impaired older adults; thus, floor effects seem unlikely.

Theoretically, ceiling effects should be no more likely on the Diagnostic and Tracking Tests than would be the case with younger age groups on tests such as the LAMB. However, it is possible that the use of easier task materials and/or a self-read procedure could have made the tasks so easy that mildly impaired individuals might perform at levels that are not distinguishable from intact performance. Although this seems unlikely, the question of whether ceiling effects are present on the G-LAMB tests must remain unanswered until further research is conducted with cognitively impaired samples.

Whether the three potential limitations or difficulties described above will turn out to be problems for the G-LAMB will depend, in part, upon future research. The final section of the dissertation briefly describes some areas of future research that are being considered.

Future Research

Four lines of research are planned with the G-LAMB Tests. These include: (a) a validation of the tests with clinical samples, (b) an examination of rates of forgetting, (c) the development of alternate figures for the Geometric Figure subtests, and (d) an analysis of constructional and recall strategy by intact older adults on the Geometric Figure subtests. Each of these lines of research will be discussed briefly.

The most pressing research needed is to validate the use of the Diagnostic and Tracking Tests with clinical samples. The Diagnostic Test needs to be administered to intact, as well as mild to moderately memory-impaired, individuals to determine the degree that performance on the tests can differentiate between these groups (e.g., sensitivity and specificity). In addition, the expected relationships between different types of memory impairment and different disease processes needs to be demonstrated. For example, normal rates of forgetting are expected with diencephalic amnesia, whereas rapid rates of forgetting typically are found with bitemporal amnesia and Alzheimer's disease (Huppert & Piercy, 1978; Squire, 1982). Another example is that use of a recognition format tends to improve memory performance more for individuals suffering from some disorders (e.g., Huntington's disease; Butters et al., 1985) than others (e.g., Korsakoff's disease; Butters & Cermak,

1980).

Another area of research that can be addressed using the Diagnostic and Tracking Tests is an examination of rates of forgetting over long delay intervals among cognitively intact and impaired older adults. In fact, data collection has already begun using the Diagnostic Test to examine rates of forgetting over periods of 20 minutes, 1 day, 1 week, 3 weeks, and 6 weeks among older adults. Such data will not only provide valuable information about the amount and type of information that is retained over these delay intervals, but the relative memorability of verbal and visuospatial materials as well.

A third line of research involves the development of alternate figures for the Geometric Figure subtests. Practice effects, or the finding that performance improves over repeated administration of the same figure, has been shown with complex figures (Spreeen & Strauss, 1991). Thus, the development of alternate figures for the both Diagnostic and Tracking figures would be recommended for retesting purposes. Moreover, if the Tracking Test is shown to be useful with a cognitively impaired sample, more than one alternate figure may be required to track memory strengths and weaknesses over time.

Finally, another line of research, for which the data collection has already begun, involves a qualitative analysis of the types of strategies used by intact older

adults on both copy and recall trials of the Geometric Figure subtests. Because the same score can be assigned for very different types of copy or recall performance, considerable interest has been expressed in assessing more qualitative aspects (e.g., analysis of error types and strategies) of figure performance (Janowsky & Thomas-Thrapp, 1993; Loring, Lee, & Meador, 1988; Shorr, Delis, & Massman, 1992; Waber & Holmes, 1985). More specifically, recent research by Ska and Nespoulous (1988) with adults aged 20-84 years suggests that the developmental stages described in Osterrieth's (1944) classic paper may be overstated as they found just as many adults used a piecemeal approach as used a global approach to copying a complex figure. In addition, the examination of copy and recall strategies of cognitively intact individuals has implications for interpreting the strategies used by dementing and head-injured individuals. For example, the findings of Ska and Nespoulous (1988) may have implications for the finding that individuals with right-sided lesions tend to use a piecemeal approach to copy complex designs (Gainotti & Tiacci, 1970).

Endnotes

¹ The variance accounted for by a factor is the amount of variance in the original variables that has been condensed into the factor. It should be noted, however, that when oblique rotation is used the variance accounted for is only a rough estimate since it is calculated before rotation and hence before variance among the factors is shared (Tabachnick & Fidell, 1989).

² When the scores are normally distributed, standard scores usually are calculated by converting the raw scores into z-scores and then into standard scores of a given mean and standard deviation (Aiken, 1994).

³ Test validity refers to the validation of inferences made from test scores rather than to the validity of the test itself.

⁴ The last acquisition trial on the LAMB complex figure is Trial 4. However, performance is reported for Trial 3 on this subtest to allow for an adequate comparison of performance on the LAMB and G-LAMB figures. The percent correctly recalled on total acquisition (Tot. Acq.) for the LAMB complex figure has been adjusted to reflect performance over three acquisition trials rather than the original four trials. Caution must be taken in interpreting the percent correctly recalled on the delay trial for the LAMB complex figure, however, as this reflects performance after the original four acquisition trials.

⁵ The issue of a restricted range of scores, and hence decreased internal consistency estimates, is relevant to tasks that are either overly easy (as in the present case) or overly difficult. Both cases can lead to a decreased variability in the observed scores.

⁶ This would not *necessarily* be the case when a restricted range of scores is the result of a very difficult test. Poor performance might result in high inter-scorer reliability because there would be very little on which the scorers could disagree. Alternatively, poor performance might result in decreased inter-scorer reliability because the assignment of points to a very poor drawing could occur in many different ways.

References

- Aiken, L. R. (1994). Psychological testing and assessment. (8th ed.). Boston, MA: Allyn & Bacon.
- Anastasi, A. (1988). Psychological testing. (6th ed.). New York: Macmillan Pub. Co.
- Anders, T. R., Fozard, J. L., & Lillyquist, T. P. (1972). Effect of age upon retrieval from short-term memory. Developmental Psychology, 6, 214-217.
- Ardila, A., & Rosselli, M. (1989). Neuropsychological characteristics of normal aging. Developmental Neuropsychology, 5, 307-320.
- Atkinson, J. W., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence and J. T. Spence (Eds.), The psychology of learning and motivation (Vol. 2). New York: Academic Press.
- Baddeley, A. D. (1986). Working memory. New York: Oxford University Press.
- Barak, B., & Schiffman, L. G. (1981). "Cognitive age: A nonchronological age variable". In K. B. Monroe (Ed.), Advances in consumer research (Vol. 8) (pp. 602-606). Ann Arbor: Association for Consumer Research.
- 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). Revised Visual Retention Test

(4th ed.). New York: Psychological Corporation.

Benton, A. L., Eslinger, P. J., & Damasio, A. R. (1981). Normative observations on neuropsychological test performance in old age. Journal of Clinical Neuropsychology, 3, 33-42.

Bigler, E. D. (1988). Frontal lobe damage and neuropsychological assessment. Archives of Neurology, 3, 279-297.

Bigler, E. D., Rosa, L., Schultz, F., Hall, S., & Harris, J. (1989). Rey-auditory verbal learning and Rey-Osterrieth complex figure design performance in Alzheimer's disease and closed head injury. Journal of Clinical Psychology, 45, 277-280.

Binder, L. M. (1982). Constructional strategies on complex figure drawings after unilateral brain damage. Journal of Clinical Neuropsychology, 4, 51-58.

Birren, J. E. (1974). Translation in gerontology - from lab to life: psychophysiology and speed of response. American Psychologist, 29, 808-815.

Boone, K. B., Lesser, I. M., Hill-Gutierrez, E., Berman, N. G., & D'Elia, L. F. (1993). Rey-Osterrieth Complex Figure performance in healthy, older adults: Relationship to age, education, sex, and IQ. The Clinical Neuropsychologist, 7, 22-28.

Bornstein, R. A., & Chelune, G. J. (1989). Factor structure of the Wechsler Memory Scale - Revised in relation

to age and educational level. Archives of Clinical Neuropsychology, 4, 15-24.

Bornstein, R. A., Chelune, G. J., & Prifitera, A. (1989). IQ-memory discrepancies in normal and clinical samples. Psychological Assessment, 1, 203-206.

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

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: A Bayesian analysis. Experimental Aging Research, 8, 189-193.

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

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

Butters, N., & Cermak, L. S. (1980). Alcoholic Korsakoff's syndrome. New York: Academic Press.

Butters, N., Salmon, D. P., Cullum, C. M., Cairns, P., Troster, A. I., Jacobs, D., Moss, M., & Cermak, L. S. (1988). Differentiation of amnesic and demented patients with the Weschler Memory Scale - Revised. The Clinical Neuropsychologist, 2, 133-148.

Butters, N., Wolfe, J., & Martone, M., et al., (1985). Memory disorders associated with Huntington's disease: Verbal recognition and procedural memory. Neuropsychologia, 23, 729-743.

Camp, C., & McKittrick, L. (1989). The dialectics of forgetting and remembering across the adult lifespan. In D. A. Kramer and M. Bopp (Eds.), Transformation in clinical and developmental psychology. New York: Springer-Verlag.

Cerella, J., Poon, L., & Fozard, J. (1982). Age and iconic read-out. Journal of Gerontology, 37, 197-202.

Chiulli, S. J., Haaland, K. Y., LaRue, A., and Garry, P. J. (1995). Impact of age on drawing the Rey-Osterrieth Figure. The Clinical Neuropsychologist, 9, 219-224.

Claussen, C.-F, & Patel, N. P. (1990). Sensory changes in later life. In M. Bergener and S. I. Finkel (Eds.), Clinical and scientific psychogeriatrics (Vol 1: The holistic approaches). New York: Springer.

Cliff, N. (1987). Analyzing multivariate data. San Diego, CA: Harcourt Brace Jovanovich Pub.

Cohen, J. (1992). A power primer. Psychological Bulletin, 112, 155-159.

Corgiat, M. D., Templer, D. I., & Newell, T. C. (1989). The effects of presentation modality, task demand, and content structure on age-related memory differences for prose. International Journal of Aging and Human Development, 29, 53-65.

Corkin, S. (1982). Some relationships between global amnesia and the memory impairments in Alzheimer's disease. In S. Corkin, K. K. Davis, J. H. Growden, et al. (Eds.), Alzheimer's disease: A report of progress in research. New York: Raven.

Corkin, S., Growdon, J. H., Sullivan, E. V., Nissen, M. J., & Huff, F. J. (1986). Assessing treatment effects: A neurological battery. In L. W. Poon (Ed.), Handbook for Clinical Memory Assessment of Older Adults. Washington, DC: American Psychological Association.

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

Craik, F. I. M. (1986). A functional account of age differences in memory. In F. Klix and H. Hagendorf (Eds.), Human memory and cognitive capacities. Amsterdam: Elsevier.

Craik, F. I. M. (1991). Memory function in normal aging. In T. Yanagihara and R. C. Petersen (Eds.), Memory disorders: Research and clinical practice. New York: Marcel Dekker.

Craik, F. I. M., & McDowd, J. M. (1987). Age differences in recall and recognition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 13, 474-479.

Craik, F. I. M., & Rabinowitz, J. C. (1984). Age

differences in the acquisition and use of verbal information. In J. Long and A. Baddeley (Eds.), Attention and performance. (Vol. X). Hillsdale, NJ: Erlbaum.

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. Fozard, L. Cermak, D. Arenberg, & L. W. Thompson (Eds.), New directions in memory and aging: Proceedings of the George Talland memorial conference. (pp. 95-112). Hillsdale, NJ: Lawrence Erlbaum Assoc.

Danziger, W. L., & Storandt, M. (1982). Psychometric performance of healthy and demented older adults: A one-year follow-up. Paper presented at the 35th Annual Meeting of the Gerontological Society of America, Boston, MA.

Darlington, R. B. (1990). Regression and linear models. New York: McGraw-Hill.

Davis, H. P., & Bernstein, P. A. (1992). Age-related changes in explicit and implicit memory. In L. R. Squire and N. Butters (Eds.), Neuropsychology of memory. New York: The Guilford Press.

Delbecq-Derouesne, J., & Beauvois, M.-F. (1989). Memory processes and aging: A defect of automatic rather than controlled processes? Archives of Gerontology and Geriatrics, (Suppl. 1), 121-150.

Delbecq-Derouesne, J., Beauvois, M.-F., & Shallice, T. (1990). Preserved recall versus impaired recognition: A case study. Brain, 113, 1045-1074.

Delis, D. C. (1989). Neuropsychological assessment of learning and memory. In F. Boller and J. Grafman (Eds.), Handbook of neuropsychology, Vol. 3, (pp. 3-33). Elsevier Science Pub.

