

A SYSTEMATIC REVIEW OF ASSESSMENT PROTOCOLS FOR THE
DISCRIMINATION BETWEEN MILD COGNITIVE IMPAIRMENT
AND NORMAL COGNITIVE ABILITY
IN THE AGING POPULATION

Except where reference is made to the works of others, the work described in this thesis is my own work or was done in collaboration with my advisory committee. This thesis does not include proprietary or classified information.

Jessica Suzanne Lindsay

Certificate of Approval:

Michael J. Moran
Professor
Communication Disorders

Nancy J. Haak , Chair
Assistant Professor
Communication Disorders

William O. Haynes
Professor
Communication Disorders

Joe F. Pittman
Interim Dean
Graduate School

A SYSTEMATIC REVIEW OF ASSESSMENT PROTOCOLS FOR THE
DISCRIMINATION BETWEEN MILD COGNITIVE IMPAIRMENT
AND NORMAL COGNITIVE ABILITY
IN THE AGING POPULATION

Jessica Suzanne Lindsay

A Thesis

Submitted to

the Graduate Faculty of

Auburn University

in Partial Fulfillment of the

Requirements for the

Degree of

Master of Science

Auburn, Alabama
May 10, 2008

A SYSTEMATIC REVIEW OF ASSESSMENT PROTOCOLS FOR THE
DISCRIMINATION BETWEEN MILD COGNITIVE IMPAIRMENT
AND NORMAL COGNITIVE ABILITY
IN THE AGING POPULATION

Jessica Suzanne Lindsay

Permission is granted to Auburn University to make copies of this thesis at its discretion, upon requests of individuals or institutions and at their expense. The author reserves all publication rights.

Signature of Author

Date of Graduation

THESIS ABSTRACT

A SYSTEMATIC REVIEW OF ASSESSMENT PROTOCOLS FOR THE
DISCRIMINATION BETWEEN MILD COGNITIVE IMPAIRMENT
AND NORMAL COGNITIVE ABILITY
IN THE AGING POPULATION

Jessica Suzanne Lindsay

Master of Science, May 10, 2008
(B.A., University of South Florida, 2006)

154 Typed Pages

Directed by Nancy J. Haak

Mild cognitive impairment (MCI) is a concept that an aging individual has reached a point of cognitive decline that is not yet severe enough to be termed *dementia*, but is not considered normal. Subtle declines in cognitive ability are many times difficult to discriminate from normal ability. This thesis took a systematic approach to reviewing the literature to obtain all available research on the various assessment protocols available to discriminate between MCI and normal cognitive ability. This systematic review was accomplished by searching the following databases: MEDLINE, PsycINFO, ERIC, CINAHL and DISSERTATION ABSTRACTS, putting the gather articles through multiple levels of rigorous sorting for of the study, analyzing and synthesizing the data and rating the methodological design of each study. Conclusions drawn to the scope of

practice of a speech-language pathologist yielded a multitude of appropriate diagnostic protocols with adequate support from the research. Specific screening protocols and short, focused formal tests were found to be appropriate for the use of a speech-language pathologist, although these tests may be more appropriate to determine who needs further evaluation and who does not. The most rigorous, powerful research was found in formal, more exhaustive test batteries that may require a referral to a neuropsychologist. Overall, speech-language pathologists have an active role in the early identification of MCI and should be aware of the available tests and how sensitive they are to detect the subtle changes associated with mild cognitive impairment.

ACKNOWLEDGMENTS

I would like to thank Dr. Nancy J. Haak for all of her guidance and support throughout this process. Also, many thanks to my advisory committee, Dr. Michael Moran and Dr. William Haynes for the time spent in designing and carrying out the study. To Leigh Reynolds, I would have never gotten through this without your everyday support and understanding of the process. Thank you for your time spent helping with the quality assessment of the articles. I would like to thank my family, Mom, Dad and Whitney for their encouragement and prayers as I have taken on this endeavor. Finally, thank you to my fiancé, Thomas, for the ongoing support throughout my education. All of my thanks go to God for enabling me to accomplish this and giving me the support I need on a daily basis.

Style manual or journal used American Psychological Association Publication
Guide, 5th Edition

Computer software used Microsoft Word 2003 and Microsoft Excel
2003

TABLE OF CONTENTS

LIST OF TABLES	x
CHAPTER 1. INTRODUCTION	1
CHAPTER 2. LITERATURE REVIEW	4
Normal cognitive and communication skills in aging adults	5
Abnormal cognitive and communication skills in aging adults	9
Dementia	9
Mild cognitive impairment	11
Evidence-based practice	15
Systematic Review	16
Justification for research	22
CHAPTER 3. METHODOLOGY	25
Resources used	25
Search strategies	27
Inclusion and exclusion criteria	29
Evaluation of the evidence	29
Quality assessment	29
Data extraction and synthesis	32
CHAPTER 4. RESULTS	34
Method of inclusion	3

Review of articles	37
Formal test batteries.....	38
Screening tools.....	56
Computer-based assessments.....	67
Equipment-based measures.....	72
Miscellaneous assessment tools.....	74
Summary tables.....	86
CHAPTER 5. DISCUSSION.....	105
Patterns in quality assessment.....	105
Patterns in types of tests.....	111
Evidence-based recommendations.....	115
Speech-language pathology scope of practice	117
Directions for future research	119
REFERENCES	121
APPENDIX A.....	133
APPENDIX B.....	136

LIST OF TABLES

Table 1. Coding for assessment of study (SIGN, 2001)	21
Table 2. Hierarchy of study type (SIGN, 2001).....	21
Table 3. Overall assessment of article, adapted from SIGN, 2001	31
Table 4. The sorting process according to each database	35
Table 5. The breakdown of number of articles not meeting criteria for inclusion at the full-text level of sorting.....	36
Table 6. Tally of number of articles in each diagnostic test group.....	38
Table 7. Test in question, reference standard, inclusion and exclusion criteria	88
Table 8. Study and participant information	95
Table 9. Summary of articles reporting parametric statistics sorted by quality assessment.....	99
Table 10. Summary of articles reporting nonparametric statistics sorted by quality assessment.....	101
Table 11. Sensitivity, specificity and area under the curve measures for studies sorted by quality assessment score	102

CHAPTER 1

INTRODUCTION

As the baby-boomer generation steps into the “aging population,” there is a growing need to understand both normal and abnormal cognitive decline so that appropriate measures can be taken to serve this age group. With the interrelation of cognitive ability and language (American Speech-Language Hearing Association, ASHA, 2005), it is important that professionals in the area of speech-language pathology be able to distinguish between normal and abnormal decline of abilities in aging and also the transition stage between normal and abnormal. The term *mild cognitive impairment* has become a term of interest in recent years as a decline in cognitive ability, many times memory, with age that does not meet the criteria for the more severe type of progressive cognitive decline of dementia (Petersen, et al., 2001). There is a growing awareness of MCI in the aging population that is progressively gaining more attention due to the importance of early identification of cognitive decline such as dementia. Most research has focused on discriminating between normal cognitive decline and dementia due to the increase in awareness and need for early identification. The task becomes more difficult when trying to discriminate between the transition period of MCI; when the mild forgetfulness is not considered “normal” any longer. And as stated by Chapman, Ulatowska, King, Johnson, and McIntire (1995), the line between normal and abnormal when discussing communication ability can easily become blurred.

As stated previously, communication and cognitive ability are related, and although changes in these areas may be subtle in the beginning stages of cognitive decline, it is vital that health professionals can distinguish normal from abnormal. Identifying the individuals who have crossed the line into being mild cognitively impaired is also important to consider in terms of early intervention and prevention since it has been reported that individuals with MCI are at a higher risk of progressing to Alzheimer's disease at a rate of 10-12% per year that is contrasted with the 1-2% conversion rate to AD of normal aging adults (Petersen, et al., 1999). Determining who has MCI while the individual is exhibiting only mild cognitive deficits may make it easier to work in conjunction with the individual and family while the individual still retains the awareness of his/her own deficits (Lu, Haase, & Farran, 2007). There are various ways to assess cognitive ability in the elderly, and studies have shown that many professional disciplines provide assessment procedures to distinguish MCI from normal cognitive decline, ranging from medical, psychological, and many other procedures that can be used across health related fields (Galluzzi, Cimaschi, Ferrucci, & Frisoni, 2001; Molloy, Standish, & Lewis, 2005; Jefferson, et al., 2006; Zhang, et al., 2007). Since there are many different assessment tools in various professional fields to distinguish MCI from normal cognition, knowing what is available and the evidence supporting each would be important for all fields involved to understand, especially in an age of evidence-based practice.

A systematic approach should be taken to distinguish what is normal aging versus abnormal in the early stages (MCI) in order to bridge the gap of the gray area between normal cognitive deterioration and dementia. This study was designed as a systematic

review (SR), using a cross-disciplinary approach to determine the available types of assessment tools and the research that supports the tools distinguishing between MCI and normal cognitive decline. A careful and thorough SR is thought to be a helpful resource for a professional because the information is in one paper and the need to carry out additional searches is reduced (Griffer, Hargrove, & Lund, 2005). Knowing what assessments are available and supported by research will help health professionals know what to look for in this population and the procedures they can use in their specific field to assess the individual.

CHAPTER 2

LITERATURE REVIEW

There is a rapid growth in the aging population since the baby-boomer generation is large in number and is gradually entering the “senior adult” age range. According to the Alzheimer’s Association (2007), 78 million people in the baby-boomer generation are growing closer to the senior age range, with the oldest baby-boomers turning 60 last year. What is considered “senior” may vary according to individual lifestyles, but overall, the age of 65 is a landmark for many individuals as a change in lifestyle with the advent of retirement and the use of government programs such as Social Security and Medicare. As this population ages, for some there may be a steady decline in certain abilities, which in turn requires healthcare professionals to understand and recognize the symptoms stemming from this decrease in abilities. Many abilities of the aging population may be affected including the sensory, special sensory, motor, cognitive, and language systems (Civil & Whitehouse, 1991). The function and development of the language and cognition systems are correlated and a disruption in one system can affect the other and vice versa (ASHA, 2005a). For the purposes of this study the focus will be on the interrelated relationship between the cognition and language systems and the parameters that distinguish normal aging from aging with mild cognitive decline. According to Shadden (1988), communication involves sending and receiving information between at least two people using verbal and/or nonverbal skills for a particular purpose. Healthcare

professionals specializing in communication, speech-language pathologists need both an understanding of communication disorders and typical communication in the aging population to effectively manage problems with communication (Shewan & Henderson, 1988).

Normal Cognitive and Communication Skills in Aging Adults

Many studies have been conducted in order to determine what normal communication skills of the elderly are, but there is great variability in the findings. To begin, communication can be defined as the act of sharing information through a shared symbol system, be it a linguistic or nonlinguistic system (Bayles & Kaszniak, 1987; Shadden, 1988). Due to the individual variability of age-related changes including the person's health, cognitive skills, social, and economic status, plus numerous other factors, defining normal communication of the elderly is a difficult undertaking (Shadden, 1997). Au and Bowles (1991) discussed the interrelated system of communication and cognition of a normal aging adult in terms of the influences of the cognitive domain of memory on discourse, defined as "the ability to produce meaningful messages orally or in writing" (p. 296). The authors concluded that different types of memory deficits in normal aging adults including both episodic and semantic memory deficits may impact an elderly individual's ability to process and produce language. This chapter defines as *episodic memory* as information being stored that the individual has experienced, either long or short term. Working memory, which requires a manipulation of the information being stored for further processing, is included as a function of episodic memory. *Semantic memory* on the other hand is defined by Au and Bowles (1991), is "storage of knowledge about one's language" (p. 293). Both types of memory are involved in the linguistic

process from comprehending and storing the message long enough to respond properly in conversation or a number of different other tasks. In summary, the authors found that the memory decline in older adults may affect language ability in comprehension from working memory deficits and also vague, common words used in discourse due to word retrieval deficits involved in semantic memory decline.

Another source that suggests that cognitive decline, specifically memory decline, in the normal aging adult affects the communication skills of an individual is Benjamin's (1988) literature review on how memory and cognitive decline in aging adults affects the semantic component of language. The literature suggests that the aforementioned decline affects word retrieval skills in various tasks. Benjamin then reviews the literature focusing on the effects memory and cognitive decline in aging adults (in many studies reviewed, adults ranging from 60-65 years-old) have on the pragmatic component of language. This review by Benjamin suggests that although normal aging adults can communicate functionally in conversation, some tend to use a scattered, less organized structure of discourse and have difficulty recalling past discourse topics (Benjamin, 1988). The results from this study show that even minimal cognitive decline such as memory can impact language and communication. Benjamin, among others (Bayles & Kaszniak, 1987; Shadden, 1988), completed nearly 20 years ago raises an interesting question now that there is a growing awareness of early cognitive decline and the transition to dementia: could what was considered "normal aging" language and cognition in the past now be what is considered mild cognitive impairment (MCI) and possibly an early transitional state to dementia?

Another study of communication in the aging normal population was conducted by Shewan and Henderson (1988) that analyzed language samples from 60 normal, English-speaking adults ranging in ages from 40-70 with the *Shewan Spontaneous Language Analysis* (SSLA) system in order to outline parameter of normal aging communication skills to use as comparison to neurologically or cognitively impaired individuals (Shewan & Henderson, 1988). This analysis included a number of variables of expressive language such as time, rate, number of utterances, paraphasias, and Communication Efficiency. The subjects were broken up into four age groups, but the findings showed few significant differences between the variables on the SSLA between age groups therefore concluding that communication skills are not greatly affected by aging. The variables that did yield significant effects from aging in this study were paraphasias and Communication Efficiency. This particular study involving the *SSLA* defined paraphasias as “percent of substitutions related to the number of utterances” and Communication Efficiency as “rate of communicating information (content units/time)” (p.142). The paraphasia variable had statistically significant increase in the three older age groups, concluding that the age of 50 may be where decline in communication begins. Communication Efficiency was also reduced in the older age groups in the study, revealing that some older individuals require more time to transmit a message. According to the authors of the study, this may be attributed to the individual having to use more words to express a message than a younger counterpart. The results from this suggested that aging adults have minimal deficits in communication, but when neurological deficits are present, deficits may become much more substantial (Shewan & Henderson, 1988).

Other sources such as Obler, Au, and Albert (1995) reported that healthy aging adults maintain their overall ability to use conversational discourse, but may tend to have a decline in naming, receptive language skills, and discourse recall, which was also found to be a factor affected by aging in a previous mentioned review of cognition and communication by Benjamin (1988). Another review by Shadden (1997) on normal communication skills of aging adults focuses specifically on how discourse behaviors are affected by aging and cognitive decline on discourse behaviors in adults in order to have a basis of what is considered “normal.” The results of the review yielded that normal aging adults produce more verbal disruptions (i.e. disfluencies, uncertainty behaviors, etc.), employ a reduced grammatical complexity and length, and use of vague references that yield less information than younger counterparts (Shadden, 1997). This review notes the difficulty in determining a standard for typical communication in a heterogeneous population of aging adults, but these results found to be a pattern shown throughout many studies including the underlying factor that cognitive decline does affect communication skills in the senior population.

The previous reviewed studies yield various results on how the interrelated communication and cognitive systems are affected by aging, but in summary, normal aging adults tend to have functional communication skills but may have subtle deficits in word-retrieval, discourse recall, and ability to produce concise, direct message (i.e. producing a more ambiguous message). It is noted that the majority of these studies and reviews of the literature on normal communication skills of the aging population were done 10-20 years ago, and after an extensive search of the literature, these references were the most recent found. This suggests that most of the research done on normal

communication was done in the 1970-1990 range, and in recent years, the research has shifted its focus to distinguish what is abnormal communication and cognitive ability in the aging population. It is also pertinent to note that since the most recent citation in this section is 1997, it suggests that what was once considered “normal aging” language and cognitive skills may now be what is labeled “mild cognitive impairment” (to be discussed in detail in the next section), since the awareness of early identification of cognitive decline such as dementia has had an increase in awareness in recent years.

Abnormal Cognitive and Communication Skills in Aging Adults

Dementia

When distinguishing what is normal from abnormal, the most distinct difference results from a contrast of the two opposite ends of the spectrum. The task becomes much more difficult and blurred when contrasting a milder form of “abnormal” from the norm, so what is known as most abnormal cognitive decline, dementia, will be discussed first.

Dementia is a progressive degenerative disorder that causes a decline in various areas such as intellect, personality, and communication skills (Bayles & Kaszniak, 1987).

According to the American Psychiatric Association (1997, p.14), dementia is “a syndrome comprised of multiple cognitive deficits including memory and at least one other area such as aphasia, apraxia, agnosia, or disturbance in executive function.”

Dementia can be a symptom of many types of disorders such as Alzheimer’s disease (AD), Parkinson’s disease, Huntington’s chorea, or drug toxicity, to name a few (Ripich, 1991), but is frequently discussed in reference as a symptom of Alzheimer’s Disease since 50-70% reported cases of dementia are due to AD (Alzheimer’s Association, 2007).

According to the Alzheimer’s Association (2007), to be classified as individual with

dementia, one must exhibit a decline in two of the following cognitive areas: 1) memory, 2) comprehension of written language and production of understandable speech, 3) ability to plan and make judgments, or 4) processing and comprehension of visual stimuli, and the declines must be enough that would affect the individual's everyday lifestyle.

Dementia is a degenerative disorder, meaning that early symptoms are mild and progressively worsen. Deficits in communication over the entire course of dementia include areas such as word-finding, comprehension of abstract language, attention, pragmatic skills such as turn-taking and topic maintenance, and use of vague, confused language (Bourgeois, 1992; Bourgeois, 2002). Early stage dementia begins with mild deficits in areas such as mild forgetfulness, trouble with attention, and mild disorientation (Bayles & Kasniak, 1987). According to Ripich (1991), the use of vague, nonspecific speech is a symptom of middle-stage dementia of the Alzheimer's type (DAT) and difficulty maintaining a conversation occurs once the disorder has progressed to the later stages. Chenoweth and Spencer's study (as cited in Bayles & Kaszniak, 1987) showed that many families have difficulty determining when they first noticed symptoms of dementia, and when asked to look retrospectively after the individual progressed to a later stage of dementia, the year prior to any specific symptoms being identified was described as being difficult due to misunderstanding between the individual suffering from dementia and family members. According to the research, the early stages of dementia are mild in symptoms, and fairly free from any observable communication deficit, making early identification difficult.

There are various rating scales available that show the progression to dementia from normal cognition, and one rating scale that is helpful in bringing this progression into

perspective is the *Global Deterioration Scale (GDS)* (Reisberg, Ferris, De Leon, & Crook, 1982). This is a 7-point rating scale starting with 1 as no cognitive decline, 2 being only mild forgetfulness that is still not considered abnormal for an elderly individual. Stages 3 and 4 are labeled the “confusional” stages, in which a person would fall into the category of MCI, meaning an individual is not cognitively normal, but the subtle changes in cognitive ability are not severe enough to be dementia. At these stages, the individual may experience a decrease in performance in work or social situations, but can still carry out activities of daily living (ADL). Stages 5-7 are the levels of dementia in which the individual can no longer carry out ADL without some level of assistance and have severe memory deficits. This rating scale helps to illustrate the many stages between normal cognition and dementia and that each stage may have very subtle differences.

Mild cognitive impairment

Normal cognitive and communication skills in the aging population were previously discussed and although memory decline and related areas of language may have a slight decline, the boundary line of normal versus impaired communication skills can become quite blurred (Chapman, Ulatowska, King, Johnson, & McIntire, 1995). Even in the beginning stages of a severe disorder causing cognitive decline such as dementia, it may be difficult to observe the subtle deficits in cognition, but it may be an even more difficult task to distinguish if the decline is considered a normal aspect of aging or is a symptom of a disorder. A disorder gaining more attention in recent years that causes a decline in cognitive ability, most often memory, as an individual ages but does not meet the criteria for the more severe type of progressive cognitive decline of dementia is termed *mild cognitive impairment (MCI)* (Petersen, et al., 2001). According to some

sources, daily living abilities are not affected in MCI, but as some of these individuals progress to dementia, activities of daily living will then become affected (Galluzi, Cimashi, Ferruchi, & Frisoni, 2001; Petersen, et al., 2001).