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

Denney, N. W. (1974). Classification abilities in the elderly. Journal of Gerontology, 29, 309-314.

Earhard, M. (1977). Retrieval failure in the presence of retrieval cues: A comparison of three age groups. Canadian Journal of Psychology, 31, 139-150.

Erickson, R. C., Poon, L. W., Walsh-Sweeney, L. (1980). Clinical memory testing of the elderly. In L. W. Poon, J. L. Fozard, L. S. Cermak, D. Arenberg, & L. W. Thompson (Eds.), New directions in memory and aging (pp. 379-399). Hillsdale, NJ: Erlbaum.

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

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

Fozard, J. L. (1990). Vision and hearing in aging. In J. E. Birren and K. W. Schaie (Eds.), Handbook of the psychology of aging. (3rd ed.) New York: Academic Press.

Freides, D., & Avery, M. E. (1991). Narrative and visual spatial recall: Assessment incorporating learning and delayed retention. The Clinical Neuropsychologist, 5, 338-344.

Fuld, P. A. (1981). The Fuld Object-Memory Evaluation. Chicago: Stoelting Instrument Company.

Fuld, P. A. (1983). Psychometric differentiation of the dementias: An overview. In B. Reisberg (Ed.), Alzheimer's disease: The standard reference. New York: Free Press.

Gainotti, G., & Tiacci, C. (1970). Patterns of drawing disability in right and left hemispheric patients. Neuropsychologia, 8, 379-384.

Gilbert, J. G., Levee, R. F., & Catalano, F. L. (1981). Guild Memory Test Manual. Bloomfield, NJ: Unico National Mental Health Research Foundation.

Glosser, G., Goodglass, H., & Biber, C. (1989). Assessing visual memory disorders. Psychological Assessment, 1, 82-91.

Gottsdanker, R. (1982). Age and simple reaction time. Journal of Gerontology, 37, 342-348.

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesia subjects. Journal of Experimental Psychology: Learning, Memory, and Cognition, 13, 45-53.

Hakstian, A. R., & Abel, R. A. (1974). A further comparison of oblique factor transformation methods.

Psychometrika, 39, 429-444.

Hart, R. P., Kwentus, J. A., Harkins, S. W., & Taylor, J. R. (1988). Rate of forgetting in mild Alzheimer-type dementia. Brain and Cognition, 7, 31-38.

Hartley, J. T., Harker, J. O., & Walsh, D. A. (1980). Contemporary issues and new directions in adult development of learning and memory. In L. W. Poon (Ed.), Aging in the 1980s: Psychological issues. (pp. 239-252). Washington, DC: APA.

Hayes, D., & Jerger, J. (1984). Neurotology of aging: The auditory system. In M. L. Albert (Ed.), Clinical neurology of aging. New York: Oxford University Press.

Howieson, D. B., Holm, L. A., Kaye, J. A., et al., (1993). Neurologic function in the optimally healthy oldest old: Clinical neuropsychological evaluation. Neurology, 43, 1882-1886.

Hubley, A. M. (1994). Assessment of depression in older adults and the development of a new measure. (Unpublished manuscript, Carleton University).

Hulicka, I. M., & Grossman, J. L. (1967). Age group comparisons for the use of mediators in paired-associate learning. Journal of Gerontology, 22, 46-51.

Hultsch, D. F. (1971). Adult age differences in free classification and recall. Developmental Psychology, 4, 338-342.

Huppert, F. A., & Piercy, M. (1978). Dissociation

between learning and remembering in organic amnesia. Nature, 275, 317-318.

Incisa della Rochetta, A. (1986). Classification and recall of pictures after unilateral frontal or temporal lobectomy. Cortex, 22, 189-211.

James, W. (1890). Principles of psychology (2 vols.). New York: Henry Holt.

Janowsky, J. S., & Thomas-Thrapp, L. J. (1993). Complex figure recall in the elderly: A deficit in memory or constructional strategy? Journal of Clinical and Experimental Neuropsychology, 15, 159-169.

Jones-Gotman, M. (1986). Memory for designs: The hippocampal contribution. Neuropsychologia, 24, 193-203.

Kaiser, H. F. (1974). An index of factorial simplicity. Psychometrika, 39, 31-36.

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). Boston Naming Test. Philadelphia: Lea & Febiger.

Kazniak, A. W., Poon, L. W., & Riege, W. (1986). Assessing memory deficits: An information-processing approach. In L. W. Poon (Ed.), Handbook for clinical assessment of older adults. Washington, DC: APA.

Kesner, R. P., Hopkins, R. O., & Chiba, A. A. (1992). Learning and memory in humans, with an emphasis on the role of the hippocampus. In L. R. Squire and N. Butters (Eds.), Neuropsychology of memory. New York: Guilford Press.

Kolb, B. & Whishaw, I. Q. (1990). Fundamentals of human

neuropsychology. (3rd ed.) New York: W. H. Freeman and Co.

Kopelman, M. D. (1985). Rates of forgetting in Alzheimer-type dementia and Korsakoff's syndrome.

Neuropsychology, 23, 623-638.

Kopelman, M. D. (1986). Clinical tests of memory.

British Journal of Psychiatry, 148, 517-525.

Kramer, J. H., Delis, D. C., Blusewicz, M. J., Brandt, J., Ober, B. A., & Strauss, M. (1988). Verbal memory errors in Alzheimer's and Huntington's dementias. Developmental Neuropsychology, 4, 1-15.

Larrabee, G. J., & Crook, III, T. H. (1995). Assessment of learning and memory. In R. L. Mapou and J. Spector (Eds.), Clinical neuropsychological assessment: A cognitive approach. New York: Plenum Press.

Larrabee, G. J., Kane, R. L., Schuck, J. R., & Francis, D. J. (1985). Construct validity of various memory testing procedures. Journal of Clinical and Experimental Neuropsychology, 7, 230-250.

La Rue, A. (1992). Aging and neuropsychological assessment. New York: Plenum Press.

Lezak, M. D. (1995). Neuropsychological assessment (3rd ed.). New York: Oxford University Press.

Loring, D. W., Lee, G. P., & Meador, K. J. (1988). Revising the Rey-Osterrieth: Rating right hemisphere recall. Archives of Clinical Neuropsychology, 3, 239-247.

Luria, A. R. (1981). Higher cortical functions in man.

New York: Basic Books.

Mandler, G. (1967). Organization and memory. In D. W. Spence and J. T. Spence (Eds.), The psychology of learning and motivation (Vol. 1). New York: Academic Press.

Martin, A. (1987). Representation of semantic and spatial knowledge in Alzheimer's patients: Implications for models of preserved learning in amnesia. Journal of Clinical and Experimental Neuropsychology, 9, 191-224.

Massaro, D. W. (1970). Preperceptual auditory images. Journal of Experimental Psychology, 85, 411-417.

Mayes, A. R. (1988). Learning and memory disorders and their assessment. Neuropsychologia, 24, 25-39.

Milner, D. (1970). Memory and the medial temporal regions of the brain. In K. H. Pribram and D. E. Broadbent (Eds.), Biology of memory (pp. 29-50). New York: Academic Press.

Milner, D. (1975). Psychological aspects of focal epilepsy and its neurosurgical management. Advances in Neurology, 81, 299-321.

Montepare, J. M., & Lachman, M. E. (1981). "'You're only as old as you feel': Self perceptions of age, fears of aging, and life satisfaction from adolescence to old age". Psychology and Aging, 4, 73-78.

Moscovitch, M. (1984). The sufficient conditions for demonstrating preserved memory in amnesia: A task analysis. In L. R. Squire and N. Butters (Eds.), Neuropsychology of

memory. (pp. 104-114). New York: Guilford Press.

Moscovitch, M. (1992). A neuropsychological model of memory and consciousness. In L. R. Squire and N. Butters (Eds.), Neuropsychology of memory. New York: Guilford Press.

Moscovitch, M., & Umiltà, C. (1990). Modularity and neuropsychology: Modular and central processes in attention and memory. In M. F. Schwartz (Ed.), Modular deficits in Alzheimer's type dementia. (pp. 1-59). Cambridge, MA: MIT Press and Bradford Books.

Moscovitch, M., & Winocur, G. (1992). The neuropsychology of memory and aging. In F. I. M. Craik and T. A. Salthouse (Eds.), The handbook of aging and cognition. Hillsdale, NJ: Lawrence Erlbaum Associates, Pub.

Moss, M. B., & Albert, M. S. (1988). Alzheimer's disease and other dementing disorders. In M. S. Albert and Moss, M. B. (Eds.), Geriatric neuropsychology. New York: Guilford Press.

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.

Murdock, B. B. (1974). Human memory: Theory and data. New York: Wiley.

Murphy, K. R., & Davidshofer, C. O. (1994). Psychological testing: Principles and applications.

Englewood Cliffs, NJ: Prentice Hall.

Norusis, M. J. (1986). Advanced statistics SPSS/PC+ for the IBM PC/XT/AT. Chicago, IL: SPSS Inc.

Osborne, D. P., Brown, E. R., & Randt, C. T. (1982). Qualitative changes in memory function: Aging and dementia. In S. Corkin, K. L. Davis, J. H. Growdon, E. Usdin, & R. L. Wurtman (Eds.), Aging, (Vol. 19, Alzheimer's disease: A report of progress in research). New York: Raven Press.

Osterrieth, P. A. (1944). Le test du copie d'une figure complexe. Archives of Psychology, 30, 206-356.

Owsley, C., & Sloane, M. F. (1990). Vision and aging. In F. Boller and J. Grafman (Eds.), Handbook of neuropsychology. (Vol. 4). (pp. 229-249). Amsterdam: Elsevier.

Perlmutter, M., & Mitchell, D. B. (1982). The appearance and disappearance of age differences in adult memory. In F. I. M. Craik and S. Trehub (Eds.), Aging and cognitive processes (pp. 127-143). New York: Plenum Press.

Petersen, R. C., Smith, G., Kokmen, E., Ivnik, R. J., & Tangalos, E. G. (1992). Memory function in normal aging. Neurology, 42, 396-401.

Petri, H. L., & Mishkin, M. (1994). Behaviorism, cognitivism, and the neuropsychology of memory. American Scientist, 82, 30-37.

Read, D. E. (1987). Neuropsychological assessment of memory in the elderly. Canadian Journal of Psychology, 41,

158-172.

Rey, A. (1959). Manuel: Test du copie d'une figure complexe. Paris: Centre de Psychologie Appliquee.

Rey, A. (1964). L'examen clinique en psychologie. Paris: Presses Universitaires de France.

Ritchie, K. (1988). The screening of cognitive impairment in the elderly: A critical review of current methods. Journal of Clinical Epidemiology, 41, 635-643.

Rosselli, M., & Ardila, A. (1991). Effects of age, education, and gender on the Rey-Osterrieth Complex Figure. The Clinical Neuropsychologist, 5, 370-376.

Sahakian, B. J., Morris, R. G., Evenden, J. L., Heald, A., Levy, R., Philpot, M., & Robbins, J. W. (1988). A comparative study of visuospatial memory and learning in Alzheimer-type and Parkinson's disease. Brain, 111, 695-718.

Salmon, D. P., Shimamura, A. P., Butters, N., & Smith, S. (1988). Lexical and semantic priming in patients with Alzheimer's disease. Journal of Clinical and Experimental Neuropsychology, 10, 477-494.

Salthouse, T. A. (1976). Speed and age: Multiple rates of decline. Experimental Aging Research, 2, 349-359.

Salthouse, T. A. (1982). Adult cognition: An experimental psychology of human aging. New York: Springer-Verlag.

Schaie, K. W. (1994). The course of adult intellectual development. American Psychologist, 49, 304-313.

Schmidt, J. P., & Tombaugh, T. N. (1995). LAMB: Learning and Memory Battery Manual. Toronto: Multi-Health Services.

Schmitt, F. A., Murphy, M. D., & Sanders, R. E. (1981). Training older adults free recall rehearsal strategies. Journal of Gerontology, 36, 329-337.

Schonfield, D., & Robertson, B. A. (1966). Memory storage and aging. Canadian Journal of Psychology, 20, 228-236.

Shorr, J. S., Delis, D. C., & Massman, P. J. (1992). Memory for the Rey-Osterrieth figure: Perceptual clustering, encoding, and storage. Neuropsychology, 6, 43-50.

Ska, B., & Nespoulous, J.-L. (1988). Encoding strategies and recall performance of a complex figure by normal elderly subjects. Canadian Journal on Aging, 7, 408-416.

Smith, A. D. (1977). Adult age differences in cued recall. Developmental Psychology, 13, 326-331.

Smith, A. D. (1980). Age differences in encoding, storage, and retrieval. In I. Poon et al., (Eds.), New directions in memory and aging: Proceedings of the George A. Talland Memorial Conference. Hillsdale, NJ: Erlbaum.

Smith, D. E. (1986). Training programs for performance appraisal: A review. Academy of Management Review, 11, 22-40.

Sperling, G. (1960). The information available in brief

visual presentations. Psychological Monographs, 74, (11, whole no. 498).

Spinnler, H., Della Sala, S., Bandera, R., & Baddley, A. (1988). Dementia, aging, and the structure of human memory Cognitive Neuropsychology, 5, 193-212.

Spreeen, O., & Strauss, E. (1991). A compendium of neuropsychological tests: Administration, norms, and commentary. New York: Oxford University Press.

Squire, L. R. (1982). The neuropsychology of memory. Annual Review of Neuroscience, 5, 241-273.

Squire, L. R. (1986). Mechanisms of memory. Science, 232, 1612-1619.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. Psychological Review, 99, 195-231.

Squire, L. R., Shimamura, A. P., & Amaral, D. G. (1989). Memory and the hippocampus. In J. Byre & W. Berry (Eds.), Neural models of plasticity. (pp. 208-238). San Diego, CA: Academic Press.

Storandt, M., Botwinic, and humans. k, J., & Danziger, W. L. (1986). Longitudinal changes: Patients with mild SDAT and matched healthy controls. In I. W. Poon (Ed.), Handbook for clinical memory assessment of older adults. Washington, DC: American Psychological Association.

Syndulko, K., & Tourtellotte, W. W. (1989). What is neurologically normal? In T. L. Munsat (Ed.), Quantification

of neurologic deficit. Stoneham, MA: Butterworth Pub.

Tabachnick, B. G., & Fidell, L. S. (1989). Using multivariate statistics. New York: Harper & Row Pub.

Talland, G. A. (1968). Age and the immediate memory span. In G. A. Talland (Ed.), Human aging and behavior. New York: Academic Press.

Taub, H. A. (1976). Method of presentation of meaningful prose to young and old adults. Experimental Aging Research, 2, 469-474.

Taub, H. A., & Kline, G. E. (1978). Recall of prose as a function of age and input modality. Journal of Gerontology, 33, 725-730.

Taylor, L. B. (1969). Localisation of cerebral lesions by psychological testing. Clinical Neurosurgery, 16, 269-287.

Tombaugh, T. N. (1994). CMAR Memory Change Questionnaire. Unpublished manuscript.

Tombaugh, T. N., Faulkner, P., & Schmidt, J. 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.

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., & Schmidt, J. P. (1992). The Learning

and Memory Battery (LAMB): Development and standardization. Psychological Assessment, 4, 193-206.

Tuokko, H., & Crockett, D. (1989). Cued recall and memory disorders in dementia. Journal of Clinical and Experimental Neuropsychology, 11, 278-294.

Van Gorp, W. G., Satz, P., & Mitrushina, M. (1990). Neuropsychological processes associated with normal aging. Developmental Neuropsychology, 6, 279-290.

Verhaeghen, P., Marcoen, A., & Goossens, L. (1993). Facts and fiction about memory aging: A quantitative integration of research findings. Journal of Gerontology, 48, P157-P171.

Waber, D. P., & Holmes, J. M. (1985). Assessing children's copy productions of the Rey-Osterrieth Complex Figure. Journal of Clinical and Experimental Neuropsychology, 7, 264-280.

Waugh, N. C., & Norman, D. A. (1965). Primary memory. Psychological Review, 72, 89-104.

Wechsler, D. (1987). Manual for the Wechsler Memory Scale - Revised. San Antonio, TX: Psychological Corporation.

Welsh, K., Butters, N., Hughes, J., Mohs, R., & Heyman, A. (1991). Detection of abnormal memory decline in mild cases of Alzheimer's disease using CERAD neuropsychological measures. Archives of Neurology, 48, 278-281.

West, R. L. (1986). Everyday memory and aging. Developmental neuropsychology, 2, 323-344.

Wickelgren, W. A. (1975). Age and storage dynamics in continuous recognition memory. Developmental Psychology, 11, 165-169.

Wilkinson, R. T., & Allison, S. (1989). Age and simple reaction time: Decade differences for 5,325 subjects. Journal of Gerontology, 44, P29-P35.

Wilson, R. S., & Kazniak, A. W. (1986). Longitudinal changes: Progressive idiopathic dementia. In L. W. Poon (Ed.), Handbook for clinical memory assessment of older adults (pp. 285-293). Washington, DC: American Psychological Association.

Williams, J. M. (1991). Memory Assessment Scales: Professional Manual. Odessa, FL: Psychological Assessment Resources.

Yesavage, J. A., Brink, T. I., Rose, T. L., Lum, O., Huang, V., Adey, M., & Leirer, V. O. (1983). Development and validation of a geriatric depression rating scale: A preliminary report. Journal of Psychiatric Research, 17, 37-49.

Zacks, R., & Hasher, L. (1988). Capacity theory and the processing of inferences. In I. L. Light and D. M. Burke (Eds.), Language, memory, and aging. New York: Cambridge University Press.

Table 1
Demographic Information For Six Age Groups for Study 1

		Age Groups						
		55-59 (n=32)	60-64 (n=29)	65-69 (n=32)	70-74 (n=30)	75-79 (n=30)	80-85 (n=31)	ALL (n=184)
Age								
	Mean	57.03	62.14	66.97	71.83	77.10	82.32	69.51
	SD	(1.28)	(1.33)	(1.62)	(1.44)	(1.35)	(1.56)	(8.78)
Education								
	Mean	12.41	12.41	12.78	12.33	12.67	12.23	12.47
	SD	(2.55)	(2.40)	(2.57)	(2.15)	(2.56)	(2.38)	(2.42)
Vocab								
	Mean	53.16	57.10	53.72	55.93	53.53	56.52	54.96
	SD	(8.78)	(6.65)	(10.20)	(8.52)	(9.95)	(7.93)	(8.80)
MMSE								
	Mean	28.66	29.14	28.53	28.23	28.23	27.97	28.45
	SD	(1.29)	(0.85)	(1.29)	(1.22)	(1.45)	(1.22)	(1.28)
GDS								
	Mean	4.84	2.59	3.59	3.47	4.20	4.16	3.83
	SD	(4.41)	(2.26)	(2.80)	(3.22)	(3.68)	(3.03)	(3.35)
Gender								
	Males	14	12	14	14	13	18	85
	Females	18	17	18	16	17	13	99

Note. Education is reported in number of years, Vocab = Weschler Adult Intelligence Scale - Revised (WAIS-R) Vocabulary subtest (max. score = 70), MMSE = Mini-Mental State Examination (max. score = 30), GDS = Geriatric Depression Scale (max. score = 30).

Table 2
Correlations Among Demographic Variables in Study 1

Variables	Age	Educ	Vocab	GDS	MMSE
Age	1.000				
Educ	-0.026	1.000			
Vocab	0.032	0.475 ¹	1.000		
MMSE	-0.257 ²	0.229 ¹	0.343	1.000	
GDS	0.027	-0.082	-0.042	-0.064	1.000

Note. Educ = Education and is reported in number of years, Vocab = Weschler Adult Intelligence Scale - Revised (WAIS-R) Vocabulary subtest, MMSE = Mini-Mental State Examination, and GDS = Geriatric Depression Scale.

¹ $p < .01$

² $p < .001$

Table 3
Factor Analysis Pattern Matrix for Diagnostic Test

Variables	Pattern Matrix	
	Factor I	Factor II
Paragraph		
Total Acquisition	.841	.090
Delay (FR)	.909	.070
Delay (FR + CR)	.971	.003
Delay (FR + CR + Rec)	.904	-.103
Rectangle		
Total Acquisition	.042	.944
Delay	-.023	.968

Note. For the Delay trials, FR = free recall, FR + CR = cued recall, and FR + CR + Rec = free recall + cued recall + recognition.

Table 4
Factor Analysis Pattern and Correlation Matrix for Study 1

Variables	Pattern Matrix		
	Factor I	Factor II	Factor III
Rectangle (Tot. Acq.)	.854	-.070	.101
Rectangle (Delay)	.821	-.096	.054
Abstract (Tot. Acq.)	.815	.042	-.027
Abstract (Delay)	.916	.064	-.159
Trails A	-.397	-.072	-.202
Trails B	-.537	-.140	-.150
Verbal Fluency (F)	-.004	.761	-.036
Verbal Fluency (A)	-.020	.875	-.017
Verbal Fluency (S)	.027	.820	-.081
Verbal Fluency (Anim.)	.038	.399	.219
Vocabulary	.061	.367	.265
Paragraph (Tot. Acq.)	-.062	.045	.930
Paragraph (Delay - FR)	-.038	.008	.916
Word List (Tot. Acq.)	.204	.021	.603
Word List (Delay - FR)	.133	-.027	.579

Correlation Matrix			
	Factor I	Factor II	Factor III
Factor I	1.000		
Factor II	.398	1.000	
Factor III	.507	.421	1.000

Note. Tot. Acq. = Total Acquisition, Anim. = Animals, and FR = free recall.

Table 5
Summary of Stepwise Regressions in Study 1

Subtest and Predictors	Step	R ²	R _{adj} ²	r _{sp}	F
Paragraph - Tot. Acq.		.272			22.410 ²
Age	1		.133	-.338	
Education	2		.080	.293	
Gender	3		.059	.242 ¹	
Paragraph - Delay		.207			15.651 ²
Age	1		.100	-.294	
Education	2		.068	.268	
Gender	3		.039	.197 ¹	
Rectangle - Tot. Acq.		.232			27.357 ²
Age	1		.184	-.425	
Education	2		.048	.218	
Rectangle - Delay		.161			17.393 ²
Age	1		.124	-.355	
Education	2		.037	.188	

Note. Tot. Acq. = Total Acquisition.

¹ A positive semi-partial correlation indicates better performance by females than males on the subtest.

² All F-tests are significant at $p < .001$ ($N = 184$).

Table 6
Performance on Paragraph Over Six Age Groups for Study 1

		Age Groups						
		55-59 (n=32)	60-64 (n=29)	65-69 (n=32)	70-74 (n=30)	75-79 (n=30)	80-85 (n=31)	ALL (n=184)
<hr/>								
Trial 1								
Mean		21.72	22.66	20.13	19.77	17.93	16.26	19.73
SD		(5.19)	(3.60)	(6.06)	(5.39)	(6.25)	(5.20)	(5.71)
Trial 2								
Mean		27.94	28.76	27.28	27.27	25.70	24.74	26.94
SD		(3.19)	(2.15)	(4.73)	(3.35)	(5.58)	(5.16)	(4.37)
Tot. Acq.								
Mean		49.66	51.41	47.41	47.03	43.63	41.00	46.67
SD		(7.86)	(5.24)	(10.24)	(7.95)	(11.16)	(9.88)	(9.51)
Delay 1								
Mean		25.88	26.55	25.06	25.17	22.43	21.68	24.46
SD		(4.43)	(3.50)	(5.22)	(3.80)	(7.44)	(5.67)	(5.41)
Delay 2								
Mean		28.22	28.14	27.69	27.67	25.73	25.35	27.14
SD		(2.92)	(2.80)	(4.50)	(2.47)	(5.79)	(4.03)	(4.04)
Delay 3								
Mean		30.13	30.28	29.87	30.00	28.77	28.97	29.67
SD		(1.52)	(1.00)	(2.47)	(1.20)	(2.78)	(2.36)	(2.07)

Note. Maximum score on all of the Paragraph subtest trials is 31. Maximum score on Total Acquisition (Tot. Acq.) is 62. On the Delay trials, Delay 1 = free recall, Delay 2 = free recall + cued recall, Delay 3 = free recall + cued recall + recognition.