It has been reported that individuals with MCI are at a higher risk of progressing to AD at a rate of 10-12% per year. This is a substantial contrast to the 1-2% conversion rate to AD of normal aging adults (Petersen, et al, 1999). A longitudinal study completed by the Mayo Alzheimer's Disease Research Center found that the conversion rate of individuals with MCI to AD increased to 80% over a 6 year period (Petersen, 2001). Individuals with MCI have also been found to perform similarly to normal aging adults on general cognitive tasks (i.e. IQ testing), but then performed similarly to individuals with AD on tasks involving memory (Petersen, et al., 1999). Individuals with MCI have been contrasted with people with AD in that the MCI group has an increased amount of awareness to their cognitive decline(s), such as memory, and have the ability to reason to comprehend the deficits, which is an ability and awareness that individuals with AD are lacking (Lu, Haase, & Farran, 2007). This study concluded that individuals with MCI are able to verbally describe their strengths and weaknesses and are aware that their cognitive deficits may fluctuate on a daily basis and can be unpredictable. The retention of awareness of these deficits is a stark contrast in a more severe form of dementia such as dementia of the Alzheimer's type.

Having a cognitive decline that is not indicative of dementia but is not considered normal can be a vague description of MCI, which is a reason why MCI is often used as a miscellaneous category for what is not dementia (Luis, Loewenstein, Acevedo, Barker, & Duara, 2003). The Mayo Clinic's diagnostic criteria is widely used and consists of the

following criteria: one must exhibit a decline in memory, but not be lacking in other areas of cognitive ability or be impaired in activities of daily living (i.e. the individual maintaining his/her social, professional, and familial responsibilities and roles) (Petersen, et al., 2001; Riberio, Guerreiro, & DeMendonca, 2007). One study done by Riberio, et al. (2007) researched the memory deficits in MCI by focusing on semantic clustering and verbal learning strategies, both which reflect the interrelationship of cognition and language. The term semantic clustering or semantic strategies refers to the ability to use a categorization strategy to aid in a task of memory or recall, and the study discovered that the strategy of semantic clustering was impaired in contrast to normal control subjects, suggesting that this may also contribute to the memory deficits exhibited in individuals with MCI. Another interesting finding from this study concluded that individuals with MCI could benefit from learning strategies including verbal semantic cueing during the word recall tasks. A study by Riberio, Guerreiro, & de Mendonca (as cited in Ribeiro, et al, 2007) showed that other cognitive areas other than memory may be affected in MCI including complex language skills, executive function, semantic fluency, ability to calculate, and initiation of motor skills. These results demonstrate the importance of using language ability such as semantic ability to aid SLP's in determining what is normal aging cognitive ability versus a mild decline in cognitive ability (MCI) and then what has transitioned to a significant cognitive decline (dementia).

Much research has been completed on how to define MCI according to signs and symptoms, such as memory and other cognitive domains, but little to no research has shown that communication skills are affected by MCI (Riberio, et al., 2006; Luis, et al., 2003; Petersen, et al., 2001; Petersen, et al., 1999). Through their extensive literature

review of other types of cognitive decline such as dementia, Bayles and Kaszniak (1987) have shown that cognition and communication are interrelated in function. It can then be presumed that the mild cognitive decline associated with MCI could also cause subtle deficits in communication ability that go unseen. In a study on how individuals with MCI perceive their decline in abilities, Lu, et al. (2007) reported that all participants agreed that they were uncertain if the decline in memory due to normal aging and when they did acknowledge its presence in their lives, they attributed it to causes other than a progressive neurological disease such as MCI and considered their loss manageable and repairable. This suggests that individuals suffering from MCI may not be very forthcoming in their deficits, therefore becoming less accessible to identify other areas of decline such as communication skills.

There is a growing awareness of MCI in the aging population due to the importance of early identification of cognitive decline such as dementia. There are various ways to assess cognitive ability in the elderly, and studies have shown that areas from many professional disciplines provide assessment procedures to distinguish MCI from normal cognitive decline, including medical, psychological, and many other procedures that can be used across health related fields (Galluzzi, et al., 2001; Molloy, et al., 2005; Jefferson, et al., 2006; Zhang, et. al, 2007). Early intervention for individuals with MCI in what some define as the transitional stage between normal cognition and dementia may improve the management of the disorder and help with future planning for living and dealing with the deficits, possibly saving billions of dollars in healthcare (Luis, et al., 2003). It is also important to consider early intervention while the individual is exhibiting only mild cognitive deficits with MCI because it is easier to work in

conjunction with the individual and family while the individual still retains the awareness of his/her own deficits (Lu, et al., 2007). In summary, it is imperative that there be a systematic approach to distinguish normal from abnormal cognitive ability in the aging in order to bridge the gap of the gray area between normal and dementia.

Evidence-Based Practice

In a time of rapidly changing and expanding healthcare systems, it can be a challenging and time consuming task for many healthcare professionals to maintain a fully-informed and knowledgeable basis of current research (Cox, 2005). In many healthcare fields, questions arise as to the most effective way to evaluate and treat a new case which may be answered based on the health provider's prior clinical experience, personal preference or opinions of colleagues. The weakness in this approach to clinical decision making is that it is many times unsystematic and based solely on the experience or opinion of a few clinicians. In recent years, evidence-based practice (EBP) has become a widely used systematic approach to the clinical decision-making process in many of the various healthcare professions.

The healthcare field of speech-language pathology has also adopted EBP as a systematic way in determining the most effective evaluation and treatment approach of communication disorders. EBP is useful in a clinical setting because it allows the clinician's experience of practice, the client's value, and the best, firm evidence and research to come together to make the most effective clinical decision (Sackett, Straus, Richardson, & Haynes, 2000). Five steps are involved in order to apply EBP to a clinical case, (a) beginning with formulating a specific clinical question that can be answered, (b) searching the best accessible evidence (research), (c) evaluating the evidence in regards

to its validity and relevance, (d) making a recommendation from the evidence along with clinical experience and the client variables, (e) and evaluating the results and finding areas for improvement (Cox, 2005). When specifically focusing on reviewing the current research and evidence (step b) and evaluating the evidence (step c), it can be noted that these steps can become quite in-depth and time consuming. An exhaustive review of the literature would show what is strongly supported by research and what is lacking in evidence so that consumers can make informed clinical decisions and then subsequent research can follow-up on the areas lacking evidence.

Systematic review

A systematic review (SR) is a controlled, methodological identification, evaluation and qualitative analysis of the literature on a specific topic or clinical question based on a process that is designed before the evidence search and analysis ever begin (Hargrove, et al. 2005). SRs have become quite popular in other professions such as medicine and other health-related fields, but have yet to be widely found in the communication disorders literature (Hargrove, et al., 2005). With all healthcare fields such as speech-language pathology adopting an EBP approach to clinical decision making, SRs will soon become more common as a quick reference of a systematic evaluation and analysis of the literature on a particular clinical topic. A careful and thorough SR is thought to be a valuable asset for a professional because the information is in one paper and the need to carry out additional searches is reduced (Griffer, et al., 2005). Due to the controlled, rule-governed nature of a SR, the results should be reliable and less subject to bias than a review article not using any methodology to the review (Baylor & Yorkston, n.d).

As stated, a SR is a methodological, step-by-step process of reviewing the literature and synthesizing the evidence. The steps to this process according to Kitchenham (2004) are as follows:

1. Defining the question
2. Developing a protocol or method to answer the question
 - a. Resources that will be used
 - b. Strategy to obtain primary articles
 - c. Inclusion and exclusion criteria
 - d. Reliability of application of inclusion/exclusion criteria on articles
(i.e. number of reviewers, resolution of disagreements)
 - e. Quality assessment procedures
 - f. Data extraction procedure

The first step in the SR process is to develop a well-defined question (Pai, 2002). To ensure that the question being asked is clearly focused, it is helpful to use the PICO template to frame the question (ASHA, n.d.). The first letter in this acronym stands for *patient, population, or problem* and would require the question to state the characteristics of the population or the problem being studied, such as a disorder that is being addressed. The letter *I* stands for the type of intervention or exposure is being studied which can include carrying out types of therapy, assessment, or using only observation. The third letter stands for *comparison or control*. This would include the alternative to the intervention such as using a placebo, no intervention, or a different type of intervention. The final aspect of a well-defined question according to PICO is the *outcome* the researcher is interested in discovering such as the intervention showing higher/lower

results than the comparison. All relevant outcomes should be identified in the study and be of importance to professionals (Kitchenham, 2004).

The next step is to develop a protocol or method of review, beginning with determining the resources that will be employed in the SR. The task of gathering all available research on identification of mild cognitive impairment from normally aging cognitive and communication skills then requires the reviewer to delve into the literature of various disciplines in addition to speech-language pathology such as nursing, medicine and psychology. Shadden (1994) reported that a factor that speech-language pathologists need to address when working with the aging population is learning not to be insular in an area (aging) that requires involvement from various disciplines. In order to conduct a reliable systematic review, a rigorous review of literature is needed to gather the available evidence on the topic (Griffer, et al., 2005). This required the researcher(s) seeking professional consultation on internet databases to determine which were appropriate for use in the SR. It is also important at this stage in the SR to perform trial searches of the relevant literature to determine what disciplines/professions are instrumental in the topic to then determine the databases that will be necessary for a rigorous, interdisciplinary review (Kitchenham, 2004). A SR also must take into consideration publication bias, meaning that most often the published literature only reports positive results. To get a true picture of all available evidence on a topic, unpublished material should be included in the resources used by scanning the “grey literature” and conference proceedings, contacting experts in the field for information on unpublished material and hand searching of key journals (Pai, n.d.).

The next step in the process of conducting the SR will be to retrieve the articles that are to be analyzed. According to the Scottish Intercollegiate Guideline Network (SIGN) (2001), it is important to sift through the articles through a series of steps in order to remove the articles for analysis that are irrelevant to the topic. To minimize bias and ensure inter-judge reliability of the elimination of articles, at least 2 researchers are required to independently sift through the articles and then resolve disagreements through the decision of a third judge (SIGN, 2001). It has been found throughout the literature that most commonly, the initial sift is done through the titles and/or abstracts of the articles (Kitchenham, 2004; SIGN, 2001). Full-text articles are then obtained from the included abstracts for the final sift and are analyzed by the two reviewers for relevance of the study. This analysis and sifting process is accomplished through applying inclusion and exclusion criteria to the articles in question for each level (through abstracts and then full-text articles). Criteria for inclusion and exclusion in the study should be determined from the well-defined question and trial searches done in the beginning of the study. These criteria should reflect the information that is to be obtained in the search and should be piloted before the final search to ensure that they reliably include relevant articles and exclude irrelevant articles (Kitchenham, 2004).

Once the primary articles have been selected through the sifting process, study quality is assessed through an investigation of the methodology such as how well the study minimizes bias and maximizes validity (Kitchenham, 2004). This includes taking into consideration sample size, use of appropriate statistical analysis and use of valid outcome measure. It is also important to consider internal and external validity of a study when assessing the quality of an article and the results that the article yields. There is not

one accepted method of quality assessment, but SIGN has developed a method of assessing quality of a study that has been adopted by many proponents of evidence-based practice and is used frequently in SRs (Ricci, Celani, & Righetti, 2006; Chisolm, et al., 2007; ASHA, n.d.). According to Waugh (1999), the original SIGN system was a useful way to grade evidence based on only the type of study (RCT vs. cohort study), but the development of a more complex assessment instrument to look at the quality of the study was needed to integrate all available evidence. Although a complex grading system increases subjectivity in the assessment, the author of this study insisted that it was necessary to get a true picture of the quality of evidence. In 2001, the SIGN review group published the updated grading system that included the recommendations from Waugh to develop a formal assessment of methodology of a study and combine this with the previously used hierarchy of study types to determine the level of evidence (Harbour and Miller, 2001).

The quality assessment process in a SR is also carried out individually by two researchers and involves a third researcher to settle any discrepancies in quality assessment to account for reliability of the measurement. The results from the quality assessment are then given a code as to the overall assessment of the paper, as seen in Table 1. The results from the quality assessment are then combined with the type of study as developed by SIGN to obtain an overall level of evidence (i.e. 1+ meaning a RCT or SR of RCT that met most criteria in the quality assessment of the study.) When assigning study type according to SIGN hierarchy, Level 1 is given to randomized controlled trials (RCTs), or meta-analyses/SRs of RCTs because they are considered to have the strongest control of bias in the study. See Table 2. Level 2 is given to SRs that review case-control

studies without randomization, quasi-experimental controlled trials, or cohort studies. Articles judged as Levels 3 and 4 are given to case studies/reports and expert opinion, respectively. (SIGN, 2001).

Table 1. Coding for assessment of study (SIGN, 2001).

++	The majority of the criteria were fulfilled, but where the criteria were not fulfilled, the outcome of the study was very unlikely to alter.
+	Some of the criteria were fulfilled, but where the criteria were not fulfilled, the outcome of the study was unlikely to alter.
-	Few/no criteria fulfilled, and conclusions are likely to alter.

Table 2. Hierarchy of study type (SIGN, 2001).

Study Type	Explanation
1	RCT, meta-analysis, or SR of RCT's
2	SR of case studies, case-controlled studies, cohort studies
3	Case studies, reports
4	Expert opinion (i.e. conference reports, clinical experience)

The final step in a SR is to extract the pertinent data from the primary articles, usually set by a specified set of numerical values and/or descriptive aspects of the study, such as treatment effect, number of subjects, and study type (Kitchenham, 2004). The information to be obtained from each study should be formulated onto a data extraction form that is piloted before the actual data extraction occurs to assess the completeness of the extraction. The data can then be synthesized descriptively in a tabular format that highlights the major differences and similarities found across studies and patterns of outcomes across levels of evidence/study quality, sample size, or study type (Kitchenham, 2004).

Justification for Research

From the previous sections of this review of the literature, it is apparent that the line between a normal aging adult and an adult with mild cognitive decline that is not considered severe enough to be defined as *dementia* but is too severe to be *normal* is not well defined. Being able to identify those individuals in the early stages of cognitive decline such as MCI would be helpful in allowing for early intervention for individuals with MCI in what some define as the transitional stage between normal cognition and dementia. This possibility of early intervention will help in either maintaining the skills the individual still possesses or making plans and compensate for future losses in cognitive decline if the deficit does continue to progress (Luis, et al., 2003). It is also important to consider early intervention while the individual is exhibiting only mild cognitive deficits with MCI because it is easier to work in conjunction with the individual and family while the individual still retains the awareness of his/her own deficits (Lu, et al., 2007). The concept of MCI is also becoming increasingly more important in early identification since it has been reported that individuals with MCI progress to AD at a rate of 10-12% per year that is contrasted with the 1-2% progression rate to AD of normal aging adults (Petersen, et al., 1999). With the use of an SR, an exhaustive literature review could gather evidence indicating what assessment tools can distinguish MCI from normal aging cognitive abilities and by what parameters this distinction can be made. In the practice of speech-language pathology, many individuals with cognitive decline are treated, and it would also be useful to have a study such as a systematic review to reveal what evaluation strategies for MCI versus normal are available and which procedures may be accessible for use in this particular discipline. A position statement developed by

ASHA (2005b) for the professional statement of the field of speech-language pathology on the role of an SLP with individuals with dementia includes identification of individuals at risk for dementia as well as evaluation of the cognitive-communication disturbance. ASHA (2007) also states that evaluating cognitive domains such as memory, sequencing, executive function and sequencing is considered to be within the scope of practice of an SLP. Since the literature shows that there are many types of assessment procedures spanning various professions, it is important to have a collection of all available procedures including the research that is available to support them, in order to determine which tools are appropriate in the field of speech-language pathology. These current issues posed are the questions that the current study anticipates to answer through a systematic review of the literature

The original purpose of this study was to conduct a systematic review to answer the question: When comparing neuropsychological, medical, and allied health professional cognitive assessment measures, which does research suggest are the most effective diagnostic tools to distinguish MCI from normal cognitive ability in the aging population? After piloting the systematic review for feasibility, it was determined by a consensus of the committee that the medical aspect of the research question could be excluded. For the purposes of the field of speech-language pathology, understanding the full extent of neuropsychological and other health-related assessment protocols was thought to be more manageable and professionally useful. The question was also changed from comparing the two professional fields (neuropsychology and allied health) to compiling both, since research showed overlap in assessment protocols and professions, making sorting into groups difficult. The research question was changed to: What does

research suggest as the most effective neuropsychological and allied health professional cognitive assessment measures to distinguish MCI from normal cognitive ability in the aging population?

Effective in this study was defined as the assessment procedure with the highest level of evidence found with the best balance of sensitivity and specificity measures. *Neuropsychological* assessments in this study will be defined as assessment batteries that are typically given by a neuropsychologist, which *allied health* professional assessments will be defined as any battery that is available for multidisciplinary use. The most effective diagnostic tools were analyzed as to the parameters they are measuring and conclusions were drawn as to what parameters may be most sensitive to change and should be measured to make this distinction between MCI and normal aging. Once the assessment procedures had been determined, this study drew conclusions as to the types of procedures available to be utilized in the profession of speech-language pathology to distinguish between MCI and normal aging cognitive decline.

CHAPTER 3

METHODS

The first step in this study was to seek professional educational support from the Ralph B. Draughton Library at Auburn University on the best and most efficient ways to utilize the databases and library resources. A professional contact was made with the Subject Specialist for Communication Disorders in the Ralph B. Draughton Library Reference Department at Auburn University, and was used to help educate the investigator on how to use the databases and the differences between each one (N. Noe, personal communication, January 30, 2007). After discussing the study with the library professional, it was agreed that the databases that would pull from many sources in each discipline are PSYCINFO, MEDLINE, CINAHL, ERIC, and DIGITAL DISSERTATIONS to access all evidence concerning identification of MCI in the disciplines of medicine, nursing, speech-language pathology, psychology, and education.

Resources Used

The first database that was used in this study was PSYCINFO, an academic database pulling articles from many disciplines that have psychological relevance that as of March 2007 consisted of more than 2.3 million records (American Psychological Association, 2007). This scholarly database provides citations to articles in various journals such as behavioral sciences, mental health, social work, medicine, and education, 98% of which are peer-reviewed. PSYCINFO uses EBSCO Host interface, and when

using the advanced search option, keywords can be entered into individual search boxes. Phrase searching is used with this interface, which provides a phrase box for each phrase being used in the search and does not require each phrase to be truncated using quotations.

The next database that was used in the current systematic review was MEDLINE, using the interface of OVID. This database was selected to obtain all available evidence concerning MCI identification because it includes biomedical journal citations and abstracts from over 5,000 journals (US Library of Medicine, National Institute of Health, 2002). This database was created by the U.S. National Library of Medicine and using the OVID interface, provides a direct link to Auburn University Library's catalogue. After consulting with the professional contact, it was decided that the "map term to subject heading" feature of this interface which then narrows down the search through use of a tree for each subject heading would not be conducive to getting the citations needed for a systematic review and would therefore not be used in this study. In order to search through all available citations from a set of keywords, this option should not be selected (the default is selected). When using this database, each keyword must be divided by the word "and" but will not be recognized as separate keywords if separated with a comma or an "&". This interface, unlike the previous, does require each phrase being used to be truncated during phrase searching (i.e. "mild cognitive impairment").

For the current study the database CINAHL, Cumulative Index to Nursing and Allied Health Literature, was also selected to collect all available evidence in nursing, allied health, biomedicine, and healthcare literature on MCI identification. This database provides references to various types of literature needed for a systematic review such as

journals, books, dissertations, conference proceedings, and professional standards in healthcare (CINAHL Information Systems, n.d.). CINAHL, like MEDLINE, uses the OVID interface, in which the same applications were used in navigating the database regarding the division of keywords and not using the map to subject heading feature. A helpful feature that was included in the database searching was the ability to switch from MEDLINE to CINAHL (both using the OVID interface) by using the “Change Database” option. This made searching quick and efficient and should be noted for future database searching for systematic reviews.

Another database chosen for this systematic review was ERIC, Educational Resource Information Center. ERIC provides access to educational literature and is sponsored by the U.S. Department of Education, and was chosen because of its ability to access large amounts of literature related to education which was needed to fulfill the purpose of collecting all available evidence in varying disciplines for a systematic review. The final database selected for the current study is DIGITAL DISSERTATIONS, which includes both theses and dissertations to ensure that the systematic review is exhaustive of all available avenues of evidence. Along with the databases, professional contacts in speech-language pathology and were also made for access to any unpublished studies or conference proceedings that were available.