Table 7
Performance on Rectangle Over Six Age Groups for Study 1

		Age Groups						
		55-59 (n=32)	60-64 (n=29)	65-69 (n=32)	70-74 (n=30)	75-79 (n=30)	80-85 (n=31)	ALL (n=184)
Trial 1								
Mean		17.56	18.38	16.61	15.48	13.00	12.23	15.54
SD		(4.76)	(3.98)	(4.02)	(4.21)	(4.85)	(3.84)	(4.80)
Trial 2								
Mean		22.64	23.09	22.22	20.50	18.50	17.94	20.82
SD		(3.72)	(2.91)	(3.10)	(4.86)	(5.20)	(5.11)	(4.65)
Trial 3								
Mean		23.87	24.26	23.72	22.78	21.03	20.89	22.76
SD		(3.77)	(2.66)	(2.23)	(4.24)	(4.90)	(4.55)	(4.03)
Tot. Acq.								
Mean		64.08	65.72	62.55	58.77	52.53	51.05	59.13
SD		(11.07)	(8.98)	(8.48)	(12.17)	(14.03)	(12.78)	(12.57)
Delay								
Mean		23.66	23.98	23.63	22.45	20.57	19.95	22.38
SD		(3.95)	(2.74)	(2.23)	(4.27)	(5.13)	(5.01)	(4.28)
Copy								
Mean		25.69	25.66	25.66	25.40	25.17	24.87	25.41
SD		(0.74)	(1.01)	(0.60)	(1.10)	(1.18)	(1.75)	(1.15)

Note. Maximum score on all of the Rectangle subtest trials is 26. Maximum score on Total Acquisition (Tot. Acq.) is 78.

Table 8
Demographic Information for Three Age Groups in Study 1

	Age Groups			
	55-64 (n=61)	65-74 (n=62)	75-85 (n=61)	ALL (n=184)
Age				
Mean	59.46	69.32	79.75	69.51
SD	(2.88)	(2.88)	(3.00)	(8.78)
Education				
Mean	12.41	12.56	12.44	12.47
SD	(2.46)	(2.37)	(2.46)	(2.42)
Vocab				
Mean	55.03	54.79	55.05	54.96
SD	(8.03)	(9.41)	(9.03)	(8.80)
MMSE				
Mean	28.88	28.39	28.10	28.45
SD	(1.12)	(1.26)	(1.34)	(1.28)
GDS				
Mean	3.77	3.53	4.18	3.83
SD	(3.71)	(2.99)	(3.34)	(3.35)
Gender				
Males	26	28	31	85
Females	35	34	30	99

Note. Education is reported in number of years, Vocab = Weschler Adult Intelligence Scale - Revised (WAIS-R) Vocabulary subtest (max. score = 70), MMSE = Mini-Mental State Examination (max. score = 30), GDS = Geriatric Depression Scale (max. score = 30).

Table 9
Paragraph Performance for Three Age Groups in Study 1

	Age Groups			
	55-64 (n=61)	65-74 (n=62)	75-85 (n=61)	ALL (n=184)
Trial 1				
Mean	22.16	19.95	17.08	19.73
SD	(4.49)	(5.70)	(5.75)	(5.71)
Trial 2				
Mean	28.33	27.27	25.21	26.94
SD	(2.76)	(4.09)	(5.35)	(4.37)
Total Acq.				
Mean	50.49	47.23	42.30	46.67
SD	(6.74)	(9.13)	(10.52)	(9.51)
Delay 1				
Mean	26.20	25.11	22.05	24.46
SD	(4.00)	(4.55)	(6.55)	(5.41)
Delay 2				
Mean	28.18	27.68	25.54	27.14
SD	(2.84)	(3.63)	(4.94)	(4.04)
Delay 3				
Mean	30.20	29.94	28.87	29.67
SD	(1.29)	(1.95)	(2.55)	(2.07)

Note. Maximum score on all of the Paragraph subtest trials is 31. Maximum score on Total Acquisition (Tot. Acq.) is 62. On the Delay trials, Delay 1 = free recall, Delay 2 = free recall + cued recall, and Delay 3 = free recall + cued recall + recognition.

Table 10
Rectangle Figure Performance for Three Age Groups in Study 1

	Age Groups			
	55-64 (n=61)	65-74 (n=62)	75-85 (n=61)	ALL (n=184)
Trial 1				
Mean	17.95	16.06	12.61	15.54
SD	(4.39)	(4.12)	(4.34)	(4.80)
Trial 2				
Mean	22.85	21.39	18.21	20.82
SD	(3.34)	(4.11)	(5.12)	(4.65)
Trial 3				
Mean	24.06	23.27	20.96	22.76
SD	(3.27)	(3.36)	(4.69)	(4.03)
Total Acq.				
Mean	64.86	60.72	51.78	59.13
SD	(10.08)	(10.51)	(13.31)	(12.57)
Delay (FR)				
Mean	23.81	23.06	20.25	22.38
SD	(3.40)	(3.40)	(5.03)	(4.28)
Copy				
Mean	25.67	25.53	25.02	25.41
SD	(0.87)	(0.88)	(1.49)	(1.15)

Note. Maximum score on all of the Rectangle Figure subtest trials is 26. Maximum score on Total Acquisition (Tot. Acq.) is 78.

Table 11
Standard Scores for the Paragraph and Rectangle Figure
Subtest Total Acquisition Scores in Study 1

Standard Scores	Age Groups					
	55-64 yrs.		65-74 yrs.		75-85 yrs.	
	Para.	Fig.	Para.	Fig.	Para.	Fig.
135		78	62	76	59	73
134		77.5				72.5
133				75.5		72
132		77	61			
131	60	76.5		75		71.5
130						71
129		76	60		58	70.5
128				74.5		70
127		75.5				
126			59	74	57	69.5
125		75				
124				73.5		69
123			58			68.5
122	59	74.5		73	56	68
121						67.5
120				72.5	55	67
119		74	57			66.5
118	58	73.5		72		66
117	57			71.5	54	
116		73	56	71		65.5
115	56	72.5		70.5		65
114		72		70	53	
113						64.5
112		71.5	55	69-69.5	52	
111		71		68-68.5	51	64
110	55	70.5	54	67-67.5	50	63.5
109		70		66.5		63
108		69.5	53		49	62.5
107				66		61-62
106					48	59-60.5
105	54	69	52	65.5	47	58.5
104			51	65		
103		68.5	50	64.5	46	57.5-58

Table 11 cont'd

Standard Scores	55-64 yrs.		65-74 yrs.		75-85 yrs.	
	Para.	Fig.	Para.	Fig.	Para.	Fig.
102	53	68	49	63.5-64	45	56-57
101				63	44	55.5
100				62.5		54-55
99	52	67.5	48	62	43	52.5-53.5
98		67	47	61-61.5		51.5-52
97	51			60.5	42	50.5-51
96		66.5	46	59.5-60		48-50
95	50	66		58.5-59	41	47-47.5
94		65.5	45			45-46.5
93	49		44	58	40	43-44.5
92	48	65		57.5	38-39	41.5-42.5
91	47	64.5	43	57	36-37	41
90	46	63-64		56.5-56.5	33-35	40.5
89		61-62.5		54.5-55	32	40
88	45	59.5-60.5	41-42	53.5-54		
87		58-59			31	39.5
86	44	56.5-57.5	40	53		38-39
85		56	39	52.5		36.5-37.5
84	43	55-55.5	38	49-50.5	30	35-36
83		54.5		48-48.5		33.5-34.5
82		54	37	47.5		
81	42	52.5-53.5	36	47	29	33
80	41	50.5-52	35		28	32.5
79	40	49.5-50	34			
78	39	47-49	33	46-46.5	27	31.5-32
77	38	44.5-46		45-45.5	25-26	31
76	37	44		43.5-44.5	23-24	30-30.5
75	36		32	42.5-43	22	29.5
74	35	43-43.5	30-31	41-42		28.5-29
73		42-42.5	28-29	39-40.5	21	27-28
72	34	41-41.5	26-27	37-38.5		25.5-26.5
71	33	38.5-40.5	24-25	35-36.5	20	24-25
70		35-38	22-23	31.5-34.5	19	23-23.5
69	32	31.5-34.5	20-21	28-31	18	21.5-22.5
68		28-31	17-19	25-27.5	17	20.5-21
67		24.5-27.5	15-16	21.5-24.5	16	19-20
66		21-24	13-14	18-21	15	18-18.5
65		19-20.5	11-12	16-17.5	14	17-17.5

Table 12
Standard Scores for the Paragraph and Rectangle Figure
Subtest Delay Scores in Study 1

Standard Scores	Age Groups					
	55-64 yrs.		65-74 yrs.		75-85 yrs.	
	Para.	Fig.	Para.	Fig.	Para.	Fig.
130						
129					30	
128						
127			31			
126						
125						26
124						
123	31					
122						
121						
120						25.5
119			30	26	29	
118						
117						
116						
115	30					25
114			29	25.5		
113		26				
112					28	24.5
111						
110	29					
109			28	25		24
108						
107		25.5			27	23.5
106						
105				24.5		
104	28		27			23
103					26	22.5
102						
101		25			25	22
100			26	24	24	21.5
99					23	
98	27	24.5			22	21

Table 12 cont'd

Standard Scores	55-64 yrs.		65-74 yrs.		75-85 yrs.	
	Para.	Fig.	Para.	Fig.	Para.	Fig.
97			25	23.5	21	20.5
96						
95	26		24		20	20
94		24				19-19.5
93				23	19	18-18.5
92	25			22.5		17.5
91			23		18	17
90	24	23.5		22		16.5
89		23				16
88		22.5			17	15.5
87	23		22	21.5	16	
86		22			15	15
85						
84	22	21-21.5		21	14	14.5
83	21	20.5		20.5	13	
82			21	19.5-20		
81	20	19.5-20		18.5-19	12	14
80		19	20	17.5-18		
79	19					
78		18.5	19	17	11	11.5-13.5
77	18		18	16.5	10	9.5-11
76		18	17	16		
75			16	15.5		
74	17	17.5	15	15	9	9
73		17		14.5		8-8.5
72				14		7.5
71	16	16-16.5	14	13-13.5	8	7
70		14-15.5	12-13	12.5		
69		12-13.5	11	11.5-12		
68		10-11.5	10	10.5-11	7	
67		8-9.5	8-9	9.5-10		
66		6-7.5	7	8.5-9		
65	15	5-5.5	6	8	6	6.5

Table 13
Cumulative Percentiles for Adults Aged 55-64 Years on the
Paragraph Subtest Trial-by-Trial Scores in Study 1

55-64 years

Cumulative Percentiles	Trial 1	Trial 2	Delay 1	Delay 2	Delay 3
100	29	31	31	31	31
99					
98					
97					
96					
95					
94					
93					
92	28				
91					
90					
89			30		
88					
87	27				
86					
85	26				
84					
83					
82					
81					
80			29		
79	25	30		30	
78					
77					
76					
75					
74					
73					
72					
71					
70			28		
69					
68					

Table 13 cont'd

55-64 years

Cumulative Percentiles	Trial 1	Trial 2	Delay 1	Delay 2	Delay 3
67	24				
66					
65					
64					
63					
62					
61				29	
60					
59					
58					
57					
56		29			
55					
54	23				
53					
52					
51					
50					
49					
48					
47					
46	22				
45					
44					
43					
42					
41	21			28	
40					
39					30
38		28			
37					
36					
35					
34					
33					

3

OF/DE

3

PM-1 3 1/4" x 4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

PRECISIONSM RESOLUTION TARGETS

Table 13 cont'd

55-64 years

Cumulative Percentiles	Trial 1	Trial 2	Delay 1	Delay 2	Delay 3
32					
31					
30	20			27	
29					
28	19	27	24		
27					
26					
25					
24					
23		26	23	26	
22					
21					29
20	18				
19					
18	17				
17					
16			22	25	
15					
14			21		
13	16	25			
12					
11		24	20		28
10			19	24	
9					
8	15	23		23	
7	14	22	18	22	
6					
5	13		17	21	27
4					26
3	12	21	16	20	
2	11	20	15		25
1				19	

Table 14
*Cumulative Percentiles for Adults Aged 65-74 Years on the
 Paragraph Subtest Trial-by-Trial Scores in Study 1*

65-74 years

Cumulative Percentiles	Trial 1	Trial 2	Delay 1	Delay 2	Delay 3
100	31	31	31	31	31
99	30				
98					
97	29				
96					
95	28				
94			30		
93					
92	27				
91					
90	26				
89					
88					
87	25		29		
86					
85					
84		30			
83					
82					
81				30	
80					
79					
78					
77	24		28		
76					
75					
74					
73					
72					
71					
70					
69					
68	23				

Table 14 cont'd

65-74 years					
Cumulative Percentiles	Trial 1	Trial 2	Delay 1	Delay 2	Delay 3
67					
66			27		
65		29			
64					
63				29	
61	22				
60					
59					
58					
57					
56					
55			26		
54					
53	21	28			
52				28	
51					
50	20				30
49					
48	19				
47					
46					
45					
44			25		
43					
42		27		27	
41					
40			24		
39					
38					
37	18				
36					
35	17				
34					
33					
32			23		

Table 14 cont'd

65-74 years					
Cumulative Percentiles	Trial 1	Trial 2	Delay 1	Delay 2	Delay 3
31	16				
30					
29		26			
28					
27				26	
26					
25					
24			22		
23	15				
22					
21					29
20					
19					
18	14	25		25	
17		24			
16					
15		23	21		
14				24	
13	13				28
12			20		
11		22			
10	12	21		23	
9					27
8		20	19		
7			18		
6			16-17		
5	11	19	15	22	25-26
4	9-10	18		21	22-24
3	6-8	13-17	10-14	15-20	20-21
2	3-5	8-12	6-9	8-14	18-19
1					

Table 15
Cumulative Percentiles for Adults Aged 75-85 Years on the
Paragraph Subtest Trial-by-Trial Scores in Study 1

75-85 years:

Cumulative Percentiles	Trial 1	Trial 2	Delay 1	Delay 2	Delay 3
100	29	31	30	31	31
99					
98	28				
97	27				
96					
95	26		29	30	
94					
93	25				
92	24				
91					
90					
89		30			
88					
87	23				
86					
85			28		
84					
83					
82					
81					
80	22				
79					
78					
77	21			29	
76					
75		29			
74					30
73					
72			27		
71					
70	20				
69					
68					

Table 15 cont'd

75-85 years

Cumulative Percentiles	Trial 1	Trial 2	Delay 1	Delay 2	Delay 3
67					
66					
65					
64	19			28	
63					
62		28	26		
61					
60					
59	18				
58					
57		27			
56					
55					
54			25		
53					
52	17	26	24	27	
51					
50					
49			23	26	
48			22		
47					
46					
45					
44	16	25	21		29
43					
42					
41					
40					
39			20	25	
38					
37					
36	15	24			
35					
34			19	24	
33					

Table 15 cont'd

75-85 years

Cumulative Percentiles	Trial 1	Trial 2	Delay 1	Delay 2	Delay 3
32					
31	14				28
30			18		
29					
28	13				
27					
26		23		23	
25		22			
24					
23		21	17		
22					
21	12			22	27
20	11		16		
19					
18	10	20	15	21	
17					
16			14		
15	9		13	20	
14					
13		19	12		26
12		18			
11		17			
10		16		19	25
9					24
8	8	15	11	17-18	
7		14	10	15-16	23
6	7				22
5		13	9	13-14	
4				11-12	21
3	6	10-12	7-8	9-10	19-20
2		8-9	6	7-8	18
1					

Table 16
*Cumulative Percentiles for Adults Aged 55-64 Years on the
 Rectangle Figure Subtest Trial-by-Trial Scores in Study 1*

55-64 years

Cumulative Percentiles	Trial 1	Trial 2	Trial 3	Delay	Copy
100	25.5-26	26	26	26	26
99	24.5-25				
98	24				
97					
96	23.5				
95					
94		25.5			
93	23				
92					
91					
90					
89	22.5				
88					
87		25			
86					
85	22				
84					
83					
82					
81	21.5			25.5	
80			25.5		
79					
78		24.5			
77	21				
76	20.5				
75	20				
74					
73					
72					
71					
70					
69					
68					

Table 16 cont'd

55-64 years

Cumulative Percentiles	Trial 1	Trial 2	Trial 3	Delay	Copy
67		24			
66					
65					
64					
63					
62				25	
61	19.5				
60					
59			25		25.5
58					
57					
56	19				
55		23.5			
54					
53				24.5	
52					
51					
50					
49					
48	18.5				
47					
46					
45					
44				24	
43	18	23			
42					
41	17.5				
40					
39	17		24.5		
38			24		
37					
36					
35					
34	16.5				
33					

Table 16 cont'd

55-64 years

Cumulative Percentiles	Trial 1	Trial 2	Trial 3	Delay	Copy
32					
31					
30	16	22.5	23.5		
29					
28		22			
27					
26				23.5	
25	15.5			23	
24					
23	15			22.5	
22	14.5				
21			23		
20	14	21.5		22	
19					
18		21	22.5		25
17					
16	13.5				
15	13	20.5	22	21.5	
14	12.5	19.5-20	21.5	21	
13	12	19	21	20.5	24.5
12	11.5	18.5			
11	11	18		19.5-20	
10	10.5		20.5	19	
9			19.5-20		
8	10	17.5	19	18.5	24
7		16.5-17	18.5	18	
6	9.5	15.5-16	17.5-18		23.5
5		14.5-15	15-17	17.5	
4		13-14	11.5-14.5	17	
3	6.5-9	10.5-12.5	8-11	11-16.5	22-23
2	4-6	9-10	6-7.5	5-10.5	21-21.5
1					

Table 17
Cumulative Percentiles for Adults Aged 65-74 Years on the
Rectangle Figure Subtest Trial-by-Trial Scores in Study 1

65-74 years

Cumulative Percentiles	Trial 1	Trial 2	Trial 3	Delay	Copy
100	24	26	26	26	26
99	23.5				
98	23				
97					
96					
95	22.5				
94		25.5			
93					
92	22				
91				25.5	
90			25.5		
89	21.5				
88					
87	21	25			
86					
85	20.5				
84	20				
83					
82	19.5				
81				25	
80					
79	19		25		
78					
77					
76	18.5	24.5			
75					
74		24			
73					
72				24.5	
71					
70		23.5			
69	18				
68					

Table 17 cont'd

65-74 years

Cumulative Percentiles	Trial 1	Trial 2	Trial 3	Delay	Copy
67					
66			24.5		25.5
65		23			
64					
63	17.5			24	
62					
61	17				
60					
59		22.5			
58	16.5				
57					
56					
55					
54					
53	16	22	24		
52					
51					
50					
49		21.5		23.5	
48	15.5				
47					
46					
45	15				
44		21	23.5		
43					
42					
41	14.5				
40					
39					
38					
37	14				
36					
35			23	23	
34		20.5			
33					

Table 17 cont'd

65-74 years

Cumulative Percentiles	Trial 1	Trial 2	Trial 3	Delay	Copy
32					
31	13.5	20		22.5	25
30					
29		19.5		22	
28					
27		19	22.5		
26			22		
25					
24					
23	13				
22					
21		18.5		21.5	24.5
20					
19	12.5		21.5		
18	12				
17					
16		18	21	21	
15		17.5			
14		16.5-17			
13	11.5	16	20.5	20.5	
12			19.5-20	19.5-20	
11			19	18.5-19	
10	11		18.5	17.5-18	24
9					
8	10.5	15.5	18	17	23.5
7		15		16.5	23
6	9.5-10	14.5	17-17.5	15.5-16	
5	9	13.5-14	16-16.5	15	22.5
4	8.5	12.5-13	13.5-15	14-14.5	
3	6-8	8-12	10-13	11-13.5	22
2	4-5.5	4-7.5	8-9.5	8-10.5	
1					

Table 18
Cumulative Percentiles for Adults Aged 75-85 Years on the
Rectangle Figure Subtest Trial-by-Trial Scores in Study 1

75-85 years

Cumulative Percentiles	Trial 1	Trial 2	Trial 3	Delay	Copy
100	21	26	26	26	26
99		25.5			
98	20.5				
97	20	25			
96					
95	19.5	24.5	25.5	25.5	
94					
93	18.5-19				
92	18	24			
91					
90				25	
89			25		
88					
87	17.5				
86					
85				24.5	
84					
83					
82	17	23.5	24.5		
81					
80					
79		23		24	
78					
77	16.5	22.5			
76					
75			24		
74	16	22			25.5
73				23.5	
72					
71					
70					
69	15.5	21.5			
68					

Table 18 cont'd

75-85 years

Cumulative Percentiles	Trial 1	Trial 2	Trial 3	Delay	Copy
67	15	21			
66	14.5		23.5	23	
65					
64	14				
63					
62	13.5	20.5			
61		20			
60				22.5	
59	13				
58					
57	12.5	19.5	23		
56					
55					
54	12	19		22	
53			22.5	21.5	
52					
51				21	
50					
49	11.5				
48			22		25
47					
46	11	18.5	21.5		
45				20.5	
44		18			
43		17.5	21		
42					
41		17			
40					
39	10.5		20.5		
38		16.5		20	
37				19-19.5	24.5
36	10	16		18.5	
35					
34			20		
33		15.5	19.5	18	

Table 18 cont'd

75-85 years

Cumulative Percentiles	Trial 1	Trial 2	Trial 3	Delay	Copy
32		15	19		
31		14.5			
30		14	18.5	17.5	
29				17	
28	9.5		18	16.5	
27					
26			17.5		24
25	9	13.5			
24					
23			17	16	
22					
21	8.5		16.5	15.5	
20	8	13		15	
19					
18			16		23.5
17			15.5		
16	7.5	12.5	15		
15		12	14.5	14.5	
14					
13				14	
12					
11	7	11.5	14		
10		11	13.5		23
9					22.5
8	6.5	10-10.5		11.5-13.5	
7		9.5	13	9.5-11	22
6					
5	6	9	12-12.5	8.5-9	21.5
4	5.5	8.5	10-11.5	7.5-8	
3		7-8	8-9.5	7	19.5-21
2	5	5.5-6.5	6.5-7.5	6.5	18-19
1					

Table 19
Standard Error of Measurement (SEM) for Paragraph and
Rectangle Figure Subtests in Study 1

Subtests	Age Groups			
	55-64	65-74	75-85	ALL
Paragraph¹				
Trial 1	2.22	2.33	2.41	2.32
Trial 2	1.49	1.60	1.89	1.67
Tot. Acq.	2.68	2.85	3.07	2.87
Delay (FR)	1.85	1.97	2.18	2.01
Rectangle²				
Trial 1	2.34	2.41	2.50	2.42
Trial 2	1.62	1.90	2.19	1.93
Trial 3	1.29	1.55	1.93	1.62
Tot. Acq.	3.14	3.44	3.83	3.50
Delay	1.33	1.61	1.98	1.67
Copy	0.62	0.72	1.03	0.81

¹ Maximum score on each trial of the Paragraph subtest is 31. The maximum score on total acquisition (Tot. Acq.) is 62.

² Maximum score on each trial of the Rectangle subtest is 26. The maximum score on total acquisition (Tot. Acq.) is 78.

Table 20
Demographic Information for Six Age Groups in Study 2

		Age Groups						
		55-59 (n=27)	60-64 (n=30)	65-69 (n=32)	70-74 (n=32)	75-79 (n=30)	80-85 (n=29)	ALL (n=180)
Age								
Mean		56.63	62.20	67.03	71.84	76.73	82.38	69.61
SD		(1.47)	(1.35)	(1.62)	(1.27)	(1.14)	(1.61)	(8.58)
Education								
Mean		12.74	12.73	12.84	12.56	12.97	12.28	12.69
SD		(2.35)	(2.68)	(2.62)	(1.93)	(2.77)	(2.30)	(2.43)
Vocab								
Mean		55.00	55.33	54.81	55.78	55.50	57.28	55.61
SD		(7.17)	(7.85)	(10.42)	(9.22)	(8.88)	(7.30)	(8.53)
MMSE								
Mean		28.67	29.10	28.44	28.25	28.27	27.97	28.44
SD		(1.47)	(0.88)	(1.34)	(1.16)	(1.39)	(1.40)	(1.32)
GDS								
Mean		4.30	2.57	4.75	4.31	4.67	5.28	4.31
SD		(3.73)	(2.40)	(2.87)	(2.99)	(2.94)	(2.78)	(3.04)
Gender								
Males		13	9	15	17	13	16	83
Females		14	21	17	15	17	13	97

Note. Education is reported in number of years, Vocab = Weschler Adult Intelligence Scale - Revised (WAIS-R) Vocabulary subtest (max. score = 70), MMSE = Mini-Mental State Examination (max. score = 30), GDS = Geriatric Depression Scale (max. score = 30).

Table 21
Correlations Among Demographic Variables in Study 2

Variables	Age	Educ	Vocab	GDS	MMSE
Age	1.000				
Educ	-0.035	1.000			
Vocab	0.074	0.433 ¹	1.000		
MMSE	-0.225 ¹	0.263 ²	0.311 ¹	1.000	
GDS	0.159	-0.059	0.005	-0.064	1.000

Note. Educ = Education and is reported in number of years,
 Vocab = Weschler Adult Intelligence Scale - Revised (WAIS-R)
 Vocabulary subtest, MMSE = Mini-Mental State Examination,
 and GDS = Geriatric Depression Scale.