Search strategies

The next step in the process of conducting the systematic review was to determine the key words needed to insure an extensive search of the literature on the most efficient and effective way to distinguish MCI from normal cognitive decline. The investigator ran trial searches on all databases discussed in the previous section using the basic search

strings: “mild cognitive impairment, identification” and “mild cognitive impairment, assessment” and then tallied the keywords listed under each article to determine the most commonly used throughout the literature. After tallying the keywords, 10 of the most common keywords (tallied from at least 15 different articles) were chosen for the current study to combine in order to create search strings that yield the greatest amount of articles. The keywords *mild cognitive impairment, assessment, identification, dementia, normal aging, cognitive disorders, middle age, distinguish, neuropsychological tests, and diagnosis* were combined in search strings of two or three keywords for each database to gather the articles for analysis.

The third step in the process of conducting the systematic review was to retrieve the articles for analysis. According to the SIGN (2001), it is important to sift through the articles through a series of steps in order to remove the articles for analysis that are irrelevant to the topic. The current systematic review began the initial sift by having two individuals (the investigator and thesis advisor) review the abstracts of the articles retrieved from the database searches and include the articles relevant to the study and exclude the articles that are irrelevant to the current question of determining the most efficient and effective way to distinguish MCI from normal cognitive and communicative decline in the aging population. The abstracts were put into three groups, being Group 1 (include in full text analysis), Group 2 (possibly include in full text analysis), or Group 3 (exclude from full text analysis). Groups 1 and 2 were then included in a second sifting process by having the two researchers review the full texts to determine which articles were relevant to the current topic. The full text articles were grouped into either Group A (included in systematic review) or Group B (excluded from systematic review). A third

individual, a member of the thesis committee was used as a mediator if any discrepancies between the two judges occurred at any level of the sifting process.

Inclusion and Exclusion Criteria

This study included articles that compare MCI from normal cognition, but if the study only looked at MCI and dementia, it was excluded. Studies that compared all three levels of cognitive function, normal, MCI, and demented, were included, because the comparison of MCI and normal was available for analysis, and the comparison to dementia was ignored in analysis. Many articles found in preliminary searches of the databases yielded studies differentiating between MCI and dementia, but the current study is only concerned with the diagnosis of MCI from normal cognition. In order for this systematic review to find the relevant articles for this topic of the aging population, the articles chosen required the average participant age to be 55 years or older. According to SIGN (2001), when conducting a systematic review on a topic that is rapidly growing area, it is appropriate to limit the search to the last 10-15 years. Since the diagnosis of MCI is rapidly growing more awareness in the medical and allied professional fields, articles ranging from 1992 to the present were included, and any articles before 1992 were excluded according to SIGN (2001) suggestion.

Evaluation of the Evidence

Quality assessment

Once the appropriate articles were chosen by the researchers, a methodology assessment procedure adapted from SIGN (2001) to evaluate articles of diagnostic accuracy of an assessment tool was used to systematically evaluate the evidence. (See Appendix for checklist). This analysis included quality assessment questions of the

internal validity of the article by addressing the study design and then assigned a code of ++, +, or -, to rate the assessment of the design study, as explained in the literature review. Questions in the portion of the quality assessment included:

- Was the nature of the assessment procedure explained in the study?
- Was the test being studied compared with an established reference standard that has research to support its ability to accurately make a diagnosis?
- If the reference standard was not established/validated, did the study justify the use of the chosen control that has known specificity and sensitivity?
- Was the patient population chosen randomly or measures taken to ensure that the population was not chosen to encourage a particular outcome of diagnosis?
- Were the measurements of the test under question and the reference standard control obtained independently of each other so to ensure that the examiner is not scoring the test with any bias from previous performance?
- Were the tests (experimental and control) administered as close together as possible?
- Were the results reported for all participants in the study?
- Was a pre-test diagnosis made and given in the study?

These questions were applied to each study included in the SR and judged subjectively by two trained experimenters (the investigator and another trained graduate student familiar with SIGN quality assessment) by assigning as either well covered, adequately covered, or poorly addressed. If the topic addressed in the question was not mentioned in the study, the question was assigned “not addressed,” but if the topic was mentioned but not in detail, the question was assigned a judgment of “not reported.” If the question did not

apply to the study, “not applicable” was chosen. The former three judgment codes were used as much as possible and the latter three were used only if necessary. Reliability measures were also taken into account by incorporating a third judge (thesis advisor) to make a decision about any point-by-point discrepancies between the two judges on the quality assessment. The study was then coded from the results of the quality assessment. This +/- code, along with the level of evidence (1-2) assigned according to the type of study being analyzed, was combined to give an overall score of the study. Studies falling under the categories of 3 and 4 were not included in the quality assessment, but were included in the data synthesis table to draw conclusions from all levels of evidence. See Table 3 for further explanation.

Table 3. Overall assessment of article, adapted from SIGN, 2001.

1++	RCT, meta-analysis, or SR of RCT's with excellent quality assessment and therefore, having a very low risk for bias in the outcomes
1+	RCT, meta-analysis, or SR of RCT's that met most of the quality assessment criteria, yielding a low risk of bias in the outcomes
1-	RCT, meta-analysis, or SR of RCT's that did not meet the majority of the criteria, yielding a high risk of bias in the outcomes
2++	SR of case studies, case-controlled studies, cohort studies, retrospective studies with excellent quality assessment and therefore, having a very low risk for bias in the outcomes
2+	SR of case studies, case-controlled studies, cohort studies, retrospective studies that met most of the quality assessment criteria, yielding a low risk of bias in the outcomes
2-	SR of case studies, case-controlled studies, cohort studies, retrospective studies that did not meet the majority of the criteria, yielding a high risk of bias in the outcomes
3	Case studies, reports, feasibility studies
4	Expert opinion (i.e. conference reports, clinical experience)

Data extraction and synthesis

After the quality assessment was performed on the studies, the articles were analyzed by looking at various factors that affect the outcome such as number of participants, participant characteristics, the test procedure being studied, the reference standard used, sensitivity and specificity measures, predictive value of the test, and issues that the results raise in relation to the question of the effectiveness of a test measurement in distinguishing MCI from normal cognition. Other aspects of multiple studies that arose as pertinent to the study as the analysis takes place were also included in the data extraction and synthesis. These aspects of each study, along with the overall score of the study (level of evidence plus quality assessment) were summarized in tables according to each type of discipline (i.e. neuropsychological, medical, allied health) to make comparisons between and across disciplines. If multiple articles yielded similar measures in sensitivity and specificity or other outcome measures, tables were compiled to compare these measures and the results of each will be interpreted in the discussion.

Along with creating tables to show the results of the quality assessment, descriptive results were discussed in paragraphs as to the overall impression of each study. Along with the results from the quality assessment checklist, the descriptive results discussed possible threats to internal and external validity of each study. The narrative of each study along with the tables showing the results of the quality assessment decreased the possibility of losing any valuable information by using qualitative measures and stringent tables only.

After the data was extracted, synthesized in tables and discussed descriptively in paragraphs, clinical implications were discussed as to the most effective assessment

procedures available from the level of evidence and the measures of sensitivity and specificity found in the various assessment procedures. Conclusions were also drawn as to the parameters of an assessment tools that help to distinguish MCI from normal cognitive ability in the aging population and the specific tests available for the use of a speech-language pathologist.

CHAPTER 4

RESULTS

Method of Inclusion

Thirty-eight out of 532 articles were chosen for this systematic review (SR) after two sorting processes, described fully in the methodology. In order to gather all available evidence on the diagnostic tools used to distinguish MCI from normal cognitive function, the search strings were input into PsycINFO, Medline, CINAHL, Dissertation Abstracts and ERIC databases. The databases yielded a total of 532 articles. Due to overlap from the various search strings and different databases in respect to the journals included in each database, there were a great number of duplicates. Once the 218 duplicates were pulled out of the group, the number of articles totaled 314 that were then used for the first sorting process. After applying the inclusion and exclusion criteria to each of the 314 abstracts, a total of 99 articles met the requirements to be analyzed in the second sort at the full-text level. The sorting process can be seen in more detail in reference to each step and each database in Table 4. The PsycINFO database yielded the largest amount of articles that fit the criteria, totaling 61 to be analyzed at the full-text level. The sorting process was accomplished by both the author and thesis director in order to maintain reliability and 11 abstracts were taken to a member of the thesis committee for resolution due to discrepancy in decisions between the 2 main investigators.

Table 4. The sorting process according to each database.

Database	With duplicates	Without duplicates	Criteria met after abstract analysis	Criteria met after full text analysis
PsycINFO	228	182	61	25
CINAHL	112	72	17	8
ERIC	4	2	0	0
Dissertation Abstracts	21	13	6	2
Medline	167	45	15	3
Total	532	314	99	38

The 99 full-text articles were then analyzed by both the investigator and thesis advisor to maintain reliability of findings. Five full-text articles were taken to a member of the thesis committee for resolution. In the initial stages of this project, the investigator planned to include articles in the medical field of study in the SR, but after sorting and analyzing the full-texts of the articles, it was determined by the investigator and committee that the results found in the medical journals would not directly apply to the field of speech-language pathology and therefore would not be as applicable as other areas such neuropsychology. A decision was made to exclude the 30 medically-based articles. Five progression studies were included in the full-text analysis to ensure that the results did not apply to the SR topic of discriminating between MCI and normal cognition, but were then excluded after analyzing the texts. Ten review articles were excluded at this level of the sorting process along with 16 others that did not meet other criteria such as language, inclusion of controls, or prevalence studies. The breakdown of the full-text sort by database and exclusion criteria can be seen in Table 5.

Table 5. The breakdown of number of articles not meeting criteria for inclusion at the full-text level of sorting.

Database	Medical	Review	Progression	Other	Total
PsycINFO	19	4	3	10	36
CINAHL	6	1	0	2	8
Dissertation Abstracts	1	1	0	2	4
Medline	4	4	2	2	12
Total	30	10	5	16	61

Thirty-eight articles were included in the SR after the full-text analysis. The checklist adapted and modified from SIGN (2001) was utilized to analyze each article and rate the methodological quality of the study. This checklist can be seen in the Appendix. The investigator along with a secondary reader, a fellow graduate student and researcher familiar with the modified SIGN checklist, individually rated each article and then brought discrepancies to the thesis director for resolution. This occurred for 4 articles and only when the discrepancy would result in a change in the overall rating of the article. For example, if one rater judged the reference standard as “poor” while the other rated as “adequate,” this would change the overall rating to either +/- and would therefore be taken to the thesis director for resolution. If the discrepancy was only over an “adequate” or “well addressed” rating, this was resolved among the two raters and not brought to the thesis director since it would most likely not change the overall rating of the study. Along with the checklist, each article went through a data extraction of various parameters including as inclusion and exclusion of participants along with the test in question and gold/reference standard(s) used for comparison. These parameters can be seen in detail in the summary tables section, Table 7. Also included in the data extraction process was participant and study information which can be seen in the summary tables section, Table 8.

Along with the database search, personal contacts were made with professionals in the field of speech-language pathology who specialize in the area of cognition in geriatric individuals in order to obtain any available unpublished literature in the area of MCI within the speech-language pathology field. Two contacts were made, one (R. Lubinski, personal communication, October 5, 2007) stating that she was unaware of any specific tests for MCI but that some screening tools that may appropriate would include the *Arizona Battery for Communication Disorders of Dementia (ABCD)*, *Mini mental status examination (MMSE)*, *Short portable mental status questionnaire*, *Clock Drawing Test*, *7 minute screen*, *RIPA*, *Mini Cog* and *Hopkins Verbal Learning Test (HVLIT)*. Another contact was made (K. Bayles, personal communication, October 5, 2007) suggested the use of the *ABCD* for a screening tool for MCI, and referred the investigator to a recently published book. The results in the book suggested that the *ABCD* was normed on mild, moderate and severe AD patients and the subtests Story Retelling-Delayed and Word Learning best distinguished AD from controls and were appropriate to use as screening tools for determining people at risk for developing dementia (Bayles & Tomoeda, 2007). No mention was given to an MCI population within the normative sample of the *ABCD* although this may be an area of future investigation.

Review of Articles

When compiling the data, patterns were shown in the literature as to the types of tools available, which are tallied and seen in Table 6. A large of number of studies in the formal tests group were also rated as a (+) or better, as seen in Table 6. The second group of tests was grouped as screening protocols that could be administered to identify individuals who require further evaluation and who do not. The miscellaneous group

includes informal and formal assessments that were newly-made or developed from other tests such as intraindividual variability of individuals on a variety of tasks over a period of time or a tool used to analyze discourse in a narrative. Two other groups, computer-based tests and equipment-based (other than computer) were also included in the review. The following paragraphs offer descriptions of each study involving the methodology, findings and quality assessment rating of the study.

Table 6. Tally of number of articles in each diagnostic test group.

Diagnostic test group	Number of articles tallied in group	Number of articles in tally with a QA of (+) or better
Formal Test Batteries	14	11
Screening Tool	9	6
Computer-based	4	1
Equipment-based (EEG)	1	0
Miscellaneous	10	7

Formal test batteries

Greenaway, et al. (2006) investigated the use of verbal memory in discriminating between MCI, AD and normal cognitive ability in the aging population. The *California Verbal Learning Test (CVLT)* is a 16 word learning test that gives multiple trials to evaluate if learning the words over a period of time improves performance of memory. Other measures are included in this test such as the amount of time for recall and types of errors made when recalling the word list. Using the *CVLT*, the study compared the results to a battery of neuropsychological tests to evaluate non-memory domains of 195 participants. The results found that using the *CVLT* more accurately discriminated between MCI and normal cognitive ability than the non-memory measures of the other tests (see Table 9). When analyzing the results from the *CVLT*, discriminant analyses showed that 2 measures, learning and delayed recall, discriminated accurately for 68.7%

of participants. According to the authors, other studies have shown better discriminatory power when combining multiple neuropsychological tests, concluding that multiple tests and/or data points may increase accuracy of diagnosis. When comparing the *CVLT* to the non-memory measures no statistical differences were found, although some participants were not given the non-memory measures, making the number of participants smaller. The results of this study support the use of the *CVLT*, specifically measures of learning and delayed recall when discriminating between MCI and normal cognitive ability.

Judging the study found no reported reliability and poor generalizability of the findings to the overall population due to the lack of minority inclusion and high educational status of the participants. Bigger numbers of participants may have changed the results when comparing the *CVLT* to the non-memory measures, although statistical differences were found on the test in question, the *CVLT*. Double blinding was employed in the study since the participants were not yet aware of the diagnosis during the *CVLT* and the test administrators diagnosed at the same time as giving the *CVLT*. This also decreased the chance for any mortality or maturation effects to occur. The overall rating of the study was a (+), suggesting that the *CVLT* is a sensitive tool in discriminating between MCI and normal cognitive ability, especially the areas of verbal learning and delayed recall.

Mioshi, Dawson, Mitchell, Arnold, and Hodges (2006) evaluated the diagnostic utility of a revised version of the *Addenbrooke's Cognitive Examination (ACE)*, which is a screening tool used frequently in the United Kingdom. The changes made in the revised edition allowed for easier administration and other problems that occurred in the original edition such as ceiling effects and difficulty answering questions. After piloting the new

test (*ACE-R*), the authors wanted to test the discriminatory ability of the *ACE-R* for the various diagnostic groups including MCI, dementia, and controls. Two hundred and forty-one participants were given the 12-20 minute test. The MCI group was diagnosed a priori using the typical Mayo clinic guidelines for MCI, although it was not specifically stated that the authors used these criteria. The *Clinical Dementia Rating* was used as the reference standard in comparison with the results of the *ACE-R* to determine validity. Mixed ANOVAs found significant differences between all subtests except visuospatial for detecting MCI. The authors only used subgroups of 23 from each diagnostic group for statistical analysis instead of using the larger number of participants that entered into the study. Sensitivity and specificity were measured but only for detecting dementia. A significant negative correlation of the *ACE-R* to the *CDR* was found, determining that the *ACE-R* has good construct validity. The authors concluded that the *ACE-R* is an appropriate test to help identify MCI from normal cognitive ability.

When rating this study, the reference standard was determined to not be validated when being used for comparison alone, although the methodology did state that the *MMSE* and the *WMS-R* Logical Memory II subtest were also used in the pretest diagnosis, but not included in the statistical analysis. Since additional testing was done, the study was not rated poorly in the area of the reference standard, even though additional statistical analysis would have been helpful. Blinding of the administrators was employed to the participants' *CDR* scores to control for administrator bias. Reliability and construct validity were also reported. The authors also gave explicit details of the nature of the test in the article. These factors boosted the quality assessment of the study. Overall, this study was given a score of (+). The results determined that the *ACE-R* is an

appropriate diagnostic tool for MCI evaluation, although this is used more in the United Kingdom and not in America.

Pernecky, et al. (2006) determined how assessing complex ADL could help distinguish between controls and individuals with MCI. Seventy-five participants were recruited from a university hospital for this study. The assessment tools used were the Alzheimer's disease Cooperative Study scale for ADL in MCI (*ADCS-MCI-ADL*) and the Alzheimer's disease Assessment Scale, cognitive subscale (*ADAS-cog*). The reference standard included a full neuropsychological evaluation including the *Consortium to Establish a Registry for Alzheimer's Disease-NAB* (German version), *WMS*, *Trail Making Test*, *CDR*, and neurological evaluation along with other various tests. These references to determine diagnostic groups a priori were deemed as validated reference standards in the area of MCI diagnosis due to the exhaustive nature of the testing and use of many tools to make a diagnosis. This study did vary from the Mayo Clinic criteria for MCI which included complex ADL being impaired (basic still intact) and no subjective memory complaint. Adequate sensitivity and specificity measures were found for both tests, but slightly better discriminatory ability was found in the *ADCS-MCI-ADL* test than the *ADAS-cog*. Comparison of these two tests to the reference standards used for a priori diagnosis using logistic regression analyses found significant values for both the *ADCS-MCI-ADL* and *ADAS-cog*. These results determined that the inclusion of complex ADL in the evaluation of MCI is helpful, especially since as the widely-used criteria states that basic ADL are not affected in MCI.

Judging the methodology found no reliability measures included in the article or mention of blinding of the administrators. The tests being evaluated were given 4 weeks

after initial diagnosis was made which is sufficient for controlling for testing effects (especially if given the same day) and ensuring not too much time had elapsed for maturation to set in (i.e. cognitive decline). Overall, the methodology was well-controlled, giving a rating of (+). These results suggest that both the *ADCS-MCI-ADL* and *ADAS-cog* are appropriate tests for discrimination between MCI and normal cognitive ability.

Chandler, et al. (2005) developed a total score for the *Consortium to Establish a Registry for Alzheimer's Disease (CERAD)* which is a combination of various neuropsychological tests that before had to be analyzed subtest-by-subtest. Two hundred and fifty participants were included in the study to determine the diagnostic ability of the *CERAD* to distinguish between controls, MCI and AD. The tests included in the *CERAD* are animal naming, modified *Boston Naming Test*, *MMSE*, constructional praxis, and word list memory (including word list learning, recall and recognition). Using subgroups of AD and controls to develop a way to determine the total score, addition of each subtests to obtain a total score was determined to be the method of choice due to ease of computation and accuracy according to results. A second subgroup was chosen to determine specificity and sensitivity of the total score to different diagnostic groups: controls, MDI and AD. The results showed that the total score for the *CERAD* was superior in distinguishing MCI from normal cognitive ability in comparison to the *MMSE*. Although the *CERAD* takes more time to administer, according to the authors, it has better discriminatory ability than the widely-used *MMSE*. The *CERAD* did have identical sensitivity to the word list recall subtest for MCI, but did have better specificity to the controls than did the word list recall subtest. Overall, the results suggest that using

the computation to obtain a total score of the *CERAD* is helpful in discriminating MCI from normal cognitive ability instead of having to look at each individual subtest separately.

Rating of this study yielded excellent test-retest reliability and convergent validity measures, but this was done with the original participant pool of controls and AD, not the second group that included MCI, which was the group that is in question in this SR. No minorities were included in this study, although demographics were at least reported in the study. Appropriate measures were taken to ensure methodological control with administrators and testing environment. This study overall was well-controlled and therefore received a rating of a (+), suggesting that the *CERAD* and the use of addition of subtest scores be considered an accurate diagnostic tool in discriminating MCI from normal cognitive ability.

Karrasch, Sinerva, Gronholm, Rinne, and Laine (2005) investigated the sensitivity of the *CERAD* to 15 individuals with MCI, 15 with AD and 15 with normal cognitive ability. The authors chose to include a neuropsychological evaluation with tests such as the *WMS-III-R*, *Wechsler Adult Intelligence Scale-Revised (WAIS-R)*, *Trail Making Test*, and *BNT* (among others) as reference standards to the *CERAD*. The *CERAD* includes various subtests such as Word list Learning (over three trials), Word list Delayed Recall, and Word list Recognition. Statistical analyses utilizing MANOVAs determined that only the scores on the Word list Learning subtest were significantly worse for the MCI group compared to the controls. In determining sensitivity and specificity of the tests to the different diagnostic groups, the sensitivity to MCI on all subtests of the *CERAD* was low, although specificity was high. Another interesting finding of this study was the low

sensitivity of 13% for the *MMSE* in discriminating MCI from controls. This is interesting since the *MMSE* is used often as a screening procedure and in conjunction with other assessment tools. The authors suggested that new screening procedures need to be developed with better sensitivity to these subtle cognitive changes.

When judging the study, the group number of 15 participants per diagnostic group (45 total) was quite low and a larger number could have increased statistical power that may have found more differences in the MCI group's performance compared to the control's. The article stated references as to the accuracy of a neuropsychological evaluation and sensitivity to subtle changes in cognition, which was adequate evidence to deem the reference standards as validated for this SR. The authors did suggest that increasing the delay time (usually 5 minutes for the *CERAD*) of the memory task may increase sensitivity to MCI since it is increasing complexity of the task. No blinding was mentioned in the study. The *CERAD* was performed a month after the neuropsychological evaluation which is adequate time for decreasing the chance of testing effects, but increasing the chance for morbidity or maturation. The article stated that the *CERAD* was given a month after the initial evaluation for the MCI and AD groups, but nothing was mentioned as to when the controls were given the *CERAD*. For the reasons stated above, this study was rated as (-) although other articles included in this SR have found that the *CERAD* is a useful tool in discriminating between MCI and controls. Again, the findings may have been different if a larger sample size was used and should be taken into account when applying the findings into practice.

Nordund, Rolstad, Hellstrom, Sjogren, Hansen, Wallin (2005) analyzed 21 neuropsychological tests in the domains of speed/attention, memory/learning,

visuospatial function, language and executive function to determine which best discriminated 112 individuals with MCI from 35 individuals with normal cognition in Sweden. Six test administrators were used to conduct a priori diagnoses with the use of tests such as the *Stepwise Comparative Status Analysis (STEP)*, *MMSE*, *CDR* and *I-Flex*. Little participant information such as gender, language or education was stated in the article. The age of the participants was found to be significantly different between groups, with the MCI group being significantly older than the control group. The tests that showed significantly lower scores for MCI using the non-parametric Mann-Whitney U test were *Digit symbol*, *Trail Making Test A*, *RAVLT*, *Logistic Memory Delayed Recall*, *Visual Object and Space Perception silhouettes*, *Boston Naming Test*, *Assessment of Subtle Language Disorders* repetition subtest, *Parallel Serial Mental Operation*, *Picture Word Test* (see Table 10). The conclusions of the study showed that one of the tests analyzed, the *WAIS-R* did not show significant diagnostic ability with the exception of one subtest, showing that using intelligence scores are not sensitive to the subtle cognitive changes in MCI. The results of the study also showed that MCI is a heterogeneous population and that people exhibit various deficits, so using a large battery of tests spanning various cognitive domains is best for diagnostic accuracy.

Some concerns when rating this study were the lack of inter-judge and test-retest reliability reported along with the increase in chance of bias due to the lack of blinding for the 6 test administrators of the participants' a priori diagnoses. The study also reported that some data was not included, but no numbers were given as to how many data points were missing. The reference standard of the *MMSE* and *CDR* along with the *STEP* and *I-Flex* tests were used to put the participants into diagnostic groups, but no rationales were

given as to why these uncommonly-used tests (i.e. *STEP* and *I-Flex*) were chosen as the standards for this study. These concerns of methodology determined the poor (-) rating for the study. This study does make an important point, though, that the MCI population is heterogeneous and may need extensive testing for an accurate diagnosis.

Woodard, Dorsett, Cooper, Hermann, and Sager (2005) investigated a group of neuropsychological tests, including the *CERAD*, and determined which would best differentiate patients who had normal cognitive ability versus those with MCI or a neurocognitive disorder. Two hundred participants were given a specific neuropsychological battery lasting 60-75 minutes in the same order which included the two measures that would be used for an a priori diagnosis (*RAVLT* and *MDRS*). The statistically different scores (using ANOVAs) were then analyzed using ROC curve analyses to determine the sensitivity and specificity of the different tests. The results revealed good sensitivities but somewhat low specificities (i.e. increase in false negatives). Corresponding to these measures, positive predictive value (a result of normal) was low and negative predictive value (obtaining a diagnosis of MCI) was high. The MCI group was then combined with a neurocognitive group to determine a diagnostic algorithm that included the Verbal Category Fluency and Word list Delayed Recall subtests from the *CERAD* that would be used to determine normal or abnormal cognitive ability and the need for further evaluation. The ROC curve for this diagnostic algorithm resulted in a good balance of sensitivity and specificity and could be used as a quick way to screen for cognitive impairment. The authors suggested that a diagnostic algorithm using the *CERAD* subtests of Word list Memory Delayed Recall and Verbal Category Fluency is appropriate to discriminate between MCI and normal cognitive ability

When rating this study, testing effects such as fatigue were in question since all tests were given in the same order. This could have been reduced with counterbalancing, but this method of consistency did implement blinding into the methodology since no diagnosis would have been made during all of the testing. Also, the administration time between the reference standards and the tests in question, including the *CERAD*, was on the same day, controlling for maturation or other outside factors along with a consistent testing environment. Reliability was addressed by having a second judge check the diagnostic results of the original test administrator, who was a clinical social worker. All methodological factors listed above resulted in a rating of a (++). These results corroborate the results of Lopez (2004), who suggested the use of semantic fluency and word list recall and also the results of Knox (2004), who suggested the use of a diagnostic algorithm in the evaluation of MCI.

Knox (2004) developed and investigated the ability of a checklist-based diagnostic algorithm in discrimination between MCI and normal cognitive function. Three hundred seventy participants were included in this study and sorted into groups using the *MMSE* and *Wechsler Adult Intelligence Scale-Revised (WAIS-R)* as reference standards. The neuropsychological tests with significant differences in performance compared to MCI and controls were included in the diagnostic algorithm which included the *Dementia Rating Scale (DRS)*, *CVLT*, *WMS-R/III*, Verbal fluency, *BNT* and *Trail Making Test-B (TMT-B)*. The first three tests listed are given a score of either -1 (higher scores obtained on the tests), 0 (normal-slightly low scores on the tests) or +1 (lower scores obtained on the tests) to determine the participants with normal cognition. If the score is greater or equal to 0, the evaluation continues with the other three measures to

determine if the patient falls into either the MCI or AD category. Diagnostic algorithms are used in medicine and other professional areas and bring an objective element to the assessment procedure. The results of this study determined that the algorithm accurately grouped individuals with only 2 out of 40 MCI diagnoses made a priori being put into either the control or AD group with the checklist-based algorithm. Good results were found with the ROC curve analyses between normal versus impaired and MCI versus AD, although the investigator determined that this is not a replacement for clinical judgment and should be used to supplement the evaluation process.

While rating this study found a much smaller number of MCI participants compared to the other groups, significant results were still found and therefore deemed appropriate. One concern in the study is the use of the *MMSE* and an intelligence scale (*WAIS*) as the methods for a priori diagnosis when many other studies have used more exhaustive measures as reference standards. Although this area was in question for the raters, the entire methodology was taken into consideration, and since more than one test was used as a reference, this was judged as adequate. The methodology did control participant selection by ensuring that the groups were similar in age and educational status and only differed in cognitive status. The testing environment was well controlled and the steps to administering the diagnostic algorithm were explained in detail in the dissertation. This was a well-controlled study with the use of a development and validation cohort to ensure adequate consistency and reliability. The study was rated as (++) in this SR determining that the use of a diagnostic algorithm with a checklist-based approach is an appropriate tool in discriminating between MCI and normal cognitive function.

Lopez (2004) investigated the ability of the paper-and pencil *Repeatable Battery for the Assessment of Neuropsychological Status (RBANS)* and newly developed computerized *Automated Neuropsychological Assessment Metrics (ANAM)* to discriminate 18 individuals with MCI from 94 individuals who were controls and which subtests or underlying constructs most accurately predicted diagnostic group to possibly eliminate the requirement of a time-consuming full neuropsychological examination. Four individuals with AD were also included in the study. These 2 tests were administered to determine how they compared to the reference standards of the *Rey Auditory Verbal Learning Test (RAVLT)*, *Mattis Dementia Rating Scale (MDRS)*, and *Lawton's IADL* which were used for a priori diagnosis. The results determined that both the *RBANS* screening measure and the *ANAM* memory measures discriminated between the two groups and even a non-memory measure of coding (symbolic understanding) accurately discriminated the groups. The purpose of the study was to determine a way to measure the differences in a shorter amount of time, so a discriminant analysis of the subtests with most significant AUC from both batteries was done determining that three measures from the *RBANS* best predicted diagnosis: Semantic Fluency, List Learning, and Delayed Memory Index. Also, subtests were combined and ROC curve analyses were performed determining that performing only the Semantic Fluency and List Recall subtests or the Coding and List Recall subtests from the *RBANS* yielded good sensitivity and specificity measures, although the full test would give more accurate results. Overall, the typical pencil and paper exam (*RBANS*) had higher discriminatory power (larger overall AUC) than the novel *ANAM*, although it was reported that previous studies showed similar results in the *ANAM* and more widely used neuropsychological tests.

These tests can be given in a much shorter time frame than a larger neuropsychological battery, increasing time and cost effectiveness. Overall, these two batteries are accurate in the subtle detection on MCI, but if a quicker screener needs to be attempted, it is recommended to use List Recall, Semantic Fluency and/or Coding from the *RBANS*.

Rating this study yielded that the MCI group was quite small compared to the control group, which is comparable to the general population, but does not bring as much statistical power to a study. Statistical differences were found though, and although the authors recommended further research with more participants, the number was judged as adequate for this systematic review. Ethnicity was predominantly Caucasian, but at least was mentioned which has been omitted from most studies in this SR. Validated reference standards were used for a priori diagnoses of the participants. Reliability and blinding were well covered in the study. Although testing effects and fatigue could have played a factor in performance since no counterbalancing was used throughout the 2 ½ hour neuropsychological exam, all testing was done in one session and therefore controlled for maturation and other outside factors. These factors gave the study an overall rating of (++) for the thorough report of methodology.

Mejia, Gutierrez, Villa and Ostrosky-Solis (2004) studied the cognitive tests that best discriminated between MCI, normal and dementia in 314 Spanish-speaking individuals in Mexico City. This is a population who has not been studied as much as English-speaking Caucasians according to the literature, so the authors of this study determined that it was important to develop standards for evaluation of people with cognitive decline in this population. The three groups were assessed on multiple tests of cognition including the Spanish version of the *MMSE*, the *Brief Neuropsychological*

Evaluation for Spanish-Speaking Subjects (NEUROPSI), the *Short Blessed Test* and two tests of ADL: the *Blessed Dementia Rating Scale (BDRS)* and the *Pfeffer Functional Activities Questionnaire*. Diagnostic groups were made by consensus of the investigators although no specific information was given as to the assessment battery used to determine cognitive function. The multivariate analyses showed that there were group differences on all tests although no post hoc analysis was done to determine where the differences were occurring. There were significant differences in performance on tests due to age and educational status, which was rather low for an average of 2.9 years. The participants who were illiterate performed significantly worse than the more educated peers. This shows that these tests are biased to education and age. An analysis was done to determine how many false positives and false negatives occurred with all tests in diagnosing MCI. All tests in question had high false negative results ranging from 57-92% of MCI participants being determined as having normal cognitive ability. The highest in false negatives was the Spanish version of the *MMSE*. These results show that these tests may not be the most powerful tests to distinguish between the subtle decline in cognition of MCI versus normal cognition, especially in a population of lower educated, Spanish-speaking older adults.

When rating this study, the lack of information about the reference test used to make a priori diagnoses was a concern, leaving no room for comparison of the results of the tests being evaluated. Also, statistical information including post hoc analyses was insufficient although the authors did suggest that these tests may not be the most accurate measures to use when trying to identify MCI in the general population. These reasons along with the low educational status of the participants resulted in a methodology rating

of (-), suggesting that more research needs to be done on these tests in the Spanish-speaking population.

de Jager, Hogervorst, Combrinck, & Budge (2003) studied the sensitivity and specificity of a group of known and novel neuropsychological tests in the diagnosis of MCI from normal aging, along with other diagnoses such as AD. The total number of participants in the study was 152. The reference standard used was the *Cambridge Examination for Mental Disorders of the Elderly (CAMDEX)* which included the *MMSE*, *CAMCOG*, and an informant interview. Sensitivity of 86% for the *CAMDEX* was reported, but this number was in identifying AD and not MCI. This study looked at a battery of tests that would take up to an hour and a half to administer and found which specific tests best discriminated between the different diagnostic groups. Using non-parametric statistics and ROC analyses, the *Hopkins Verbal Learning Test (HVLT)* resulted in the best discrimination between diagnostic groups, including MCI versus controls. Other tests other than the *HVLT* that gave accurate diagnostic results included Category Fluency, and Letter Comparison Speed; but overall, most tests did not show strong discriminatory ability between these two diagnostic groups.

When rating the methodology, no inter-judge reliability was reported, although a consensus of 2 clinicians was required for a diagnosis. The overall number of participants was sufficient for statistical analysis. The tests were given in the same order to all participants, which was good for consistency but may have resulted in testing effects such as fatigue for the tests given towards the end of the evaluation resulting in lower performance. Although there the concerns in this study were listed above, the overall

methodology was well-controlled and resulted in a rating of a (+), supporting the use of the *HVLT* in discriminating between MCI and normal cognitive ability.

Estevez-Gonzalez, Kulisevsky, Boltes, Otermin and Garcia-Sanchez (2003) compared the *RAVLT* to the *MMSE*, *GDS*, *Blessed Dementia Rating-ADL* and an Informant Questionnaire to differentiate between MCI, normal and dementia of the Alzheimer's type (DAT). DAT was confirmed over a 2 year period with the longitudinal study design as a follow-up. The total number of participants in the study was 70. The *RAVLT* consists of 5 trials of word-list learning of 15 words to determine different measures such as verbal learning (how much more was recalled at the 5th trial compared to the 1st trial), immediate recall, delayed recall (after a 20 minute delay), verbal forgetting and percentage of forgetting. As seen in Table 9, Trials 2-5 significantly differentiated between controls and MCI along with immediate and delayed recall and verbal learning. This is an interesting finding showing the possibility of decline in verbal learning ability of multiple trials in MCI whereas controls performed significantly better. A MANOVA also confirmed the verbal learning decline by charting the learning curve of all 3 groups, finding significant differences between all. Significant differences were found between groups in reference to gender and age.

When rating this study, the reference standards of short screening measures and questionnaires were found to be somewhat poor considering the other studies in this SR using more in-depth neuropsychological batteries and/or tests. Another concern was that there were significant group differences in age and gender, although this was corrected using these two factors as covariates. The influence of testing effects and fatigue on the performance of the *RAVLT* that was given last after all tests with no mention of

counterbalancing was also found during the rating process. No reliability or blinding was mentioned in this study. The average educational level for the participants in this study was 7 years whereas other studies have included participants with 12 or more years of education. This gap in educational level makes the external validity and generalizability of the results questionable. All of the factors mentioned previously are what attributed to the rating of (-) that was given to this study that concluded that MCI can be detected using the *RAVLT* paying close attention to verbal learning across trials. More powerful, controlled studies are needed to confirm this finding.

Pasqualetti, et al. (2002) compared the diagnostic abilities of the *MMSE* to the *Mental Deterioration Battery (MDB)* in a retrospective study to determine the diagnostic ability in discriminating between controls, MCI and AD. The *MDB* is a large neuropsychological battery that tests different areas of cognition including language, verbal memory, visual memory, reasoning and constructional praxis. Three hundred participants were included in the study. The methodology of the study included a factor analysis of the underlying constructs that account for the variation in performance and therefore diagnostic ability of the *MDB*. Three factors: visuospatial, verbal memory and language ability accounted for 75% of the variance of the *MDB*. These findings were consistent with factor analyses done previously by the developers of the *MDB*. These 3 factors were then applied to the *MMSE* to determine how well the correlating questions on the shorter test would account for the variance. The results yielded smaller numbers than the *MDB*, with about 62% of the variance being accounted for with these factors on the *MMSE*. The findings when using a curve-fitting procedure showed that visuospatial and language abilities on the *MMSE* had linear regressions while verbal memory had a

cubic regression. The article also suggests that traditional cut-off scores are not appropriate on the *MMSE* to determine MCI. The authors also suggested that looking at long-term memory may better indicate subtle cognitive decline. The authors concluded that the diagnostic ability of the *MMSE* overall is to be used cautiously, and looking at individual areas such as visuospatial may aid in the decision making process.

When rating this study, external validity was judged as poor considering the *MDB* is normed only on the Italian population. Also, the Italian population was used as participants, which does not generalize to other countries. The average educational status of the participants was relatively low. Although this study was retrospective, the investigators did use a consecutive series methodology to select participants to reduce bias and also controlled for the environment, which was judged to be sufficient methodological control to obtain a rating of (+). This study suggested that the *MMSE* may not be the most adequate measure to base a diagnosis of MCI on, and additional, detailed testing needs to be done.

Xiao, Yao and Zhang (2002) evaluated the use of the *World Health Organization Neuropsychological Battery of Cognitive Assessment Instruments (WHO-BCAI)* to distinguish between controls, MCI and AD. The *WHO-BCAI* includes subtests of verbal learning, trail making, sorting, concentration, language, psychomotor, visual gnosis and spatial construction. The control group consisted of 83 participants while the MCI and AD groups consisted of 27 and 26 respectively. The reference standards used for a priori diagnoses included the *MMSE*, ADL-21 scale and *Global Deterioration Scale*. The study found that the subtests from the *WHO-BCAI* that significantly discriminated between MCI and controls were verbal learning, verbal fluency, mini-token tasks, visual

reasoning, trail making B, sorting and construction although a thorough description of these subtests was not reported in the study. The authors concluded that these subtests are appropriate in the early identification of AD, such as MCI.

When rating the methodology, the group sizes were adequate for statistical analyses although statistical differences were found between age and education (i.e. age increasing with severity of cognitive decline and education decreasing with severity of cognitive decline) which decreases the internal validity of the study and the findings. The large control group compared to the other two groups is also a concern for internal validity of the findings. Also, no information was given as to how the a priori diagnosis for MCI was accomplished although very detailed information was given for AD which makes interpretation of the finding more difficult since it is vital to understand the group inclusion criteria. It was mentioned that the *MMSE* and an ADL scale were used in the diagnosis and both were found to have significant differences between groups, but these tests are not as thorough for a diagnosis when used alone when compared with other studies that used a battery of tests. This study did not report any measures reliability, blinding or counterbalancing and for these reasons was rated as (-). This study does support the use of verbal fluency and memory (including verbal learning) to distinguish between MCI and normal cognitive ability although more research is needed specifically for the *WHO-BCAI*, but further more powerful research is recommended to fully support this test.

Screening tools

Standish, Molloy, Cunje, and Lewis (2007) looked at the discriminatory ability of the *AB Cognitive Screen (ABCS)* between MCI, AD and normal cognition. Six-hundred

forty-two participants were recruited from geriatric clinics and obtained a diagnosis of MCI with a subjective memory loss, no deficits of ADL, and no dementia. Controls were mostly friends or family members of the patients with MCI or dementia. The reference standard used in this study was the *SMMSE* (standardized *Mini Mental State Exam*), which was utilized due to the frequency of use along with the ability to support other findings suggesting that the *MMSE* is not a sensitive screening tool to MCI. The 2 subtests found to be significantly different from that of the controls' scores were the Delayed Recall and Verbal Fluency subtests of the *ABCS*. Another important finding of this study is that the *SMMSE* scores did not differ significantly between groups, suggesting the lack of power in distinguishing the subtle changes in MCI. ROC curve analyses were done, but sensitivity and specificity measures were not reported. Area under the curve (AUC) was reported and was highest on Verbal Fluency (0.73), but since sensitivity and specificity were not reported, these results are difficult to compare to other studies using the ROC curve analysis (see Table 11). The conclusions of this study suggested that the Delayed Recall and Verbal Fluency subtests on the *ABCS* give the best results in discriminating MCI from normal cognitive ability.