¹ $p < .01$

² $p < .001$

Table 22
Factor Analysis Pattern Matrix for Tracking Test

Variables	Pattern Matrix	
	Factor I	Factor II
Paragraph		
Total Acquisition	.844	-.008
Delay (Free Recall)	.923	-.056
Delay (Recognition)	.718	.072
Triangle		
Total Acquisition	.055	.891
Delay	-.043	.946

Table 23
Factor Analysis Pattern and Correlation Matrix for Study 2

Pattern Matrix				
Variables	Factor I	Factor II	Factor III	Factor IV
Paragraph				
Tot. Acq.	.848	.060	.045	-.105
Delay - FR	.892	.026	.004	-.040
Delay - Rec.	.684	.078	-.005	.143
Vocabulary	.267	-.059	.077	.245
Triangle				
Tot. Acq.	.086	.901	-.008	.037
Delay	.018	.897	.015	-.004
Word				
Tot. Acq.	.008	-.004	.887	-.051
Delay	-.010	.002	.907	.037
Picture				
Tot. Acq.	.107	-.068	-.026	.785
Delay	-.079	.113	.044	.429
Correlation Matrix				
	Factor I	Factor II	Factor III	Factor IV
Factor I	1.000			
Factor II	.301	1.000		
Factor III	.431	.221	1.000	
Factor IV	.360	.274	.380	1.000

Note. Tot. Acq. = Total Acquisition, FR = free recall, and Rec = recognition.

Table 24
Summary of Stepwise Regressions in Study 2

Subtest and Predictors	Step	R	R _h	r _{sp}	F
Paragraph - Tot. Acq.		.137			9.326 ¹
Age	1		.068	-.238	
Gender	2		.034	.205 ¹	
Education	3		.036	.189	
Paragraph - Delay		.102			10.023 ¹
Education	1		.054	.236	
Age	2		.048	-.211	
Triangle - Tot. Acq.		.222			25.312 ¹
Age	1		.203	-.451	
Education	2		.019	.128	
Triangle - Delay		.201			22.301 ¹
Age	1		.175	-.422	
Gender	2		.027	-.147 ²	

Note. Tot. Acq. = Total Acquisition.

¹ A positive semi-partial correlation indicates better performance by females than males on the subtest.

² A negative semi-partial correlation indicates better performance by males than females on the subtest.

³ All F-tests are significant at $p < .001$ ($N = 180$).

Table 25
Performance on Paragraph Over Six Age Groups for Study 2

		Age Groups						
		55-59 (n=27)	60-64 (n=30)	65-69 (n=32)	70-74 (n=32)	75-79 (n=30)	80-85 (n=29)	ALL (n=180)
Trial 1								
Mean		17.96	17.80	16.50	16.25	16.73	15.21	16.72
SD		(2.43)	(2.09)	(3.53)	(2.77)	(3.60)	(3.12)	(3.09)
Trial 2								
Mean		20.04	20.00	19.00	19.22	19.43	18.72	19.39
SD		(1.60)	(1.29)	(2.96)	(1.81)	(2.19)	(2.05)	(2.10)
Tot. Acq.								
Mean		38.00	37.80	35.50	35.47	36.17	33.93	36.11
SD		(3.58)	(2.75)	(6.20)	(3.83)	(5.38)	(4.74)	(4.73)
Delay 1								
Mean		19.26	18.77	17.84	18.03	18.17	17.17	18.19
SD		(1.72)	(2.25)	(3.25)	(2.29)	(2.49)	(3.27)	(2.67)
Delay 2								
Mean		20.93	20.83	20.62	20.69	20.66	20.31	20.66
SD		(0.27)	(0.38)	(0.87)	(0.54)	(0.62)	(1.00)	(0.69)

Note. Maximum score on all of the Paragraph subtest trials is 21. Maximum score on Total Acquisition (Tot. Acq.) is 42. On the Delay trials, Delay 1 = free recall, Delay 2 = recognition.

Table 26
Performance on Triangle Over Six Age Groups for Study 2

		Age Groups						
		55-59 (n=27)	60-64 (n=30)	65-69 (n=32)	70-74 (n=32)	75-79 (n=30)	80-85 (n=29)	ALL (n=180)
Trial 1								
Mean		20.35	20.20	19.39	17.47	16.78	15.22	18.22
SD		(4.07)	(4.04)	(3.40)	(4.18)	(5.04)	(4.18)	(4.52)
Trial 2								
Mean		23.98	23.80	24.33	21.91	20.47	20.05	22.42
SD		(2.65)	(3.05)	(1.61)	(3.91)	(5.17)	(4.20)	(3.96)
Trial 3								
Mean		25.50	25.13	25.33	23.27	23.07	22.29	24.09
SD		(1.22)	(1.57)	(1.03)	(3.26)	(3.26)	(3.23)	(2.75)
Tot. Acq.								
Mean		69.83	69.13	69.05	62.64	60.32	57.57	64.74
SD		(6.89)	(7.53)	(5.11)	(10.65)	(12.29)	(10.12)	(10.18)
Delay								
Mean		25.35	25.07	25.30	22.88	23.10	21.86	23.92
SD		(1.62)	(1.80)	(1.25)	(4.08)	(2.91)	(3.51)	(3.04)
Copy								
Mean		25.96	25.97	25.88	25.84	25.80	25.76	25.87
SD		(0.19)	(0.18)	(0.42)	(0.37)	(0.48)	(0.69)	(0.43)

Note. Maximum score on all of the Triangle subtest trials is 26. Maximum score on Total Acquisition (Tot. Acq.) is 78.

Table 27
Demographic Information for Three Age Groups for Study 2

Age Groups				
	55-64 (n=57)	65-74 (n=64)	75-85 (n=59)	ALL (n=180)
Age				
Mean	59.56	69.44	79.51	69.61
SD	(3.13)	(2.82)	(3.16)	(8.58)
Education				
Mean	12.74	12.70	12.63	12.69
SD	(2.50)	(2.29)	(2.55)	(2.43)
Vocab				
Mean	55.18	55.30	56.37	55.61
SD	(7.47)	(9.77)	(8.12)	(8.53)
MMSE				
Mean	28.89	28.34	28.12	28.44
SD	(1.21)	(1.25)	(1.39)	(1.32)
GDS				
Mean	3.39	4.53	4.97	4.31
SD	(3.19)	(2.92)	(2.85)	(3.04)
Gender				
Males	22	32	29	83
Females	35	32	30	97

Note. Education is reported in number of years, Vocab = Weschler Adult Intelligence Scale - Revised (WAIS-R) Vocabulary subtest (max. score = 70), MMSE = Mini-Mental State Examination (max. score = 30), GDS = Geriatric Depression Scale (max. score = 30).

Table 28
Paragraph Performance for Three Age Groups in Study 2

	Age Groups			
	55-64 (n=57)	65-74 (n=64)	75-85 (n=59)	ALL (n=180)
Trial 1				
Mean	17.88	16.38	15.98	16.72
SD	(2.24)	(3.15)	(3.43)	(3.09)
Trial 2				
Mean	20.02	19.11	19.08	19.39
SD	(1.43)	(2.44)	(2.14)	(2.10)
Total Acq.				
Mean	37.89	35.48	35.07	36.11
SD	(3.14)	(5.11)	(5.16)	(4.73)
Delay (FR)				
Mean	19.00	17.94	17.68	18.19
SD	(2.02)	(2.79)	(2.92)	(2.67)
Delay (Rec.)				
Mean	20.88	20.66	20.46	20.66
SD	(0.33)	(0.72)	(0.84)	(0.69)

Note. Maximum score on all of the Paragraph subtest trial is 21. Maximum score on Total Acquisition (Tot. Acq.) is 42. On the Delay trials, FR = free recall and Rec. = recognition.

Table 29
Triangle Figure Performance for Age Groups in Study 2

	Age Groups			
	55-64 (n=57)	65-74 (n=64)	75-85 (n=59)	ALL (n=180)
Trial 1				
Mean	20.27	18.43	16.02	18.22
SD	(4.02)	(3.90)	(4.66)	(4.52)
Trial 2				
Mean	23.89	23.12	20.26	22.42
SD	(2.84)	(3.21)	(4.69)	(3.96)
Trial 3				
Mean	25.31	24.30	22.69	24.09
SD	(1.41)	(2.61)	(3.24)	(2.75)
Total Acq.				
Mean	69.46	65.84	58.97	64.74
SD	(7.18)	(8.90)	(11.26)	(10.18)
Delay (FR)				
Mean	25.20	24.09	22.49	23.92
SD	(1.71)	(3.23)	(3.25)	(3.04)
Copy				
Mean	25.96	25.86	25.78	25.87
SD	(0.19)	(1.39)	(0.59)	(0.43)

Note. Maximum score on all of the Triangle Figure subtest trials is 26. Maximum score on Total Acquisition (Tot. Acq.) is 78.

Table 30
Standard Scores for the Paragraph and Triangle Figure
Subtest Total Acquisition Scores in Study 2

Standard Scores	Age Groups					
	55-64 yrs.		65-74 yrs.		75-85 yrs.	
	Para.	Fig.	Para.	Fig.	Para.	Fig.
135						
134						
133						
132				77		
131						
130						
129	42	78				77
128			42	76.5		
127						
126					42	76.5
125		77.5				
124				76		
123						76
122				75.5		75.5
121						74.5-75
120		77	41	74.5-75		73.5-74
119				74		
118	41	76.5		73.5		73
117						72.5
116				73	41	
115		76	40	72.5		71.5-72
114						71
113				72		70.5
112		75.5			40	69-70
111	40	75		71.5		67.5-68.5
110		74.6	39			67
109		74			39	66.5
108		73.5		71		65.5-66
107						65
106		73				64.5
105			38	70.5	38	64
104	39	72.5		70		
103				69.5		61.5-63.5

Table 30 cont'd

Standard Scores	55-64 yrs.		65-74 yrs.		75-85 yrs.	
	Para.	Fig.	Para.	Fig.	Para.	Fig.
102		72		69	37	60.5-61
101			37	68.5		60
100		71.5			36	59.5
99		71	36	68		59
98				67.5	35	58.5
97	38	70.5				57-58
96		70	35	66.5-67	34	55.5-56.5
95				66		55
94		69.5		65.5	33	54-54.5
93		69	34	65		53-53.5
92		68-68.5		64-64.5	32	51.5-52.5
91	37	67-67.5		62.5-63.5		50.5-51
90		66.5	33	62		50
89		65-66		61-61.5	31	49.5
88		63.5-64.5		60-60.5	30	48.5-49
87	36	63	32	59-59.5		47.5-48
86		62.5		58.5	29	47
85		61.5-62		58		46-46.5
84		60.5-61				45.5
83	35	60	31	57-57.5		44.5-45
82		59.5		55.5-56.5	28	44
81		59		52.5-55		
80	34	58.5	30	50-52		43.5
79		58	29	49-49.5		43
78		57-57.5	28	48.5	27	
77	32-33	55.5-56.5	27	47.5-48		42.5
76	30-31	52.5-55	26	47		
75	29	51-52		46.5		
74		50.5	24-25	45.5-46	26	41.5-42
73			22-23	45		40-41
72		50	20-21	44-44.5		38.5-39.5
71	28	49.5	19	42-43.5	25	37-38
70				40.5-41.5		35.5-36.5
69			18	38.5-40	24	34.5-35
68				36.5-38		33-34
67			17	35-36	23	31.5-32.5
66	27	49		33-34.5		30.5-31
65			16	32-32.5	22	29.5-30

Table 31
Standard Scores for the Paragraph and Triangle Figure
Subtest Delay Scores in Study 2

Age Groups						
Standard Scores	55-64 yrs.		65-74 yrs.		75-85 yrs.	
	Para.	Fig.	Para.	Fig.	Para.	Fig.
125						
124					21	
123						
122			21			
121						
120						
119						
118						
117	21					26
116						
115						
114						
113						25.5
112					20	
111				26		
110			20			
109						
108						25
107						
106		26				
105						24.5
104	20			25.5		
103					19	
102			19			24
101						
100						23.5
99						
98		25.5				23
97				25	18	
96			18			22-22.5
95	19					21.5
94						21
93					17	20.5