Judging this study, the methodology was relatively high-powered due to the large participant size of 642. Blinding of the test administrators was done, which improves the overall quality of the study, although no mention of reliability was reported. The participants of the study did have a slightly lower average education level of 11 years, which could impact external validity but according to statistical analyses did not correlate with either the *SMMSE* or the *ABCS*. , The use of family members as controls could have created bias, but this was mentioned as a concern by the authors of the article. Overall,

this well-controlled study was rated as a (+). The results of support the use of the Delayed Recall and Verbal Fluency subtests on the *ABCS* to identify MCI from normal cognitive, but the entire *ABCS* needs more research to determine sensitivity results before utilizing it clinically.

Borson, Scanlan, Watanabe, Tu, and Lessig (2006) compared physician recognition of a cognitive impairment compared to the diagnostic ability of the *Mini-Cog*. Two hundred and thirty-one participants were included and put into a control, MCI or AD group. Medical records from 199 physicians were reviewed for indication of a cognitive impairment from information including a diagnosis of cognitive impairment, prescription of anti-dementia medication or the administration of a cognitive screening procedure that yielded an abnormal result. The participant population was pulled from a larger study focusing on minorities in the Pacific Northwest region, which limits the generalizability of the findings to the greater population. The results showed that although the majority of the physicians were Caucasian, all minorities were represented in the sample of physicians. The study found that the *Mini-Cog* discriminated between normal cognitive ability and impaired cognitive ability significantly more than the general practitioner recognition. The results from the study concluded that a screening tool could be utilized by general care physicians to aid in the identification of mild cognitive impairment.

While rating the study, the reference standard, the *Mini-Cog*, screening procedure developed by Borson, Scanlan, Brush, Vitaliano and Dokmak (2000) showed excellent sensitivity and specificity when determining if the patient was demented or not. No mention of MCI was given in the validation reference, but was given a rating of

“adequately covered” in this SR, since the *Mini-Cog* was found to be sensitive to cognitive change in multicultural and multi-linguistic populations. This study did have internal validity strength in the methodology of administering all assessments and analyzing medical records simultaneously, giving the study a rating of a (+). Overall, the use of a standard screening protocol is recommended to increase identification of MCI in the general practitioner setting.

Cummings, Raman, Ernstrom, Salmon, and Ferris (2006) looked at behavioral changes including anxiety, irritability, apathy and depression (using the *Geriatric Depression Scale*) in patients with MCI versus normal controls. The study also determined if these factors could be more effectively assessed in the clinical, face-to-face setting or through distance evaluation (home using telephone contact). Although randomization of the 644 participants was utilized in the methodology in reference to being included in either the face-to-face vs. distance evaluation, the study was judged as a level 2 since the randomization was not in reference to the diagnostic groups. It has been found throughout this SR that randomization is not appropriate for diagnostic research. Using a multivariate logistic regression analysis, participants with MCI (determined by a *CDR* of 0.5) predicted changes in all measures of mood and behavior. No significant difference was found between the clinical assessment setting versus distance assessment via the telephone (aka telehealth), concluding that telehealth is a viable option in assessment. The authors of this study suggested that behavior and depression be included in a standard assessment of cognition to distinguish MCI from normal cognitive ability.

This study was high-powered considering the randomization and power analysis done a priori to determine the suitable number of participants to be 600. Test-retest

reliability of 0.64 was reported and the methodology also reported minority inclusion (1 for every 5 participants). Blinding was not addressed in this study, but the overall methodology was well-controlled. The study was given a rating of (+). Overall, the findings support the inclusion of a depression scale such as the *Geriatric Depression Scale* to help discriminate between MCI and normal cognitive ability.

Giaquinto and Parnetti (2006) conducted a two-part study that consisted of a validation study of 200 participants' performance on the *Basic Italian Cognitive Questionnaire (BICQ)* and then a randomized observational study of 963 participants who were administered the *BICQ* by their general practitioners. This questionnaire is quick and easy to administer, consisting of questions on orientation, personal/background information (i.e. age, date of birth, etc.), family information and ecological (i.e. hypothetical shopping situations). The preliminary study was done to obtain a cut-off score which was found to be adequate at 10 to determine if the patient had normal cognitive ability or if it had declined. The first study did not include a population of MCI so will not be further evaluated in this SR but the random trial of 963 participants did include a subgroup of MCI and will be further analyzed. Out of the 963 randomly selected patients given the *BICQ*, 130 obtained a score of 10 or less and were then further evaluated and put into groups of MCI, normal, or demented. Since the randomization stopped at the initial choosing of participants, it will not be judged as an RCT for the purposes of this SR since it did not randomize the diagnostic groups. The findings showed that the people who "failed" the screening procedure of the *BICQ* consisted of 40% cognitively normal, 33% MCI and 27% individuals with dementia. This is a large number of false positives although this is a quick screening procedure that is easy to

administer and states that it is just a sorting tool to grant further need for cognitive evaluation. According to the authors, the *BICQ* is appropriate for screening measures, but should not be solely relied on for a diagnosis of MCI.

When rating this study there was little control over the administration of the *BICQ*, since this was done by various general practitioners. The generalizability of the findings to other countries is in question since the study was completed in Italy with Italian-speaking individuals. The authors did randomize and employed the use of blinding. This study also included a large participant pool and was therefore rated as (+). The findings suggest that the *BICQ* is not the most powerful discriminatory tool but still useful in the general practice setting as a screening procedure for MCI.

Kirkpatrick, et al. (2006) investigated the diagnostic ability of olfaction with the use of the *University of Pennsylvania Smell Identification Test (UPSIT)* compared to the *Addenbrooke's Cognitive Examination (ACE)* to distinguish between MCI, AD and controls. The *UPSIT* is a 40-item identification task using four choices for each "scratch and sniff" item. Fifty-four community-dwelling volunteers were given both tests on the same day using randomization of the test order (counterbalancing), which took a total of one hour. Age and education had significant impacts on both tests, which partly could be due to the larger range of age (starting with age 45) and the higher educational average of the participants. When beginning this SR, the age-cut off was 55 as the lowest in the range, but was changed to the average age being 55 in order to include multiple studies such as this one with younger outliers. Gender was also a factor in the *ACE*, with males performing significantly better than females. These findings are interesting since the validation for using this particular standard for comparison was chosen since other

studies have found that these factors do not influence performance. Sensitivity and specificity were poor when using the *UPSIT*. Results from the stepwise regression determined that mint, chocolate, cheddar cheese and lime increased predictive value to 61%, the sensitivity to 57% and specificity to 97%. The authors concluded that olfaction is a viable measure to help discriminate between MCI and normal cognitive ability, although larger, more controlled studies are needed.

Rating the study concluded that since participants with possible dementia were excluded from analyses, the MCI group only included 7 participants which was quite low for statistical differences. Counterbalancing, blinding, and reliability were included in the study, but the small numbers and lack of control over multiple testing settings may have played a factor into the results. Differences were seen in the study between MCI and controls, determining that olfaction is a good diagnostic indicator of subtle cognitive decline. This sensitivity is fair, but could be improved in further studies with increased power, including more participants and more control over the conditions. This was an exploratory study (Level 3) and unable to rate, but the recommendations for larger studies in the area were given by the authors although the study determined that *UPSIT* is a valid tool in detecting MCI.

Eibenstein, et al. (2005) investigated the discriminatory power of the *Sniffin' Sticks Screening Test (SSST)* in the diagnosis of MCI versus normal cognitive ability. Overall performance of the *SSST* showed that the 29 participants with MCI had a significantly impaired sense of smell compared to the 29 participants in the control group. The MCI group was more unaware of their deficits than the controls that did have impairments in olfaction. Using the *MMSE* as the reference standard in comparison with

the results of the *SSST*, the MCI group with higher scores of the *MMSE* also had higher scores on the *SSST* and vice versa. Another comparison of both groups' scores on the 2 tests (*MMSE* and *SSST*) was significantly correlated using the Spearman rho correlation (see Table 10). According to this study, evaluation olfaction, specifically with the use of the *SSST* is appropriate in the identification of MCI.

One concern when judging this study is the heavy reliance on the *MMSE* alone when comparing results with the *SSST*. Many of the studies that have been included in this SR show evidence that the *MMSE* is not an accurate, detailed measure and should only be used to supplement a more in-depth evaluation of cognition. No mention of inter-judge reliability, test-retest reliability, or blinding was included in the article. This study received a rating of a (-). The results of the *SSST* in this article show that people with MCI do have a decrease in olfaction ability along with a lower awareness of their deficits, although these results should be interpreted cautiously.

Lam, Lui, Tam, & Chiu (2005) investigated the validity of a short, 5-item questionnaire focusing on subjective memory complaint to discriminate between MCI, AD and normal cognitive ability. The study included a total of 306 participants and used the *ADAS-cog* as a reference standard. The questionnaire being evaluated was adapted from the *Memory Inventory for the Chinese (MIC)* which consists of 27 questions. The investigators employed the questions that involved subjective memory complaint and formulated a 5-item test. A MemScore was computed according to the results of the 5 questions that included areas such as forgetting where objects are placed, inability to recall or follow a conversation and difficulty remembering the names of friends. The MCI group was split into a group scoring a 0.5 on 2 subscales of the *Clinical Dementia*

Rating (CDR) (memory and orientation) deemed “MCI-not demented (MCIND)” and a group scoring a 0.5 on 3 or more subscales of the *CDR* deemed “MCI-possible incipient dementia (MCIID).” This made possible the analysis of the heterogeneity of this population of MCI individuals to determine if certain memory complaints were indicative of a more serious cognitive impairment but not yet considered dementia. No significant differences were found between these two MCI groups, but the control group had a significantly lower amount of memory complaints than both groups. It was also found that specific questions involving the inability to follow and remember a conversation and subjective memory problems better discriminated the groups while questions involving misplacing objects and forgetting where they were had poor discriminatory ability. The results determined that using a questionnaire involving memory complaint questions is an appropriate screening tool that is highly correlated to the *ADAS-cog* total and delayed word recall scores. The sensitivity and specificity was fair for this test with a large amount of false positives being obtained with a cut-off score of 3 or more complaints.

When rating this study, the nature of the test being described was well described, an adequate participant size for statistical analysis was used and validated reference standards for comparison and a priori diagnoses were used. The external validity of this study was rated as poor due to the Chinese language used along with the low educational status of the participants. Although the investigators did state that a number of individuals with intact cognition will also have subjective memory complaints resulting in false positives if used exclusively, this is a screening tool and false positives are likely to occur. Due to many of the reasons listed above, this study was rated as (+) in this SR determining that the use of subjective memory questions in either a formal questionnaire

or interview is an appropriate screening procedure to determine MCI from normal cognitive ability.

Geda, et al. (2004) investigated the use of the *Neuropsychiatric Inventory (NPI)* to discriminate between controls, MCI and AD. The *NPI* is scale used to determine the types of neuropsychiatric behaviors (12 total) a patient is exhibiting and can be administered by a variety of professionals as illustrated in this study (i.e. a research nurse and psychometrist administered the *NPI*). A thorough neuropsychological exam was administered and the results taken to a consensus meeting for a priori diagnosis on participants who were recruited from the Mayo Alzheimer's Disease Patient Registry (ADPR) which follows a set protocol for referral into the registry. The *NPI* was administered to a large population of 655 and began by asking the patient if he/she exhibited any of these behaviors and if so, ranges of severity and frequency were determined in which a maximum score for each behavior was a 12 (frequency maximum score of 4 x severity maximum score of 3). The results from the statistical analysis showed significant differences between each behavior for all groups although when looking at the summary statistics, the MCI group exhibited a larger amount of night-time behaviors, irritability, anxiety and apathy than the control group. The authors suggested that these behaviors may be helpful additions to a case history to distinguish between MCI and normal cognitive ability.

When judging this study, the large number of participants may have accounted for why statistical differences were found in areas that were not as clinically significant as others, but these results do show that inquiring about psychiatric behaviors and using the *NPI* is helpful in the diagnosis of MCI. No mention of blinding was found in the study,

but the study was strong in the fact that it accounted for age and educational differences in groups in the analyses along with the use of a 4 hour neuropsychological evaluation for an a priori diagnosis. The use of the ADPR for patient selection helped increase the power of this study by controlling the participant selection which can be difficult when seeking specific populations. These factors all contributed to a rating of (+) for this study, suggesting that these behaviors may be helpful to include in a case history or screening protocol to distinguish between MCI and normal cognitive ability.

Artero and Ritchie (2003) developed a 5-10 minute screening protocol that could be utilized in the general practice setting such as copying an abstract figure, assessing verbal fluency, and recalling details from a story after a time delay to distinguish between controls and MCI. Activities of daily living (ADL) were also suggested to be included as a part of the assessment such as toileting, walking and using the phone. The reference standard used for the 368 participants was the *DECO* questionnaire. The tasks to be included in the protocol were determined by analyzing the results of multiple tasks from a large neuropsychological battery from a logistic regression model. The authors stated that the sensitivity of 73% and specificity of 99% was a model for how well the assessment tool predicted the progression to dementia.

Some issues that arose when rating the article included the authors' use of the term "cognitive impairment" throughout the study and then applying the results to the term "mild cognitive impairment." Although the discussion stated that the suggested screening tool could be used to distinguish between MCI and normal, the reference standard that was used to determine the groups at the beginning of the study was a questionnaire, the *DECO* (Ritchie, et al., 1992) that did not include Petersen's (1999)

criteria, and did not make an a priori diagnosis of MCI. No blinding was mentioned to prevent any bias and no reliability measures were included. The study was done in France with a group of people recruited from different general practices across a region, so the external validity was rated as good considering the use of multiple sites across a region to obtain participants. The external validity is lacking, though, in generalization from the French speaking to the English speaking population and the unknown socioeconomic status of the population. These reasons listed above yielded a rating of (-), suggesting that this screening tool be used cautiously until further, more controlled testing and research is done.

Computer-based assessments

Doniger, et al. (2006) evaluated the ability of the *Mindstreams* computerized test battery to discriminate between MCI, AD and normal cognitive ability without being affected by the presence of depression. One hundred and sixty participants were given the test battery that included subtests of memory, executive function, visual spatial, verbal, attention and motor skills, taking 45 minutes to complete. The *RAVLT* and *Clock Drawing Test* were also utilized as reference standards. This study developed 2 cohorts who were given either the *Geriatric Depression Scale* or *Cornell Scale of Depression for Dementia* to determine the impact of depression on the performance of *Mindstreams* along with the prevalence of depression in these populations. As severity of cognitive decline increased, the prevalence of depression significantly increased which shows the need for a test that is not influenced by depression. This computerized test yielded significant differences in performance of the MCI and controls groups on memory, executive function and verbal subtests for both cohorts (*Geriatric Depression Scale* or

Cornell Scale of Depression for Dementia). The visual spatial subtest showed significantly lower scores for the MCI group for one cohort, which was determined by the authors to be the most helpful subtest when discriminating normal cognition from subtle cognitive decline. The only subtest that was influenced by the presence of depression was motor skills in the *Mindstreams* computerized battery. Overall, the study supports the use of the *Mindstreams* computerized test battery in the evaluation of MCI.

When rating this study, the heterogeneity of the participants and the use of 2 different depression scales helped to improve the overall internal validity of the methodology. The inclusion of blinding and control for maturation in the methodology were judged as adequately covered. Validated reference standards of the *RAVLT* and *Clock Drawing Test (CDT)* were used in determining group membership a priori and the use of consensus diagnosis was employed. The only concern with these finding is the poor external validity due to the lack of information included on the ethnicity, socioeconomic status and language of the participants, especially since the study was done in multiple countries. Due to the reasons listed above, this study was rated as (+). Another study by Dwolatsky, et al. (2004) also found this battery to be an effective diagnostic tool and this study added more evidence that this is an appropriate test.

Gualtieri and Johnson (2005) investigated the ability of a computerized neurocognitive battery, *CNS Vital Signs*, to discriminate between MCI normal cognition and AD in elderly individuals. Previous research by the authors of the current study determined that the *CNS Vital Signs* battery had adequate reliability and concurrent validity. According to the authors, this battery can be given by anyone and is not specific to profession. This study included 178 participants that were given this user-friendly test

battery with subtests such as verbal and visual memory, finger tapping, symbol-digit coding, Stroop task, shifting attention and continuous performance. A battery of neuropsychological tests was also given as reference standards. The test scores are configured by the computer into 5 domain scores that include memory, complex attention, reaction time, cognitive flexibility and psychomotor speed. The results determined that 3 domain scores including memory, psychomotor speed (i.e. processing speed) and cognitive flexibility (i.e. executive function), distinguished MCI from controls with a sensitivity of 90% and a specificity of at least 75%. See Table 11 for more details. The single tests of symbol-digit coding and shifting attention also obtained good sensitivity and specificity measures while the verbal and visual memory subtests did not. These results vary from some studies that found that memory is the main deficit in this population of MCI; although this does concur with other studies that have found that other areas of cognition may be impaired and should be included in the evaluation and identification of MCI.

Rating this study found that the authors did include specific information on which neuropsychological measures were used to determine diagnostic group to make comparisons of this test other than the *MMSE*. No reliability or blinding measures were mentioned in the methodology. There was also a lack of participant information compared to other studies rated in this SR. These reasons gave this study a rating of (-) in the current SR, although the findings do suggest that this is a valid tool to use and that including other areas of cognition such as attention, processing speed and memory are important when assessing for subtle cognitive decline as found in MCI. More well

controlled research is needed in the area of non-memory measures of cognition in the identification of MCI.

Inoue, et al. (2005) researched the ability of a computerized test developed from the *MMSE* to distinguish MCI, normal cognitive ability and AD. One hundred and six participants were assigned to diagnostic groups from a neuropsychological evaluation and neuroimaging, although no specific diagnostic tools were reported in the article. The rationale for creating a test on the computer was to minimize administrator bias, although face-to-face interviewing is still appropriate and needed in a full diagnostic evaluation. According to the authors, the computer adds a non-bias assessment tool that could be utilized in a larger battery. The tests that showed significantly different scores for the MCI group were the age and year of birth validity, visual working memory with 3-D objects, a second delayed recall and the total score. The 3 word memory, time orientation, first delayed recall and visual working memory with 2-D objects resulted in non-significant differences in performance. Using ROC curve analyses, a sensitivity of 82% was found with a specificity of 87% which is quite good compared to many studies evaluated in this SR. See Table 11 for more details. The results from this study show great promise for a computerized method to screen patients for subtle cognitive decline.

When rating the study, reliability was not reported, although using a computerized test lowers the chance for any bias or mistakes, Technological errors may still occur, so reliability of the computer or any chances of technological failure should have been reported. The generalizability of the finding is in question as well, since the study was in Japan, the ethnicity, language, and socioeconomic status were not reported for participants and the overall age of the participants was quite high although the range was

very wide with young outliers in their 40's. The comparator test(s) are vital to the diagnostic ability of a new test, and the article did not describe which assessment tools were used. This could have a great impact on the results, and so further information is needed to adequately evaluate the results. All of the above mentioned factors, especially the lack of reporting of reference standards and the reliance on the *MMSE* to determine the areas to be tested on the computer, are reasons why this study was rated (-). Apart from the rating, the study does show validation of another computerized test procedure in the evaluation of MCI, although more research is needed to support its use.

Dwolatzky, et al. (2004) investigated the discriminatory ability of a computer-based diagnostic battery, *Mainstreams*, between controls, MCI and AD. Widely-used neuropsychological tests such as the *ADAS-cog* were used as comparison with the 98 participants from the Israel-based clinic, which resulted in similar or slightly lower discriminatory power than the computerized battery. The benefits of using a computerized system such as *Mindstreams* include the ability to increase or decrease in difficulty depending on performance therefore increasing sensitivity. This test battery also only reportedly takes 30-45 minutes to complete, and results are computed instantaneously, which is more efficient for time compared to many neuropsychological batteries. Also information can be obtained on reaction time, which could also be used to assess cognition. Using AUC as a measure of effect size, all cognitive domains tested using the *Mainstreams* system had a strong discriminatory ability with AUC higher than 0.800 (see Table 11).

When rating this study, it was found that the authors did blind the administrators to minimize bias and used a widely-used cognitive test battery as a reference standard

(*ADAS-cog*). No mention was given to reliability and there were between-group differences in age and education, therefore decreasing the study's internal validity. It was not stated if any participants dropped out or the procedure in how the various tests were administered (i.e. what order, all on the same day, etc.) and for this reason along with the concerns listed previously, this study was rated as a (-), suggesting more research is needed before use of *Mindstreams* in the evaluation of MCI.

Equipment-based measures

Grunwald, et al. (2002) studied 51 participants with EEG measures of theta-power during haptic tasks and during rest to distinguish between controls, MCI and AD. Haptic tasks are complex perceptual and cognitive tasks that involve the participant palpating an object with eyes closed to determine the shape of the object and then recreating the object's shape through drawing with eyes open. Two different measures were taken during this task aside from EEG measures to determine the ability of the task alone to distinguish diagnostic groups. These measures included exploration time (ET) needed to perceive the object and recreate and then quality of reproductions (QR) that the participants drew of their perceptions of the object. The EEG measure, theta-power, is an EEG power that is observed mostly during sleep and rest conditions using the standard international 10-20 system of electrode placement. The *MMSE* was the only reference standard used in the study. The results found that theta-power decreased in all participant groups significantly during the haptic task involving more cognitive power than at rest. The only significant difference between the MCI and control group with EEG measures during the haptic task was an increase in theta power over the right occipital regions (O2) for MCI that was not present in controls. Analysis of the haptic tasks measures alone

revealed that the quality of reproductions (QR), which did include inter-judge reliability to account for subjectivity of the rating, was significantly lower in MCI vs. controls, although ET did not differ significantly. Poor correlation between EEG and the *MMSE* determined that EEG theta-power measures are not appropriate for diagnosis. The authors concluded that EEG reveals the differences in brain activity between participant groups, but haptic tasks are good indicators of subtle cognitive change due to the complexity of the task, although the sensitivity of the task still needs to be evaluated.

The rating of the methodology found that the study did include reliability for the haptic task but not EEG electrode placement, which was reported in the study to increase bias or error. The participant group was on average older than many other studies and no ethnicity or language information was reported. See Table 8 for more details. Although no language information was reported for participants, the tasks involved did not require language, which improves the generalizability of the findings to any language population. The groups were small and if larger may have revealed more differences, but the study did account for multiple comparisons by setting an alpha level that was Bonferroni adjusted. Another important factor to analyze was the use of the *MMSE* as the only reference standard to compare the results of EEG and other tasks. For these reasons listed above, this study was rated as a (-), suggesting that haptic tasks may need more investigation to determine the accuracy of diagnosis. According to these results, EEG measures are not appropriate for discrimination between these participant populations, although this may require more high-powered research as well.

Miscellaneous assessment tools

Dixon, et al. (2007) looked at the ability of two neurocognitive markers: 1) inconsistency (intraindividual variability) and 2) speed of reaction to discriminate between normal cognitive ability and MCI. Two studies were reported in the article, with the second being an extension and validation of the first. This SR will focus on the second study which took a cognitive battery, the Project *Mental Inconsistency in Normal Dementia (MIND)* battery, given 5 times over the course of 3 months to 304 participants. The computerized *MIND* battery included 3 reaction time tasks: 1) Simple reaction time, 2) Choice reaction time and 3) Choice reaction time one-back. The Choice reaction time one-back is the most complex, asking the participant to choose the response from the previous stimuli. Multiple cognitive domains were assessed a priori for group placement with tests such as verbal fluency, semantic memory, digit substitution and word list free recall. This study used two groups of MCI, a mild group performing below the mean on only one cognitive domain, and a moderate group, scoring below the mean on 2-5 cognitive tests. The results showed that through logistic regression analyses, when the 2 neurocognitive markers (intraindividual variability and speed of response) were combined, speed of response did not account for the variation as much as intraindividual variability over 5 sessions in 3 months. This result was seen consistently for the MCI-moderate group on all subtests of the *MIND*, but only on the most cognitively complex subtest (choice reaction time one-back) of the *MIND* for the MCI-mild group in comparison to the control group. The results determined that neurocognitive markers such as intraindividual variability and speed of response are appropriate measures to use in the

identification of MCI, but if the cognitive decline is extremely mild, more complex tasks may be needed.

When rating this study, reference standards including multiple, validated tests (see Table 7) were used. The methodology was well-controlled, giving specific times for reassessment and control of testing environment. An adequate number of individuals for statistical comparisons were also used in the study, making the results more accurate and precise. The rating of this well-controlled study is a (+) due to the reasons listed above. This article supports the use of intraindividual variability when discriminating between MCI and normal cognitive ability, especially when the MCI may involve deficits in multiple cognitive domains.

Wylie, Ridderinkhof, Eckerle and Manning (2007) investigated response inhibition using a computerized flanker task to discriminate between 20 individuals with MCI and 20 individuals with normal cognitive ability. The flanker task consisted of 309 experimental trials that randomly would display an image of a target arrow (in the middle of the screen) along with 4 distractors (either arrows pointing in the same or opposite direction or a neutral distractor such as a diamond) on a computer screen. The participant would be advised to push one of two buttons (right/left) depending on the direction the target arrow was pointing (i.e. pushing the right button if the target arrow was pointing right). This task is similar to the Stroop task except it does not require the ability to read. The Stroop test, along with the *CVLT* and *RBANS* were used as reference standards for the flanker task. The results showed that the incongruent flanker condition (i.e. flanker arrows pointing in the opposite direction than target) increased reaction time in both groups but no significance was found between groups. The results from this study

determined that the Stroop task did not accurately distinguish between groups. The use of slope analysis of delta plots did show that as response time increased, the MCI group had more difficulty with response inhibition of the flankers. The authors determined that the flanker task did show differences in groups, although the differences were small. The results suggest that response inhibition may not always be an accurate indicator of mild cognitive decline.

When judging this study, validated reference standards including the *CVLT* and *RBANS* were used. The participant groups were also similar with the exception of cognitive decline, increasing the control of the study. The results determined that the groups performed similarly on the flanker task, although small differences were found. The results may be due to the small number of participants in this study and more significant findings may have occurred with a larger participant size. This study was given a rating of (+) even though the results may have been stronger with a larger sample size due to the control and use of appropriate reference standards. The use of flanker tasks in the evaluation of cognition with this population will require more extensive research and should be used cautiously since this study also reported the non-significant findings of the Stroop task.

Bonney, et al. (2006) investigated the ability of inspection time (IT) in discriminating between 28 participants with MCI and 28 participants with normal cognitive ability. IT is defined as the speed of information taken in and is different than reaction time since it is not related to how fast the person reacts but how long the person needs the stimulus to make the right judgment. A group of neuropsychological tests were given as reference standards such as the *MMSE*, *CVLT* and clock drawing test. The task

was computer-based and the participants were given visual stimuli at brief intervals and asked to discriminate if they were the same or different. The benefits of using a task that uses shapes as stimuli are the objectivity in the results and lack of educational bias. Statistical analysis determined that participants with MCI had a significantly longer IT than controls. Several neuropsychological tests were moderately correlated to the IT scores, although the regression model showed that these tests only accounted for 35% of the variation. The findings suggest that IT is a viable option in the diagnosis of MCI, although the authors determined that this should be researched further.

When judging this study small participant size and small group numbers could have affected statistical results, especially when comparing the reference standards. No numbers were given, but the discussion of the article mentioned the significant overlap of MCI and controls' performance on the IT task, determining that more in-depth studies need to be carried out to determine sensitivity of this tool. No mention of blinding or reliability was given, decreasing the overall quality of the methodology, although the 2 groups were well matched for age, gender and education, increasing internal validity. Overall, the study was well-controlled giving the overall methodology of the study a rating of (+). Larger studies do need to be carried out to support the use of IT to identify MCI.

Ribeiro, de Mendonca and Guerreiro (2006) investigated other areas of cognition affected in individuals with MCI versus normal cognitive ability as assessed by the *Battery of Lisbon for the Assessment of Dementia*. Using this battery to assess multiple cognitive domains in 179 Portuguese participants, the authors found that not only was memory significantly impaired in the MCI group, but also performance on semantic

fluency tasks, motor and graphomotor initiative, calculation and *The Token Test* which tests complex language ability. Another interesting finding of this study is that 52.6% of individuals had a cognitive domain other than memory that was as severely impaired than memory. No predictive analyses (ROC curve analysis, logistical regression analysis, etc.) were done to determine the extent of diagnostic ability of these tests, but it was determined by the authors that memory should not be the only area assessed in MCI. This is a change from the original theories of MCI being a memory impairment only, which is becoming more common throughout literature as the MCI population is being found to be heterogeneous.

Rating of this study found that the authors chose participants randomly from Lisbon, increasing the quality assessment of the methodology. The *WMS* Logical Memory test, an adequate reference standard, was used for a priori diagnosis of the participants to ensure accurate comparison. The generalizability of the findings was poor due to the language, low educational status and lack of SES information. Overall the methodology was well controlled, giving this study a rating of (+). This study supports the inclusion of cognitive domains other than memory when discriminating between MCI and normal cognitive ability.

Ritter, Despres, Monsch and Manning (2006) found that the *Topographical Recognition Memory Test (TRMT)* is a useful tool in distinguishing MCI from normal cognitive ability in the aging population. Forty-five participants from Switzerland were included in this study. Subgroups of MCI and normal were formed as depression/no depression to determine if topographical recognition was sensitive to depression. This study analyzed the results of the *TRMT* with multiple ANOVAs and Newman-Keuls post

hoc analyses and found that both subgroups of MCI (depression/no depression) performed significantly worse than the normal controls. The results found no significant difference in the performance of topographical recognition in the depression/no depression subgroups. The authors determined that depression did not play a factor in the results of topographical recognition, yet cognitive impairment did and therefore is an appropriate measure to help discriminate between MCI and normal cognitive ability.

Judging the study limitations were found, including no mention of blinding of test administrators or reliability measures. Also, the participants were French-speaking from Switzerland with no mention of socioeconomic status or ethnicity which decreases the generalizability of the finds of this study. The participant group was also quite younger than many of the other studies being analyzed in this SR, but education is similar to most other studies. Overall ratings including appropriate reference standards and methodological control gave this study a (+), suggesting that the *TRMT* may be appropriate for discriminating subtle cognitive decline, although it would be wise to research this test on other populations.

von Gunten, et al. (2006) investigated the validity of the *Protocole d'Examen Cognitif de la Personne Agee-Lausanne (PECPA)* cognitive test to discriminate between 237 normal cognitive individuals, 115 individuals with dementia and a convenience sample of 27 individuals with MCI. This assessment tool was developed from the *MMSE* which was also the test used for a priori diagnosis. The developers developed the questions for the *PECPA* with more detail to assess ten cognitive abilities which include temporal orientation, spatial orientation, attention/calculation, immediate recall, language, remote (long-term) memory, judgment/abstraction, gnosis, praxis and delayed

recall. This test was specifically developed for the French-speaking population in Switzerland from where the 379 participants were recruited. The main purpose of the study was to determine if the *PECPA* could discriminate between dementia and normal cognitive ability with a small MCI group included as well. Good discriminatory ability was found for dementia but was less powerful with MCI, with an AUC of 0.769 (see Table 11). The investigators also stated that there may have been some overlap of MCI individuals in the control group due to inclusion/exclusion criteria. The study concluded that the *PECPA* is an appropriate test when distinguishing MCI from normal cognition, but more research is needed on the MCI population.

When judging this study, the MCI group was significantly smaller than the rest and weaker results were found for discrimination of this population. Also, the investigators mainly used the *MMSE* as a reference standard and a template to create their test, which has been found to be a lower-powered assessment tool for MCI. External validity was also poor due to the language of the test and lack of specific educational, gender and ethnicity information in the article. This study received a rating of (-) in this SR due to the reasons listed above. This battery needs further evaluation of validity of discrimination of MCI versus normal cognitive ability.

Leritz (2004) investigated the ability of associative priming of novel and semantic word pairs to discriminate between 18 individuals with MCI and 18 individuals with normal cognitive ability. Semantic word pairs were related (i.e. nurse-doctor) and were therefore hypothesized to be easier to prime (remember over trials although not asked to directly to remember) than novel word pairs (i.e. captain-snail). It was also hypothesized that priming scores of novel word pairs would be worse for MCI participants although

this was not found to be true in the results of this study. Although differences were found between reaction time and priming condition (intact or recombined), no differences in any analysis were found between the MCI and control groups, suggesting that people with MCI can form novel associations to increase memory ability even when this is tested indirectly (i.e. priming task). The multitude of neuropsychological tests used as reference standards for diagnosis and comparison showed strong diagnostic results with a composite score of 9 tests including the *HVLT*, *Wechsler Memory Scale III (WMS-III)* Logical Memory I and II, and *Brief Visuospatial Memory Test (BVMT)* that was significantly lower for the MCI group than the control group. The results found that individuals with MCI can form novel memory associations, which was unexpected according to the hypothesis. Priming with novel and semantic word pairs was determined to not be an accurate diagnostic tool, although larger studies are needed to evaluate priming as a diagnostic tool for MCI.

The rating of the methodology of the study determined that it was very well controlled including blinding of administrators, reliability of diagnosis using a consensus of 7 professionals and counterbalancing of the word lists during the priming task, Although the participant numbers were somewhat small, the study used an a priori effect size calculation that called for that number for a medium effect size. More statistical significance for novel versus semantic priming may have been found with larger group numbers. The group numbers were equal, though, which does make comparison more accurate. These reasons listed above are the reasons that this study obtained an overall rating of (+). The results do suggest that using priming with novel and semantic word

pairs may not be the most powerful diagnostic tool in discriminating between the subtle decline in MCI from normal cognitive ability.

McCoy (2004) studied the intraindividual variability (IIV) of 15 participants with MCI and 53 participants with normal cognition on various memory and non-memory tasks. These tasks were administered daily over a 31-day period to determine if IIV could accurately distinguish between diagnostic groups. A full neuropsychological exam was given to each participant to assign group membership and then the *Daily Cognitive Assessment Battery (DCAB)* was given by a research partner (usually a spouse, family member, neighbor, etc.) to the participant. This 10-20 minute test that was to be given every day for 31 days consisted of a digit span task (modified from *WMS-III*), symbol digit substitution and number copy, list learning (modified from the *RAVLT*), a sleep diary, positive and negative affect scale, environmental distractions and a stress ladder. Since these were to be given on a daily basis, multiple variations were made for each task and counterbalanced across participants. IIV is defined as a fluctuation in an individual's performance on a task for a short time and was calculated as the average amount of individual variability around a best fitting regression line which was termed as the IIV residual index (IRI). The hypotheses suggested that IIV would be greater for MCI participants than controls, but the results showed varying results. On the *RAVLT* List 1, the controls had greater IIV than the MCI group, while on *RAVLT* Percent Retained the MCI group had greater IIV than controls. No other significant findings were obtained and no pattern was found between IIV and cognitive function. Another interesting finding showed that the *RAVLT* alone classified 85% of participants correctly but when IIV was added, the correct classification dropped slightly to 84%. These findings showed that IIV

may not be a factor that is influenced enough by the subtle cognitive decline of MCI to be an accurate diagnostic indicator although more powerful research is needed.

Rating this study found the MCI group quite small and with a larger subgroup, more significant differences may be found with IIV. An interesting finding showed that the *Boston Naming Test (BNT)* score, which was used as a reference standard, was significantly worse for the MCI group compared to the control group. This is different from most studies that have found the *BNT* to not discriminate between the two, but may require future research using different levels of education since this participant group was fairly highly educated. This study was very controlled and included counterbalancing, reliability, compliance monitoring and the use of a consensus diagnosis for a priori diagnosis, giving this study a rating of (+), although no significant results were found for IIV as a discriminating factor between MCI and normal cognitive ability.

Chapman, et al. (2002) investigated gist level (i.e. the overall point) and detail level recall of a paragraph with a time delay in distinguishing between MCI, normal, and AD. The *MMSE* along with other test batteries were used as reference standards. Sixty-nine participants were included in the study. This cohort study studied three probes for gist level processing of a narrative biography of 578 words: 1) giving the main idea of the story, 2) determining the lesson and 3) giving a summary. Four raters judged the discourse sample and inter-judge reliability was reported on a random selection of 20% of the sample that was analyzed by a 5th rater to insure accurate coding and scoring. The results determined that participants with MCI performed significantly worse than the controls on all gist level probes. Also, significant differences were found on the 2 measures of detail level recall: recalling of details and recognition of details. Significant

positive correlations between gist and detail level recall and the MMSE scores were found. The gist level score did diagnose 5 MCI participants out of 25 participants as having AD who performed above the MMSE cutoff for AD during a priori diagnosis. Although sensitivity and specificity measures were calculated for gist and detail level recall, it was done so for AD vs. no AD. On the lesson task (gist level processing) a sensitivity of 92% and a specificity of 76% were found. On the main idea task (gist level processing), a sensitivity of 92% and a specificity of 96% were found. Inter-judge reliability of the coding was 94% for gist level and 95% on detail level using point-by-point agreement.

When rating this study, it was found that although the judging of discourse recall was somewhat subjective, the methodology did include point-by-point agreement which was reported and found to be fairly accurate. The study also controlled well for extraneous variables and used a wide selection of tests for reference standards. A few external validity concerns of the study include the small sample size of 69 and the use of the test in a population of lower educational status where reading level may be lower; especially considering that the assessment included having the individual read the passage silently. The fact that the *MMSE* diagnosed 5 individuals as MCI that the discourse recall tasks diagnosed as AD illustrates the variability of the *MMSE* and reinforces the fact that the *MMSE* should only be a part of the diagnostic battery and not solely relied on for a diagnosis. Overall, this study's methodology was rated as a (+) for this SR, suggesting that discourse recall may be included in an evaluation of MCI.

Barbeau, et al., (2004) evaluated the diagnostic ability of a visual recognition memory task, the *DMS48*, created by the investigators, compared to a verbal memory

test, the *Free and Cued Selective Reminding* test (*FCSR*). Twenty-three MCI patients were recruited from memory clinics in France to participate, along with 40 controls and 50 individuals with AD and Parkinson's disease (PD). The *DMS48* evaluated immediate and delayed recall of visual stimuli with distractors (using a forced choice selection of original pictures shown) that used both concrete and abstract figures. The results determined that the patients with MCI performed significantly worse than the controls on the *DMS48* and the scores of the *FCSR* were significantly correlated. There was much variability in the MCI group on both tests, with some participants performing as normal on one and impaired on the other and vice versa, but 83% of the participants were put into the same category on both tests.

When rating this study, it was found that sensitivity and specificity measures were reported for all diagnostic categories except MCI, which leads the investigator to infer that the sensitivity of the test to MCI may not be adequate. No reliability or blinding was reported for the study, decreasing the quality of the methodology. The small numbers of participants within each group, especially the MCI group, is of concern since more statistical power could have been added to the study with an increase in participants, and therefore better sensitivity of the test to MCI. This study did mention that a benefit of using visual memory rather than verbal memory is the ability of the test to be used cross-culturally and cross-linguistically, which increases this study's external validity even though the study used French-speaking participants. Overall, the quality assessment rated this study as a (-), suggesting that more research is needed to support the use of visual memory as a distinguishing factor between MCI and normal cognitive ability.

Summary Tables

The tables in this SR compiled various aspects of each article in order to make quick comparisons across studies. Table 7 describes each study's test in question, the reference standard used as comparison and inclusion and exclusion criteria for participants is included in this SR. Important parameters to consider when evaluating this table include the reference standard used as well as specific inclusion and exclusion criteria. Many studies gave explicit criteria while others were non-specific, which is important to consider when evaluating a studies methodology. Also, when comparing across studies, it is important to consider that many studies used the Mayo clinic criteria for MCI, while others used different criteria. Reference standard comparison is also interesting considering some studies used multiple, detailed tests for reference while others used only screening measures such as the *MMSE*.