Table 31 cont'd

Standard Scores	55-64 yrs.		65-74 yrs.		75-85 yrs.	
	Para.	Fig.	Para.	Fig.	Para.	Fig.
92				24.5		
91			17			20
90	18			24		
89		25		23.5	16	
88			16	23		19.5
87	17	24.5		22.5	15	19
86				22		18.5
85			15	21.5		
84		24		21		18
83		23.5		20.5	14	
82			14	20		
81	16	23		19.5		17.5
80		22.5		19		
79		22			13	
78			13	18.5		
77	15	21.5			12	17
76		21		18		
75				17.5		
74		20.5	12	17		
73	14	20	11		11	16.5
72				16.5		
71		19.5	10	15.5-16		
70				14-15	10	
69		19		13-13.5		16
68		18.5		11.5-12.5	9	
67				10-11		
66	13	18		9-9.5		
65			9	8-8.5	8	

Table 32
Cumulative Percentiles for Adults Aged 55-64 Years on the
Paragraph Subtest Trial-by- Trial Scores in Study 2

55-64 years

Cumulative Percentiles	Trial 1	Trial 2	Delay 1	Delay 2
100	21	21	21	21
99				
98				
97				
96				
95				
94				
93	20			
92				
91				
90				
89				
88				
87				
86				
85				
84				
83				
82				
81				
80				
79				
78				
77				
76				
75			20	
74				
73				
72	19			
71				
70				
69				
68				

Table 32 cont'd

55-64 years

Cumulative Percentiles	Trial 1	Trial 2	Delay 1	Delay 2
67				
66				
65				
64				
63				
62				
61				
60				
59				
58				
57				
56	18			
55				
54				
53				
52				
51		20		
50				
49				
48				
47			19	
46				
45				
44				
43				
42				
41				
40				
39				
38				
37				
36				
35	17			
34				
33				

Table 32 cont'd

55-64 years

Cumulative Percentiles	Trial 1	Trial 2	Delay 1	Delay 2
32				
31				
30				
29				
28			18	
27				
26				
25	16			
24				
23				
22				
21		19	17	
20				
19				
18				
17				
16	15		16	
15				
14				
13				
12		18		20
11			15	
10				
9				
8				
7	14	17		
6	13			
5			14	
4	12	16		
3		15		
2	11	14	13	
1				

Table 33
Cumulative Percentiles for Adults Aged 65-74 Years on the
Paragraph Subtest Trial-by- Trial Scores in Study 2

65-74 years

Cumulative Percentiles	Trial 1	Trial 2	Delay 1	Delay 2
100	21	21	21	21
99				
98				
97				
96				
95				
94	20			
93				
92				
91				
90				
89				
88				
87				
86			20	
85				
84				
83	19			
82				
81				
80				
79				
78				
77				
76				
75				
74				
73	18			
72				
71				
70				
69				
68				

Table 33 cont'd

65-74 years

Cumulative Percentiles	Trial 1	Trial 2	Delay 1	Delay 2
67				
66		20		
65				
64			19	
63				
62				
61				
60				
59				
58	17			
57				
56				
55				
54				
53				
52				
51				
50				
49				
48				
47			18	
46				
45	16			
44				
43				
42		19		
41				
40				
39				
38				
37				
36	15			
35				
34				
33				

Table 33 cont'd

65-74 years				
Cumulative Percentiles	Trial 1	Trial 2	Delay 1	Delay 2
32				
31			17	
30				
29				
28				
27	14			
26				
25		18	16	20
24				
23				
22				
21				
20				
19	13	17	15	
18				
17				
16				
15				
14				
13	12		14	
12				
11	11			
10				
9	10	15-16	13	
8		14		
7	9			
6				19
5	8	13	12	
4			11	18
3	7	11-12	10	
2		9-10	9	17
1				

Table 34
Cumulative Percentiles for Adults Aged 75-85 Years on the
Paragraph Subtest Trial-by- Trial Scores in Study 2

75-85 years

Cumulative Percentiles	Trial 1	Trial 2	Delay 1	Delay 2
100	21	21	21	21
99				
98				
97				
96				
95				
94				
93				
92				
91				
90	20		20	
89				
88				
87				
86				
85				
84				
83				
82				
81				
80				
79				
78	19			
77				
76				
75				
74				
73	18			
72				
71				
70				
69				
68				

Table 34 cont'd

75-85 years				
Cumulative Percentiles	Trial 1	Trial 2	Delay 1	Delay 2
67				
66			19	
65				
64				
63		20		
62				
61	17			
60				
59				
58				
57				
56				
55				
54	16			
53				
52				
51				
50				
49				
48				
47			18	
46	15			
45				
44				
43				
42		19		
41				
40				
39	14		17	
38				
37				20
36		18		
35				
34				
33				

Table 34 cont'd

75-85 years

Cumulative Percentiles	Trial 1	Trial 2	Delay 1	Delay 2
32				
31				
30				
29				
28				
27			16	
26				
25				
24	13	17		
23				
22				
21				
20			15	
19				
18				
17			14	
16				
15	12	16		
14				
13				
12	11			19
11				
10				
9				
8			13	
7	10	15	12	
6				
5			11	18
4			10	
3	9	14	9	
2		13	8	
1				

Table 35
Cumulative Percentiles for Adults Aged 55-64 Years on the
Triangle Figure Subtest Trial-by-Trial Scores in Study 2

55-64 years

Cumulative Percentiles	Trial 1	Trial 2	Trial 3	Delay	Copy
100	26	26	26	26	26
99					
98	25.5				
97					
96					
95	25				
94					
93					
92					
91	24.5				
90					
89					
88					
87					
86	24				
85					
84					
83					
82					
81					
80		25.5			
79	23.5				
78					
77	23				
76					
75					
74					
73					
72	22.5				
71					
70	22				
69					
68					

Table 35 cont'd

55-64 years

Cumulative Percentiles	Trial 1	Trial 2	Trial 3	Delay	Copy
67					
66					
65				25.5	
64					
63			25.5		
62					
61					
60		25			
59					
58					
57					
56					
55					
54	21.5				
53	21				
52					25.5
51		24.5			
50					
49					
48	20.5				
47					
46					
45					
44					
43					
42	20	24			
41					
40					
39					
38					
37		23.5			
36					
35	19.5				
34					
33					

Table 35 cont'd

55-64 years

Cumulative Percentiles	Trial 1	Trial 2	Trial 3	Delay	Copy
32	19	23			
31					
30				25	
29					
28		22.5			
27					
26	18.5		25		
25	18				
24	17.5		24.5		
23	17	22		24.5	
22	16.5				
21	16	21.5	2.		
20					
19					
18	15.5	21			
17					
16		20.5		24	
15					
14	15			23.5	
13		20			
12			23.5	23	
11	14.5	19.5	23	22.5	
10		19		22	
9	13.5-14		22.5		
8	12.5-13	18.5	22	21.5	
7	12	18		21	
6	11.5		21.5	20.5	
5	11			20	
4		16.5-17.5	20.5-21	19.5	25
3		13.5-16	19.5-20	18.5-19	
2	10.5	12-13	19	18	
1					

Table 36
Cumulative Percentiles for Adults Aged 65-74 Years on the
Triangle Figure Subtest Trial-by-Trial Scores in Study 2

65-74 years

Cumulative Percentiles	Trial 1	Trial 2	Trial 3	Delay	Copy
100	25	26	26	26	26
99	24.5				
98					
97	24				
96					
95					
94	23.5				
93					
92					
91	23				
90					
89					
88	22.5				
87		25.5			
86	22				
85					
84					
83					
82					
81					
80	21.5				
79					
78	21		25.5		
77				25.5	
76					
75					
74					
73		25			
72					
71					
70					
69	20.5	24.5			
68					

Table 36 cont'd

65-74 years

Cumulative Percentiles	Trial 1	Trial 2	Trial 3	Delay	Copy
67					
66	20				
65					
64		24			
63					
62					
61	19.5				
60					
59					
58					
57					25.5
56	19				
55			25		
54					
53				25	
52	18.5				
51					
50					
49					
48					
47	18				
46					
45					
44					
43					
42					
41					
40					
39					
38	17.5				
37					
36	17	23.5			
35					
34		23	24.5		
33					

Table 36 cont'd

65-74 years

Cumulative Percentiles	Trial 1	Trial 2	Trial 3	Delay	Copy
32					
31		22.5	24	24.5	
30					
29					
28		22		24	
27	16.5		23.5		
26		21.5		23.5	
25					
24					
23		21		23	
22	16		23		
21					
20			22.5	22.5	
19				22	
18					
17	15.5	20.5	22	21.5	
16	15	20		21	
15				20.5	
14	13.5-14.5	19.5	21.5		
13	12-13	19		20	25
12		18.5		19.5	
11		18	21	19	
10		17.5	20.5	18.5	
9			19.5-20		
8	11.5	17	19	18	24.5
7	11	16.5		17.5	
6	10.5	16		17	
5	10	15.5	18-18.5		
4	9.5		17-17.5	16.5	
3	8.5-9	13-15	15-16.5	12-16	
2	8	11-12.5	13-14.5	8-11.5	24
1					

Table 37
Cumulative Percentiles for Adults Aged 75-85 Years on the
Triangle Figure Subtest Trial-by-Trial Scores in Study 2

75-85 years

Cumulative Percentiles	Trial 1	Trial 2	Trial 3	Delay	Copy
100	25	26	26	26	26
99					
98					
97					
96					
95	24.5				
94		25.5			
93	23.5-24				
92	23				
91	22.5				
90	22				
89					
88	21.5	25		25.5	
87			25.5		
86	21				
85					
84					
83		24.5			
82					
81	20.5				
80	20	24			
79					
78	19.5				
77					
76					
75	19			25	
74					
73	18.5	23.5	25		
72					
71	18				
70					
69					
68			24.5		

Table 37 cont'd

75-85 years

Cumulative Percentiles	Trial 1	Trial 2	Trial 3	Delay	Copy
67					
66		23			
65					
64			24	24.5	
63	17.5			24	
62					
61					
60		22.5			
59	17				
58					25.5
57			23.5	23.5	
56	16.5				
55					
54	16	22			
53	15.5				
52					
51	15	21.5			
50					
49		21	23	23	
48					
47					
46		20.5	22.5		
45					
44	14.5				
43					
42	14	20	22		
41		19.5		22.5	
40				22	
39	13.5	19		21.5	
38		18.5			
37		18		21	
36	13				
35					
34			21.5	20.5	
33					

Table 37 cont'd

75-85 years

Cumulative Percentiles	Trial 1	Trial 2	Trial 3	Delay	Copy
32		17.5	21	20	
31					
30					
29		17			
28			20.5		
27					
26					
25		16-16.5			
24		15.5	20	19.5	
23					
22	12.5			19	
21					
20	12	15			
19		14.5		18.5	
18	11.5				
17				18	
16					
15	11	14	19.5		25
14			19		
13	10.5				
12				17.5	
11			18.5		
10	10				24.5
9					
8	9-9.5		18	17	
7	8-8.5	13-13.5	17-17.5		
6		11.5-12.5	16.5	16.5	
5		10.5-11	15.5-16		24
4			14.5-15		23.5
3		10	13-14	16	
2	7.5	9.5	11.5-12.5		23
1					

Table 38
Standard Error of Measurement (SEM) for Paragraph and
Triangle Figure Subtests in Study 2

Subtests	Age Groups			
	55-64	65-74	75-85	ALL
Paragraph¹				
Trial 1	1.51	1.74	1.78	1.68
Trial 2	0.91	1.20	1.22	1.13
Tot. Acq.	1.77	2.12	2.16	2.03
Delay (FR)	1.24	1.49	1.50	1.42
Triangle²				
Trial 1	2.27	2.59	2.70	2.53
Trial 2	1.58	1.76	2.25	1.89
Trial 3	1.02	1.42	1.87	1.49
Tot. Acq.	2.99	3.42	3.99	3.51
Delay	0.98	1.42	1.89	1.49
Copy	0.19	0.39	0.49	0.39

¹ Maximum score on each trial of the Paragraph subtest is 21. The maximum score on total acquisition (Tot. Acq.) is 42.

² Maximum score on each trial of the Rectangle subtest is 26. The maximum score on total acquisition (Tot. Acq.) is 78.

Table 39
Percent Recalled on the Diagnostic Test Paragraph and
Geometric Figure by Final Three Age Groups

Subtests	Age Groups		
	55-64 yrs	65-74 yrs	75-85 yrs
Paragraph:			
Trial 1	72%	64%	55%
Trial 2	91%	88%	81%
Tot. Acq.	81%	76%	68%
Delay	85%	81%	71%
Rectangle:			
Trial 1	69%	62%	49%
Trial 3	93%	90%	81%
Tot. Acq.	83%	78%	66%
Delay	92%	89%	78%

Table 40
Percent Recalled on the Tracking Test Paragraph and
Geometric Figure by Final Three Age Groups

Subtests	Age Groups		
	55-64 yrs	65-74 yrs	75-85 yrs
Paragraph:			
Trial 1	85%	78%	76%
Trial 2	95%	91%	91%
Tot. Acq.	90%	85%	84%
Delay	91%	85%	84%
Triangle:			
Trial 1	78%	71%	62%
Trial 3	97%	94%	87%
Tot. Acq.	89%	84%	76%
Delay	97%	93%	87%

Appendix A

Scoring System for the Rectangle Geometric Figure

Instructions:

Consider each of the 13 components or details separately. Appraise the accuracy of each component and its position relative to the whole design. Maximum score = 26 points. Scoring is as follows:

Correct	→ Placed properly	→ 2 points
	→ Placed poorly	→ 1 point
Distorted or incomplete but recognizable	→ Placed properly	→ 1 point
	→ Placed poorly	→ 1/2 point
Absent or not recognizable	→	0 points

Scoring:

<u>Component</u>	
1. Large rectangle	
2. Horizontal midline	
3. Vertical midline	
4. Crosses	
5. Semicircle	
6. Square	
7. Arrow	
8. Dots	
9. Triangle	
10. Diagonal line	
11. Vertical lines	
12. Circle	
13. Short horizontal lines in circle	

Trials

1	2	3	D	C

TOTALS:

Appendix C

Brief Description of Previously Published Tests

Used in Study 1 and/or Study 2

1) WAIS-R Vocabulary Subtest

The WAIS-R Vocabulary subtest consists of a list of 35 words for which the individual being tested must give the meaning of or define (Wechsler, 1987). The Vocabulary subtest was selected because it is highly correlated with both WAIS-R Verbal IQ and Full-Scale IQ. Each item is scored 0, 1, or 2. The maximum score is 70.

2) Mini-Mental Status Examination (MMSE)

The MMSE was designed to quickly and easily test cognitive functioning (Folstein et al., 1975). The maximum score on the test is 30. The generally-accepted cut-off score of 23/24, meaning that scores equal to or less than 23 are suggestive of cognitive impairment, was used.

3) Geriatric Depression Scale (GDS)

The GDS is a 30-item questionnaire designed to screen for depression in the elderly (Yesavage et al., 1983). A yes/no response format is used. The maximum score on the test is 30. A cut-off score of 17/18 was used, meaning that individuals with scores equal to or greater than 18 are considered to be cases with significant depressive symptomatology. In addition, the GDS is retitled the "Life Satisfaction Scale" in the present studies in an attempt to avoid a social desirability bias in responding.

Appendix C cont'd

4) Trails A & B

Trails A and B measure an individual's visual scanning abilities, motor speed, and mental flexibility. Trails A requires the person to connect consecutively a series of circles numbered 1 to 25. In Trails B, the person must alternate between circles containing numbers and circles containing letters (e.g., joining 1 to A, A to 2, 2 to B, B to 3, and son on). The score is the time required to complete each task.

5) Subjective Age Questionnaire

There are two subjective age questionnaires used in Studies 1 and 2. Half of the subjects receive Barak's subjective age scale which consists of four questions asking the age a person feels, looks, acts, and with which they would associate their interests (Barak & Schiffman, 1981). The other half of the subjects receive Montepare and Lachman's (1989) subjective age scale which consists of four questions asking the age a person feels, looks, acts and would associate their interests with, and the age they would like to be.

6) Word Fluency

The Word Fluency test examines the speed and ease of verbal production, and is a sensitive measure of brain dysfunction (Lezak, 1995). In this test, the person is given 60 seconds to say as many words as they can think of

Appendix C cont'd

that begin with a particular letter, excluding the use of proper names and the use of the same word with different endings. Three letters (F, A, and S) and one category (Animals) are used to generate words.

7) Boston Naming Test

The Boston Naming Test (Kaplan, Goodglass, & Weintraub, 1983) measures confrontational naming. It consists of 60 line drawings of objects ranging in familiarity from "bed" to "seahorse" to "protractor" to "abacus" which the person is asked to name. When the person is unable to recognize the drawings or misperceives it, a stimulus cue is given (e.g., for "globe" - it is a kind of map). If the person still cannot give the name, a phonemic cue is given (e.g., for globe - it starts with the sound "glo").

Appendix D

Description of Unpublished Tests

Used in Study 1 and/or Study 2

1) Word List

A 15-item word list, similar to that given in the LAMB, was administered. There are up to five learning trials and a delayed recall trial. Both free recall and cued recall are assessed in each trial. Because each word represents a different category, this category is then used as a reminder during cued recall. The word list and the category cues are listed below:

<u>Word List</u>	<u>Category</u>
TIGER	Animal
GREEN	Colour
DESK	Furniture
HAND	Part of the Body
SWORD	Weapon
NURSE	Profession
SWEATER	Clothing
BASEMENT	Part of a Building
GUITAR	Musical Instrument
SPINACH	Vegetable
PIGEON	Bird
CANOE	Boat
SISTER	Relative
DAY	Unit of Time
TENNIS	Sport

After free and cued recall are administered on the delayed recall trial, a recognition task is given. The person is given a sheet containing a list of 45 words and is asked to circle the 15 words they had to learn and remember. Listed below are the 45 words presented to the person.

Appendix D cont'd

sweater	foot	river
drum	basement	tennis
coffee	mile	teacher
red	tiger	corn
spoon	baseball	hand
roof	guitar	oil
desk	pear	yacht
pigeon	book	hour
cabin	lion	canoe
chair	pepper	green
day	aunt	tulip
eagle	spinach	knife
sword	whiskey	moth
hammer	nurse	wind
wool	shirt	sister

2) Recognition Memory for Abstract Drawings

This test examines recognition memory for abstract drawings. Twenty-four abstract drawings are presented to the person in a list-like fashion every 3 seconds. Immediately afterward, a four-item multiple choice test is administered in which the person must select the abstract drawing they were shown from among three distractors (i.e., new drawings). There are two learning trials and a delayed recall trial. The score on each trial of the test is the total number of abstract drawings correctly recognized; the maximum score is 24.

3) Recognition Memory for Line Drawings

This test examines recognition memory for simple line drawings of common objects. Forty-eight drawings of different objects are presented to the person in a list-like fashion every 3 seconds. Immediately afterward, a four-item multiple choice test is administered in which the person

Appendix D cont'd

must select the drawing they were shown from among three distractors (i.e., new drawings). There are two learning trials and a delayed recall trial. The score on each trial of the test is the total number of drawings correctly recognized, with a maximum score of 48.

4) Recognition Memory for Common Words

This test examines recognition memory for common words. Forty-eight different words are presented to the person in a list-like fashion every 3 seconds. Immediately afterward, a four-item multiple choice test is administered in which the person must select the word they were shown from among three distractors (i.e., new words). There are two learning trials and a delayed recall trial. The score on each trial of the test is the total number of words correctly recognized, with a maximum score of 48.

5) Chronological Age Questionnaire

The Chronological Age Questionnaire consists of six questions designed to help interpret responses to the Subjective Age Questionnaire described in Appendix A. These questions ask how often people think about their age, how happy or satisfied they are with being their age, how often they feel their age, what age they feel physically, what age they feel mentally, and the age at which they were happiest.

6) Hubley Depression Scale

The Hubley Depression Scale is a 16-item

Appendix D cont'd

questionnaire designed to screen for depression in the elderly (Hubley, 1994). A yes/no response format is used. The maximum score on the test is 16. This scale is retitled the "CMAR Life Satisfaction Scale" in the present studies in an attempt to avoid a social desirability bias in responding.

7) Memory Change Questionnaire

A 36-item self-report questionnaire was developed in order to assess if an individual's memory has changed in the past year, and if so, how (Tombaugh, 1994). Sample items include:

1. Over the past 12 mos., my ability to:
Remember the names of people recently introduced to me is...
18. Over the past 12 mos., my ability to:
Remember the words I wish to use is...
36. Over the past 12 mos., my ability to:
Remember to turn off appliances or personal items, such as a stove or iron, when I am finished with them is...

A five-point Likert scale using the following responses is used: Much Worse, Worse, Same, Better, Much Better.

Appendix E

Ethics Approval

*Psychology 72-710**93-014*

DEPARTMENT OF PSYCHOLOGY PROPOSAL FOR RESEARCH WITH HUMAN SUBJECTS

Please submit two copies of this form and two copies of the supporting materials to the Department Committee on the Use of Humans in Research (B550 Loeb Building).

1. Date of this submission..... *September 10, 1993*
Faculty Sponsor..... *Dr. T.M. Timbault* Phone..... *728-2457*
Principal Investigator..... *Dr. Hukley* Phone..... *728-2682*
Other research personnel..... *Melissa Pivetti, Pat Cross, Rita Mayr, Kate Oakley, Helen*
Project title..... *A Study of Visual and Verbal Memory*
Type of research (e.g., Faculty, M.A. thesis, etc.)..... *Faculty / Independent*
Approximate starting and completion dates..... *May 25, 1993 - May 30, 1993*
Approximate length of testing session..... *90 min*
2. Number, age, and source (e.g., 49.100 students) of subjects. Note, if subjects are being recruited from an organization outside of Carleton University attach a copy of their permission and/or their Ethics Committee approval. *Approximately 500 subjects between 17 & 29 years of age... will be recruited from within the university as well as from the community*
3. Will subjects be paid or given course credit? *on remuneration but paper credit* (Indicate length of experimental session e.g., 1/2 hr.) *if in 49.100 (1 1/2 hr)*
4. Check list: Are the following included?

Description of Purpose	☒ YES	☐ NO	☐ N/A
Procedure (including materials)	☒ YES	☐ NO	☐ N/A
Informed Consent	☒ YES	☐ NO	☐ N/A
Written Debriefing	☒ YES	☐ NO	☐ N/A
Announcement for Recruiting (re: community)	☒ YES	☐ NO	☐ N/A
5. Does the study involve anything that might cause subject anxiety, pain, or embarrassment? *YES* ☒ NO
If yes, attach a description of the precautions taken to safeguard the subject's interest.
6. Does the study involve deception? If yes, attach a description of the nature of the deception and the steps that will be taken to protect, inform and debrief the subjects. *YES* ☒ NO

We (I) acknowledge that subjects will be treated in accordance with the ethical guidelines of the Canadian Psychological Association.

[Signature]

Principal Investigator

[Signature]

Project/Course Supervisor

The Department Ethics Committee has..... *approved*Date: *13/9/93*Committee Chairman *[Signature]*

Appendix F

Study 1: Order of Test Administration

Informed Consent
Mental Status Test (MMSE)

Word List (acquisition)
Abstract Recognition Memory (acquisition)

Subjective Age Questionnaire
Chronological Age Questionnaire
Geriatric Depression Scale
Trails A & B

Word List (delay)
Abstract Recognition Memory (delay)

G-LAMB Diagnostic Paragraph (acquisition)
G-LAMB Diagnostic Geometric Figure (acquisition)

Word Fluency
Hubley Depression Scale
WAIS-R Vocabulary

G-LAMB Diagnostic Paragraph (delay)
G-LAMB Diagnostic Geometric Figure (delay)

Boston Naming Test
Memory Change Questionnaire

Appendix G

Study 2: Order of Test Administration

Informed Consent
Mental Status Test (MMSE)
Subjective Age Questionnaire

G-LAMB Tracking Paragraph (acquisition)
G-LAMB Tracking Geometric Figure (acquisition)
Picture Recognition Memory (acquisition)

G-LAMB Tracking Paragraph (delay)
G-LAMB Tracking Geometric Figure (delay, copy)
Picture Recognition Memory (delay)

Geriatric Depression Scale
Chronological Age Questionnaire

Word Recognition Memory (acquisition)
WAIS-R Vocabulary
Word Recognition Memory (delay)

Appendix H

Learning and Memory Battery (G-LAMB):

Paragraph and Complex Figure

Tom Wellington, aged 63, is an electrician at Brighton University. He lives in a gray 8 room house on Beach Avenue. Every day at 7:45 A.M. he takes his white, six-speed motorcycle from under the porch. Then he drives over the Walnut Street Bridge to the university cafeteria for his morning muffin and cup of tea. On Tuesdays, he leaves his second floor office at 1:30 to visit Ted's Restaurant. Here he has a glass of red wine, plays checkers, and reads the Northern Vista magazine.

END

3 1 - 0 5 - 9 | 6

FIN