Table 8 compares study and participant information for each study. Aspects of this table include participant numbers, where the study was done and in what language the participants were tested as well as participant information such as gender, age, education and how the participants were recruited. It is important to note that most studies, the average participant age was in the range of 60-upper 70's. Also, a few studies reported differences between groups in age, education or gender, which is illustrated by asterisks. Also, significant differences between groups lower validity of the study, which is reported in each descriptive analysis. Another important aspect when evaluating these tables is the average education level. Many participants had at least a high school education (12 years), although a few studies were very high and a few were very low, which is also noted in each descriptive paragraph.

The results were split into three tables, Table 9, a parametric statistics table, Table 10, a non-parametric statistics table and Table 11 that compares the sensitivity, specificity and area under the curve of various diagnostic tests in different studies. For the parametric and nonparametric statistics tables (Tables 9 and 10, respectively), different statistics were used in each study, so comparison between each is difficult. The significance level for each was reported as well as the type of statistical analysis used (i.e. ANOVA, t-test, etc.). Along with these tables as well as the descriptive paragraphs, each study is reported by the major parameters involved and the evaluation of the methodology of each.

Table 11, comparing the sensitivity and specificity of each study is easier to compare across studies, with higher sensitivity and specificity signifying better diagnostic ability. Only tests that had sensitivity and specificity measures were included in this table. Area under the curve, also reported in Table 11, is a measure of effect size. When interpreting effect size, it is widely accepted that 0.2 is a small effect size, 0.5 is a medium effect size and 0.8 is a large effect size (Cohen, 1992). This can therefore be used to interpret area under the curve measures, with higher effect size signifying better diagnostic ability. When interpreting sensitivity and specificity measures, it is not only important for each measure to be as high as possible, but also that the two are balanced for the best sensitivity and specificity that can be reached. These two measures are often viewed as a trade-off, and that it is important to have a balance of each. For example, having a sensitivity of 98% with a specificity of 40% is not as diagnostically powerful as having a balanced sensitivity of 89% and a specificity of 90%. These factors are all important to consider when comparing the results of various studies.

Table 7. Test in question, reference standard, inclusion and exclusion criteria.

Author	QA	Test	Reference Standard	Inclusion	Exclusion
Dixon, et al., 2007	2+	Simple RT, Choice RT, Choice RT one-back	Cognitive reference tasks: Digit Symbol Substitution, Lesster Series, WL free recall, VF, SM	MCI-mild: 1 SD or more below mean on 1 test; MCI-mod: 1 SD or more below mean on 2-5 tests	Major medical illness, sensory impairment, A/SA, inpt. psychiatric tx, MMSE< 24, ESL
Standish, et al., 2007	2+	ABCS: orientation, registration, visuospatial, ST, verbal memory, VF	sMMSE	English speaking MCI: SMC, no loss of ADL, no dementia	GDS >7 younger than 55
Wylie, et al., 2007	2+	Eriksen flanker task	CES-D, AMNART (for IQ), MMSE, CVLT, RBANS, Stroop-Color Word Test	MCI: MCC	Hx of stroke, untreated mood disorder, hx bipolar disorder, schizophrenia, PD
Bonney, et al., 2006	2+	Inspection Time	BDR, CDR, MMSE, CDT, CVLT	MCI: MCC: CVLT & CDR of 0.5)	Hx of stroke, MMSE < 24, geriatric depression scale ≥ 6
Borson, et al., 2006	2+	Psycian recognition (from medical records)	Mini-Cog	NS	Motor/sensory impairment, no primary doctor, no/fragmentary outpatient records
Cummings, et al., 2006	2+	Behavioral/mood changes: irritability, apathy, anxiety (NPI) 5 item Geriatric Depression Scale (GDS) 4 word delayed recall	CDR, mMMSE, Free and Cued Recall Selective Reminding Test	Research partner MCI: CDR= 0.5 mMMSE \geq 88 with 8 years ed., mMMSE=80 if less 8 years ed., Free Cued Recall > 44	A/SA, hx mental retardation, active/major PD, Use of antipsychotic, Parkinson's or dementia drugs, over 75 years old, dx of dementia, poor health
Doniger, et al., 2006	2+	Mindstreams: memory, executive function, visuospatial, verbal fluency, attention, information processing, motor skills	RAVLT & CDT (Global Depression Scale and Cornell Scale for Depression in Dementia given in 2 cohorts)	MCI: MCC Normals: no cog decline, AD-DSM IV	Hx of depression, ND/PD other than AD, colorblind, previous testing with Mindstream, program not in primary language

Table 7. Test in question, reference standard, inclusion and exclusion criteria.

Author	QA	Test	Reference Standard	Inclusion	Exclusion
Giaquinto & Parnetti, 2006	2+	Basic Italian Cognitive Questionnaire (BICQ)	MMSE, neuropsychological evaluation (NS)	Active life	Hx of ND, psycho-active drugs
Greenaway, et al., 2006	2+	CVLT	DRS, MMSE, BNT, FAS, Animal Fluency, TMT, WAIS-III/WAIS-R	Controls: ≥ 50 yrs. MCI: MCC	ND/PD, ESL, left handed
Kirkpatrick, et al., 2006	3	University of Pennsylvania Smell Identification Test (UPSIT)	Addenbrooke's Cognitive Examination (ACE) includes MMSE, along with other cognitive tasks	MCI: normal cog function, ACE=83, DR on ACE > 1 SD from the mean	ACE<84, VL/OM=AD, nasal infection, hx smoking, stroke, diabetes, brain tumor, head trauma, Parkinsonism, topical nasal vasoconstrictors, cocaine, Nifedipine, Cancer meds
Mioshi, et al., 2006	2+	ACE-R: domains of: pattention/orientation, memory, fluency, language, visuospatial	CDR	Could perform assessment, CDR in the last 90 days MCI: MCC	PD, concomitant dementia process, causes of cognitive impairment other than ND
Pernecky, et al., 2006	2+	ADAS-MCI-ADL (interview: assess impairment in everyday living) & ADAS-cog	CERAD-NAB German (including MMSE), B-ADL, WMS-LM, TMT A & B, CDT, IQCODE, CDR, neuro eval, lab tests, MRI	CDR 0.5, normal ADL (can if more complex), no SMC, mem in tact if other cog domian affected	Dx criteria for dementia met, CDR 1 or higher, clinically sig. psychiatric/neurological disease
Riberio, et al., 2006	2+	Battery of Lisbon for the Assessment of Dementia	WMS-LM, MMSE, BDS	NS	NS
Ritter, et al., 2006	2+	Topographical Recognition Memory Test (TRMT)	MMSE, verbal/nonverbal reasoning, verbal IQ, phonological fluency, recall, IADL	MMSE ≥ 26 absence of ND/PD other than depression	Heart attack, fainting fits, hypoxia, prolonged headaches, severe general illnesses, antidepressants

Table 7. Test in question, reference standard, inclusion and exclusion criteria.

Author	QA	Test	Reference Standard	Inclusion	Exclusion
von Guten , et al., 2006	2-	PECPA-L: Protocole d'Examen Cognitif de la Personne Agee-Lausanne (A Cognitive Assessment Tool for the French-Speaking Elderly in Switzerland)	MMSE	Controls: MMSE \geq 24, independent living, French first language, normal hearing/vision MCI: MCC	Controls: disturbing mem impairment, previous/ongoing ND/PD, brain injury, psychoactive/anticholinergic drugs (high doses), GDS(depression) > 12
Chandler, et al., 2005	2+	CERAD total score (Sum of subtest scores)	MMSE BDRS CERAD	50 yrs or older English speaking Controls: CDR: 0 MCI: MCC AD: NINCDS/ADRDA & CDR: 1	No minorities Co-morbid conditions affecting cognition Institutionalized
Eibenstein, et al., 2005	2-	Sniffin' Sticks Screening Test (SSST)	MMSE, MDB, neuropsychological battery (Specific tests NS)	MCI: MMSE>24; CDR 0-0.5; GDS>6; SMC	ND/PD, head trauma, COPD, maxillofacial surgery, pathologies of nasal sinuses, asthma, hepatitis, cirrhosis, chronic renal failure, vitamin B12 deficiency, A/SA, CVA, diabetes, hypothyroidism, Cushing syndrome
Gualtieri & Johnson, 2005	2-	CNS Vital Signs: verbal/visual mem, finger tapping, Stroop symbol-digit coding, SA, continuous performance Domain scores: Mem, PM speed, RT, Cognitive flexibility & Complex attention	MMSE Neuropsychological Tests: (Specific tests NS)	MCI: new onset or progressive cog impairment, mild deficits on neuropsychological tests, no impaired ADL; MMSE > 24	Controls: medical, ND/PD psychoactive drugs, impairment of ADL
Inoue, et al., 2005	2-	Computerized screening test: age & year of birth validation, 3 word mem test, time orientation, 1st modified DR, visual WM, 2nd modified DR	NS	MCI: MCC AD: DSM IIIR/NINCDS-ADRDA	

Table 7. Test in question, reference standard, inclusion and exclusion criteria.

Author	QA	Test	Reference Standard	Inclusion	Exclusion
Karrasch, et al., 2005	2-	CERAD: WLL, WL delayed recall, VF, Naming, MMSE, Constructional praxis, CDT	WMS-R, WAIS-R, TMT A & B, Benton Visual Retention Test, BNT	NS	Controls: no head trauma, depression, or SMC
Lam, et al., 2005	2+	Abbreviated Memory Inventory for the Chinese (MIC) resulting in MemScore	CMMSE, Digit span, Category Verbal Fluency Test, ADAS-Cog	CDR < 2	Hx of ND/PD, major depressive episode
Nordund, et al., 2005	2-	Speed/attn: Digit symbol, TMT A/B, Digit Span Mem/learning: RAVLT DR, LM: DR, Rey complex figure DR, Face recognition Visuospatial: VOSP silhouettes, Rey complex figure copy, Block design Lang: TTT, ASLD rep., BNT, Similarities, FAS word fluency EF: PaSMO, Dual task, WCST-CV64 correct, Stroop, Pic word test	STEP (stepwise comparative status analysis), I-Flex interview MMSE, CDR	Controls: physically/mentally healthy, no cog impairment	For MCI-without positive outcomes on all ref. standard measures
Woodard, et al., 2005	2++	Dx algorithm: CERAD, MMSE, CDT	RAVLT & MDRS	Independently living: MCI: 1 SD below mean on RAVLT, not > 2 SD MDRS	ESL, viusal/auditory acuity impaired, PD, delirium defined by DSM-IV
Barbeau, et al., 2004	2-	DMS58: visual recognition memory task	FCSR (Free and Cued Selective Reminding) test	MCI: ADL normal, WAIS-III, FCSR \geq 1.5 SD below mean Controls: MMSE \geq 27	MCI: deficit in one or more cog. domains other than memory
Dwolatzky, et al., 2004	2-	Mindstreams: memory, executive function, visuospatial, verbal fluency, attention, information processing, motor skills	MMSE & ADAS-cog Participants from Israel: WAIS-III subtests Digit Symbol & Block design, WMS LM & Mental Control, RAVLT, CDT, TMT-A, BNT	English-speaking MCI: MCC AD: DMS IV criteria	Hx of PD Major depression ND

Table 7. Test in question, reference standard, inclusion and exclusion criteria.

Author	QA	Test	Reference Standard	Inclusion	Exclusion
Geda, et al., 2004	2+	Neuropsychiatric Inventory (NPI) scale	4 hour neuropsychological battery, MMSE & DRS	MCI: MCC AD: NINCDS/ADRDA Controls: no CC, function independently	Active ND/PD, psychoactive drugs, comorbid disease interfering with cognitive ability
Knox, 2004	2++	Dx algorithm: DRS, CVLT, BNT, Verbal Fluency, WMS-R/III, TMT-B	MMSE, WAIS-R	MCI: MCC AD: NINCDS/ADRDA	NS
Leritz, 2004	2++	Novel & Semantic Associative Priming (words from Nelson Norms of Free Association)	MMSE, WASI, NART, HVLT, BVMT, WMS-R LM I & II & Digit Span, TMT A & B, COWA, VF, CDR, MAC-Q, GDS	Age 60 or older	Hx of ND, CHI with loss consciousness, A/SA, PD requiring hospitalization, heart attack past 6 months
Lopez, 2004	2++	Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) & Automated Neuropsychological Assessment Metrics (ANAM)	MDRS Long term % retention (RAVLT) Lawton's IADL	MCI: 1 SD below on RAVLT, normal on MDRS & IADL Controls: >16th percentile all 3	ESL, limited visual/hearing acuity, PD, delirium, prior dx of MCI or dementia, unable to live independently due to disability
McCoy, 2004	2+	Intraindividual Variability (IIV) on measures of: attention processing speed, working & episodic mem (daily over 31 days)	MMSE, TMT A-B, BNT, HVLT, COWA, GDS, CDR, BDS, North American Adult Reading Test, RPR, Brief Visuospatial Memory Test-Revised, CES-D, Rey-Osterrieth Complex Figure	MCI: mem deficit >1.5 SD on HVLT, SMC, CDR<1, MMSE>23, <1 SD of non-mem, normal ADL Controls: within 1 SD on memory (HVLT)	Severe dementia, hx CHI w/ loss consciousness, ND/PD, uncorrected vision/hearing, A/SA, telephone interview of cognitive status (TICS) < 30
Mejia, et al., 2004	2-	MMSE (Spanish), Brief Neuropsychological Test Battery, Short Blessed Test Pfeffer Functional Activities Questionnaire, BDS	NS	MCI: MCC Controls: independent living & no cog impairment, normal neurologic exam	Controls: ND/PD, use of psychoactive drugs
Artero & Ritchie, 2003	2-	ECO & ECA ADL	NS	Cog impairment grp: CC Controls: no CC	Not meeting DSM IIR dx for senile dementia

Table 7. Test in question, reference standard, inclusion and exclusion criteria.

Author	QA	Test	Reference Standard	Inclusion	Exclusion
de Jagar, et al., 2003	2+	HVLT, TPT, RPR, BNT, CF from CERAD, TTT, Pattern & Letter Comp, Letter Canc A & B, Map Search from TEA, Incomplete letters & drawings, CLOX, Bisecting lines, Spatial rotation	Cambridge Examination for Mental Disorders of the Elderly (CAMDEX) informant interview & MMSE	Controls: MMSE >24 No SMC	NS
Estevez-Gonzalez, et al., 2003	2+	RAVLT: correct responses immediate recall, verbal learning, verbal forgetting, learning curve % of forgetting	MMSE IQCODE ADL-BDRS, GDSS	MCI: MCC; Controls: no SMC, no cog/mem impairment, MMSE \geq 27, normal social function	Marked ND/PD, Did not meet criteria for controls, DAT, MCI
Chapman, et al., 2002	2+	Gist & Detail-level Processing	MMSE, CDR, WMSII Hachinski Scale, HDRS, NINCDS/ADRDA	English first language	Hx of head injury w/ loss of consciousness, ND other than AD, major depression in last 2 years, A/SA
Grunwald, et al., 2002	2-	EEG: haptic task & rest	MMSE	Ages 75-85, Right handed MMSE > 18	ND/PD, neuroleptic/anti-depressive drugs for 6 weeks
Pasqualetti, et al., 2002	2+	MMSE	MDB	SMC	MMSE < 9
Xiao, et al., 2002	2-	WHO-BCAI	MMSE, ADL Scale, GDS Hachinski Ischaemic Scale	N/A: Only given for AD	N/A: Only given for AD

ABCS: AB Cognitive Screen; ACE-R: Addenbrooke's Cognitive Examination Revised; AD: Alzheimer's disease; ADAS: Alzheimer's disease Assessment Scale; ADCS: Alzheimer's disease Cooperative Study; ADL: activities of daily living; AMNART: American National Adult Reading Test; A/SA: alcohol/substance abuse; ASLD: assessment of subtle language disorders; B-ADL: basic activities of daily living; BDRS: Blessed Dementia Rating Scale; BNT: Boston Naming Test; BVMT: Brief Visuospatial Memory Test; CC: cognitive compliant; CDR: Clinical Dementia Rating; CDT: Clock Drawing Test; CERAD: Consortium to Establish a Registry for Alzheimer's disease; CES-D: Center for Epidemiologic Studies Depression Scale; CF: Category Fluency; CHI: closed head injury; CLOX: executive clock drawing task; COWA: controlled oral word association; CVLT: California Verbal Learning Test; Canc: Cancellation; Cog: cognitive; Comp; DAT: Dementia of the Alzheimer's Type; DMS58; DR: delayed recall; DRS: Dementia Rating Scale; DSM: Diagnostic and Statistical Manual of Mental Disorders; dx: diagnosis; ECO: Examen Cognitif par Ordinateur; ECA: Echelle Comportement et Adaptation scale; EEG: electroencephalography; ESL: English as a second language; ed: education; FAS: verbal fluency task; FCSR: Free and Cued Selective Reminding; grp: group; GDS/GDSS: Geriatric Depression Scale; HDRS: Hamilton Depression Rating Scale; HVLT: Hopkins Verbal Learning Test; Hx: history; IADL: instrumental activities of daily living; I-Flex: short form of the executive interview test; IQCODE: informant questionnaire on cognitive decline in the elderly; inpt.: inpatient; LM: logical memory; MAC-Q: memory complaint questionnaire; MCC: Mayo clinic criteria; MCI: mild cognitive

impairment; MDB: Mental Deterioration Battery; MDRS; Mattis Dementia Rating Scale; MMSE: Mini Mental State Exam; mMMSE: modified Mini Mental State Exam; MRI: magnetic resonance imaging; mem: memory; N/A: not applicable; NAB; neuropsychological battery; NART; Nelson Adult Reading Test; ND: neurological disease; NINCDS-ADRDA: National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association; NS: not stated; PD: Parkinson's disease; PaSMO; parallel serial mental operations; Pic: picture; RAVLT: Rey Auditory Verbal Learning Test; RPR: Rivermead paragraph recall test; RT: reaction time; SMC: subjective memory complaint; sMMSE: Standardized Mini Mental State Exam; SD: standard deviation; SM: semantic fluency; ST: short term; TEA; TMT: Trail Making Test; TPT: The Placement Test; TTT: The Token Test; VF: verbal fluency; VL/OM; verbal language/orientation memory ratio; VOSP; visual object and space perception; WAIS-III: Wechsler Adult Intelligence Scale; WHO-BCAI: World Health Organization Neuropsychological Battery of Cognitive Assessment Instruments; WL: word list; WMSII: Wechsler's Memory Scale II

Table 8. Study and participant information.

Study Information			Diagnostic group numbers				Participant Information					
Author	Year	Country	Total	MCI	Control	AD or other	% Female	Avg. Age	Avg. Ed (years)	Ethnicity	Language	Recruit
Dixon, et al.	2007	Canada	304	mild: 82 mod: 78	144	NA	NS	64-73: 51% 74-92: 49%	15.2	NS	English	Ad
Standish, et al.	2007	Canada	642	166	174	302	57	78	11	Caucasian	ESL 12%	MC
Wylie, et al.	2007	USA	40	20	20	NA	58	72.3	15.8	NS	NS	UH/MC
Bonney, et al.	2006	Australia	56	28	28	NA	57	74.2	11.4	NS	NS	MCI: MC Cont: Ad
Borson, et al.	2006	USA	231	77	140	154	NS	NS	NS	48% Asian Am. 22% AfAm 17% Hispanic 7% White non-Hispanic 6% Native Am/other	NS (some non-English)	HF & SS
Cummings, et al.	2006	USA	644	147	497		56	77.1	14.9	23% Non-white	NS	ADCS site
Doniger, et al.	2006	USA/Israel/Canada	160	61	66	33	60	76.7	12.9	NS	NS	MC & AL
Giaquinto & Parnetti	2006	Italy	103	34	41	28	55	≤65:49% 66-80: 41% >80:10%	>5: 51%	NS	Italian	GP

Table 8. Study and participant information.

Study Information			Diagnostic group numbers				Participant Information					
Author	Year	Country	Total	MCI	Control	AD or other	% Female	Avg. Age	Avg. Ed (years)	Ethnicity	Language	Recruit
Greenaway et al.	2006	USA	195	65	65	65	54	72.2	14.9	97% Caucasian	NS	NS
Kirkpatrick et al.	2006	USA	54	7	34	13	69	67	15	NS	NS	UH & CV
Mioshi, et al.	2006	NS	241	36	63	142	40	66.3	12.5	NS	NS	MCI: UH Cont: Spouse
Perneckzy, et al.	2006	Germany	75	45	30	NA	47	67.92	11.93	NS	German	UH
Riberio, et al.	2006	Portugal	179	116	63	NA	63	68.1	7.5	Caucasian	Portuguese	MCI: UH Cont: CV
Ritter, et al.	2006	Switzerland	45	20	25	NA	NS	61.9	13.1	NS	French	Lecture
von Guten, et al.	2006	Switzerland	379	27	237	115	NS	75.5	NS	NS	French	Ad
Chandler, et al.	2005	USA	250	60	95	95	55	73.8	14.6	Caucasian	NS	MCI: UH Cont: CV
Eibenstein, et al.	2005	Italy	58	29	29	NA	55	70.2	≥3	NS	Italian	UH
Gualtieri & Johnson	2005	USA	178	36	89	53	NS	NS	NS	NS	NS	MCI: MC Cont: PP
Inoue, et al.	2005	Japan	106	22	55	29	75	74.2	NS	NS	NS	MCI: MC Cont: HF
Karrasch, et al.	2005	Finland	45	15	15	15	69	69.2	9	NS	Finnish	MCI: UH Cont: CV
Lam, et al.	2005	China	306	ND:66 ID: 75	94	71	NS	79*	2.5*	NS	Chinese/ Cantonese	CV & AL

Table 8. Study and participant information.

Study Information			Diagnostic group numbers				Participant Information					
Author	Year	Country	Total	MCI	Control	AD or other	% Female	Avg. Age	Avg. Ed (years)	Ethnicity	Language	Recruit
Woodard, et al.	2005	USA	200	18	161	21	NS	75.1	15.7	NS	English	GP & MC
Nordund, et al.	2005	Sweden	147	112	35	NA	NS	65.5***	NS	NS	NS	MCI: GP Cont:CV
Barbeau, et al.	2004	France	113	23	40	50	55	72.2	9.2	NS	French	MCI: MC Cont: Ad
Dwolatzky	2004	USA and Israel	98	30	39	29	55	NS	NS	NS	English	MC
Geda, et al.	2004	USA	655	54	514	87	55	79.1**	12.9	NS	NS	MCI: PP Cont: CV
Knox	2004	USA	370	40	126	204	55	71.7**	14.3	94% Cauc. 3% AfAm 2% Asian 1%Hispanic	NS	UH
Leritz	2004	USA	36	18	18	NA	55	74.2	16.8	100% Caucasian	NS	MCI: UH Cont: CV
Lopez	2004	USA	120	18	98	4	55	74.6	14.4	94% Cauc. 3.4% AfAm .9% Asian .9% Hisp.	English	Ad
McCoy	2004	USA	68	15	53	NA	55	78	16	93% Caucasian 7% other	NS	Cont: CV Ad MCI: PP
Mejia, et al.	2004	Mexico	314	74	185	55	55	76.1	2.9	NS	Spanish	Survey
Artero & Ritchie	2003	France	368	308	60	NA	55	76.3	10.5	NS	NS	NS

Table 8. Study and participant information.

Study Information			Diagnostic group numbers				Participant Information					
Author	Year	Country	Total	MCI	Control	AD or other	% Female	Avg. Age	Avg. Ed (years)	Ethnicity	Language	Recruit
de Jagar, et al.	2003	NS	152	29	51	72	55	75.75	14	NS	NS	MCI: GP Cont:CV
Estevez-Gonzalez, et al.	2003	Spain	70	26	17	27	55	67.4*	7.8	NS	NS	Neurologi cService
Chapman, et al.	2002	USA	69	20	25	24	65	NS	NS	NS	NS	NS
Grunwald, et al.	2002	Germany	51	16	20	15	55	78.3	NS	NS	NS	Random
Pasqualetti, et al.	2002	Italy	300	47	86	167	55	71.1	8.5	NS	Italian	Hospital (CS)
Xiao, et al.	2002	China	136	27	83	26	55	67.9*	9.1*	NS	NS	MCI: MC Cont:CV

ADCS: Alzheimer's disease Cooperative Study; Ad: advertisement; AfAm: African American; AL: assisted living; CS: consecutive series; CV: community volunteers;

Cauc.: Caucasian; Cont: control; GP: general practitioner; HF: health fair; Hisp.: Hispanic; MC: memory clinic; MCI: mild cognitive impairment; MMSE: Mini Mental State Exam; NA: not applicable; NS: not stated; PP: participant pools; SS: social services; UH: university hospital

* Significant difference between groups

** Significant difference between Controls and Alzheimer disease/other groups

*** Significant difference between Controls and MCI groups

Table 9. Summary of articles reporting parametric statistics sorted by quality assessment.

Author	QA	Statistical Results
Knox, 2004	2++	Dx algorithm compared to a prior diagnosis: 89.2% accuracy overall (Kappa=0.82, p<0.01) Stepwise regression compared to a priori diagnosis: 88.6% accuracy (Kappa=0.81, p<0.01)
Leritz, 2004	2++	MANOVA: No sig. interaction between group, lag, and prime type (novel/semantic); Slope analysis: RT sig. faster for intact vs.recombined (p<0.001); Linear regression not sig. for composite memory score as predictor of novel priming score
Lopez, 2004	2++	Stepwise Discriminant Function Analysis: LL, Delayed Memory. Index & SF from RBANS (Wilks lambda=0.59) correctly classified 90% of participants but only 63.5% of MCI correctly identified while 96% of controls identified
Bonney, et al., 2006)	2+	t-tests: MCI sig. higher IT than controls (99.5 vs. 70.1) p<0.001; Sig. correlations b/w references and IT: BDI, CDR, CVLT; Logistic regression of reference standards to predict IT only accounted for 35% of IT score variability
Chandler, et al., 2005	2+	Correlation: All reference standards sig. correlated to CERAD total (all p<0.0001); WLL most highly correlated
Cummings, et al., 2006	2+	Logistic regression: CDR 0.5 predicted all behavioral changes (no statistics given); CDR 0.5 had higher percentage of all behavioral changes that CDR 0, but no statistics reported
Dixon, et al., 2007	2+	Logistic Regression: Inconsistency and Latency only sig. predictor (p<0.025) for controls and MCI-mild on the BRT, non sig. for SRT and CRT4
Doniger, et al., 2006	2+	ANCOVA: Memory, EF & Verbal scores sig. different for MCI vs. Controls (p<0.001); Kendall's Tau-c=0.208, (p<0.001) for Depression & Severity of cognitive decline; Depression severity only affected Motor skills (p>0.05)
Estevez-Gonzalez, et al., 2003	2+	ANOVA: Sig. difference with Scheffe's post-hoc test for Controls vs. MCI: Trials 2-5, Immediate recall score, DR score (all p<0.0001), Verbal learning (p<0.001); MANOVA: MCI sig. lower Verbal learning curve than controls (p<0.0001)
Geda, et al., 2004	2+	ANOVA: NPI sig. different (p=0.0001) between groups; Chi square: Sig. difference on all neuropsychiatric symptoms except for euphoria (all p's<0.001)
Giaquinto, et al., 2006	2+	Pearson correlation coefficient: sig. correlation between MMSE & BICQ scores (p<0.0001), BICQ no sig. difference for MCI vs. Controls
Greenaway, et al., 2006	2+	MANOVA sig. differences for groups (p<0.001); Cohen's d >0.90 for MCI vs. Controls on CVLT: TWL, Trial 5, Short/long DFR, % retention, % intrusions, False positives & Recognition discriminability; Logistic regression: TWL & Long DFR classified 68.7% participants
Lam, et al., 2005	2+	Logistic Regression: Education, ADAS-Cog total & delayed scores predicted MCI vs. controls (p<0.05) MemScore also predicted b/w controls vs. MCIID (p<0.05)

Table 9. Summary of articles reporting parametric statistics sorted by quality assessment.

Author	QA	Statistical Results
McCoy, 2004	2+	IIV using t-tests: RAVLT List 1-Controls greater IIV ($p=0.037$), RAVLT % Retained-MCI greater IIV ($p=0.002$); Discriminant analysis: 85% correct classification with RAVLT, 84% with RAVLT and IIV; ANOVA across time: MCI sig. decrease in IIV from Weeks 1-2 ($p=0.019$), but no change for controls
Mioshi, et al., 2006	2+	MANOVA with t-tests: Total score and all subtests were sig. different for MCI vs. Controls except Visuospatial
Pasqualetti, et al., 2002	2+	MDB Factor analysis: Visuospatial, Verbal memory & Language accounted for 75% of variance; MMSE Linear Regression: Visuospatial: 23%, Verbal memory: 25%, Language: 6% (memory increased to 27% with cubic regression)
Perneczky, et al., 2006	2+	Logistic regression: Prediction of dx sig. for both tests: ADAS-MCI-ADL: $p=.002$; ADAS-cog: $p=.041$
Standish, et al., 2007	2+	MANOVA with sig. t-tests for MCI vs. Controls: ABCS total score, Delayed recall & Verbal fluency (all $p's<0.01$); SMMSE no sig. difference
Wylie, et al., 2007	2+	ANOVA: sig. differences on Short & Long DR (CVLT) & Delayed visual memory (RBANS) ($p's<0.001$) and SF ($p<0.01$); RT was slower for MCI overall, but not sig. ($p=0.10$); Sig. effect of flanker condition on accuracy ($p<0.001$) but no difference between group ($p=0.73$); Analysis of delta slopes using ANOVA found sig. difference between groups on delta slopes for last 2 quintiles (slowest presentation) $p<0.01$
Barbeau, et al., 2004	2-	Independent t-tests: MCI sig. lower performance on DMS48 than controls ($p<0.001$); FCSR sig. correlated with DMS48 ($p<0.01$)
Gualtieri, et al., 2005	2-	MANOVA: sig. difference for Controls vs. MCI for Memory, Psychomotor speed, Complex attention & Cognitive Flexibility domain scores
Kirkpatrick, et al., 2006	3	Correlation: UPSIT Sig. positively correlated to ACE ($p=0.005$); Logistic regression: mint, chocolate, lime, cheddar cheese odors predicted MCI ($p<0.0001$, 61% better prediction over chance)
Xiao, et al., 2002	2-	ANOVA & Duncan's tests: sig. difference for MCI vs. Controls: Verbal learning, Mini-token, Visual reasoning, Trail Making B, Sorting and Construction (all $p's<0.05$)

ACE: Addenbrooke Cognitive Exam; ADAS: Alzheimer's disease Assessment Scale; ADL: activities of daily living; ANCOVA: analysis of covariance; ANOVA: analysis of variance; BDI: Beck Depression Inventory; BICQ: Basic Italian Cognitive Questionnaire; b/w: between; DFR: delayed free recall; DR: delayed recall; dx: diagnosis; CERAD: Consortium to Establish a Registry for Alzheimer's disease; CDR: Clinical Dementia Rating; CVLT: California Verbal Learning Test; FCSR: free and cued selective reminding; IIV: Intra-individual variability; IT: inspection time; LL: list learning; MANOVA: multivariate analysis of variance; MCI: mild cognitive impairment; MDB: Mental Deterioration Battery; MMSE: Mini Mental State Exam; RAVLT: Rey Auditory Verbal Learning Test; RBANS: Repeatable Battery for the Assessment of Neuropsychological Status; RFT: RT: reaction time; SF: semantic fluency; sig: significant; TWL: total words learned; UPSIT: University of Pennsylvania Smell Identification Test; WLL: word list learning

Table 10. Summary of articles reporting nonparametric statistics sorted by quality assessment.

Author	QA	Statistical Results
Borson, et al., 2006	2+	McNemar test: MiniCog more sensitive than physician recognition ($p < 0.001$); Overall accuracy: MiniCog: 83%, Physician: 59%
Chapman, et al., 2002	2+	Jonckheere-Terpsta test: MCI significantly worse than controls on all tasks: Main idea, Lesson, Recall, Recognition of details, & all summary measures except unimportant info units (all p 's < 0.05)
de Jagar, et al., 2003	2+	MWU: MCI significantly worse than controls on: Hopkins Verbal Learning Test, RPR, Category Fluency, Boston Naming Test, CLOX, Letter Comparison, Letter Cancellation Time A (all p 's < 0.05)
Riberio, et al., 2006	2+	Chi square: Incidence of cognitive domain delays in MCI: 68.7% temporal orientation, 33.7% Token test, 30.2% Semantic fluency (All sig. different from controls, all p 's < 0.05)
Eibenstein, et al., 2005	2-	MWU: SSST-MCI sig. worse than controls ($p < 0.001$); Wilcoxon found no significant difference for MMSE vs. SSST: MMSE > 26 related to higher SSST and vice versa ($p = 0.0382$)
Inoue, et al., 2005	2-	MWU: Significant difference for Controls vs. MCI for Age & Year of Birth test, ($p < 0.01$) Visual working memory test for 3-D, Second delayed recall & Total score ($p < 0.001$)
Nordund, et al., 2007	2-	MWU with acceptable effect size ($\eta^2 > 0.15 = \text{large}$, $0.06 < \eta^2 < 0.15 = \text{medium}$): Digit symbol, Trail Making Test A, Rey Auditory Verbal Learning Test, Logistic Memory, Delayed Recall, VOSP silhouettes, Boston Naming Test, ASLD repetition, PaSMO, PWT
von Guten, et al., 2006	2-	Non-parametric ANOVA & MWU: Significant difference for MCI vs. Controls for PEPCA-L Total, Attention/Calculation, Delayed Recall (all p 's = 0.000), Immediate recall & Praxis (all p 's = 0.000, 0.05)
Grunwald, et al., 2002	2-	Druskal-Wallis Test: Quality Reproductions significantly different for Controls vs. MCI ($p = 0.000$); Sig. EEG difference during haptic tasks for Controls vs. MCI over right occipital regions (O2) only; No sig. difference on Exploration Time for groups

ANOVA: analysis of variance; MCI: mild cognitive impairment; MMSE: Mini Mental State Exam; MWU: Mann-Whitney U Test; PEPCA-L: Protocole d'Examin Cognitif; RPR: Rivermead Paragraph Recall Test; SSST: Sniffin' Sticks Screening Test

Table 11. Sensitivity, specificity and area under the curve measures for studies sorted by quality assessment score.

Author and Year	QA	#	Test	Sensitivity (%)	Specificity (%)	AUC
Knox, 2004	2++	370	Dx algorithm (DRS, CVLT, BNT, VF, WMS-R/III, TMT-B)	NE vs. impaired: 99.2 MCI vs. AD: 89.2	NE vs. impaired: 87.3 MCI vs. AD: 100	NE vs. impaired: 0.98 MCI vs. AD: 0.95
Lopez, 2004	2++	120	RBANS: Delayed Memory Index (DMI) List Learning (LL) Semantic Fluency (SF) Coding Code-Saving ANAM: CDS-MDC CDI-AC CDD-AC Coding & List Recall (LR) Semantic Fluency (SF) & LR	RBANS: DMI: 99 LL: 83 SF: 56 Coding: 83 Code-Saving: 88 ANAM: CDS-MDC: 94 CDI-AC: 63 CDD-AC: 57 Coding & LR: 55.6 SF & LR: 100	RBANS: DMI: 67 LL: 70 SF: 90 Coding: 58 Code-Saving: 56 ANAM: CDS-MDC: 47 CDI-AC: 75 CDD-AC: 87 Coding & LR: 93 SF & LR: 55	RBANS: DMI: 0.843 LL: 0.827 SF: 0.782 Coding: 0.758 Code-Saving: 0.701 ANAM: CDS-MDC: 0.702 CDI-AC: 0.718 CDD-AC: 0.750 Coding & LR: 0.74 SF & LR: 0.78
Woodard, et al., 2005	2++	200	CERAD: Category Fluency WLM Delayed Recall (DR) WLM Savings Score WLM Recog Dx algorithm of Category Fluency (CF) & WLM Delayed Recall	CERAD: Category Fluency: 72 WLM DR: 83 WLM Savings Score: 89 WLM Recog: 94 Dx algorithm of CF & WLM Delayed Recall: 70	CERAD: Category Fluency: 55 WLM DR: 60 WLM Savings: 55 WLM Recog: 35 Dx algorithm of CF & WLM DR: 83	CERAD: Category Fluency: 0.69 WLM DR: 0.76 WLM Savings: 0.77 WLM Recognition: 0.73 Dx algorithm of CF & WLM DR: 0.76
Chandler, et al., 2005	2+	250	CERAD total score MMSE Word List (WL) Recall	CERAD total: 81.4 MMSE: 66.1 WL Recall: 81.4	CERAD total: 72.6 MMSE: 61.1 WL Recall: 63.2	CERAD total: 0.823 MMSE: 0.692 WL Recall: 0.823
Chapman, et al., 2002	2+	69	Gist level processing; Main idea task Lesson task	AD/no AD: Main idea: task 92 Lesson task: 92	AD/no AD: Main idea task: 96 Lesson task: 76	

Table 11. Sensitivity, specificity and area under the curve measures for studies sorted by quality assessment score.

Author and Year	QA	#	Test	Sensitivity (%)	Specificity (%)	AUC
de Jager, et al., 2003	2+	152	Hopkins Verbal Learning Test (HVL) Category Fluency (CF) Letter Comparison CLOX2 Placing CLOX1; RPR	HVL: 76 CF: 72 Letter Comparison: 72 CLOX2: 69 Placing: 68 CLOX1 77; RPR: 62	HVL: 78 CF: 68 Letter comparison: 68 CLOX2: 60 Placing: 58 CLOX1: 49; RPR: 52	
Lam, et al., 2005	2+	306	Abbreviated Memory Inventory for the Chinese (MIC) = MemScore	MCIND: 55 MCIID: 65	MCIND: 57.4 MCIID: 57.4	MCIND: 0.6 MCIID: 0.644
Perneckzy, et al., 2006	2+	75	ADCS-MCI-ADL ADCS-cog	ADCS-MCI-ADL: 89 ADCS-cog: 78	ADCS-MCI-ADL: 97 ADCS-cog.: 100	ADCS-MCI-ADL: 0.97 ADCS-cog.: 0.93
Ritter, et al., 2006	2+	45	Recognition & Recall from the TRMT	Recognition: 65 Recall: 30	Recognition: 72 Recall: 92	
Artero & Ritchie, 2003	2-	368	ECO & ECA ADL Scale	2 year dementia prediction: 73	2 year dementia prediction: 99	
Eibenstein et. al. 2005	2-	58	Sniffin' Sticks Screening Test (SSST)	84	72	0.8427
Gualtieri, et al., 2005	2-	178	Shifting Attention (SA) Coding Memory Psychomotor (PM) speed Cognitive flexibility Reaction time (RT) Complex attention	SA: 90 Coding: 90 Memory: 90 PM speed: 90 Cognitive flexibility: 90 RT: 90 Complex attention: 90	SA: 75 Coding: 85 Memory: 82 PM speed: 77 Cognitive flexibility: 76 RT: 64 Complex attention: 65	SA: 0.638 Coding: 0.658 Memory: 0.667 PM speed: 0.679 Cognitive flex: 0.692 RT: 0.663 Complex attention: 0.741
Inoue, et al., 2005	2-	106	Computerized screening test developed from MMSE	cutoff of 16: 82%	cutoff of 16: 87%	
Karrasch, et al., 2005	2-	45	Wordlist learning (WL) WL delayed recall WL recognition	WL learning: 73 WL delayed recall: 26 WL recognition: 46	WL learning: 80 WL delayed recall: 100 WL recognition: 93	

Table 11. Sensitivity, specificity and area under the curve measures for studies sorted by quality assessment score.

Author and Year	QA	#	Test	Sensitivity (%)	Specificity (%)	AUC
Kirkpatrick, et al., 2006	3	54	UPSIT: Selective Odorants Total score	Selective Odorants: 57 Total score: 0	Selective Odorants: 97 Total score: 100	

ADL: Activities of Daily Living; AD: Alzheimer’s disease; BNT: Boston Naming Test; CDD-AC: code substitution delayed recall; CDI-AC: code substitution immediate recall; CDS-MDC: code substitution; CDT: Clock Drawing Test; CERAD: Consortium to Establish a Registry for Alzheimer’s disease; CF: Category Fluency; CLOX: clock drawing task; CVLT: California Verbal Learning Test; Canc: cancellation; cog: cognition; DR: delayed recall; DRS: Dementia Rating Scale; Dx: Diagnostic; ECO: Examen Cognitif par Ordinateur; ECA: Echelle Comportement et Adaptation scale; HVL: Hopkins Verbal Learning Test; LL: list learning; LR: list recall; MCIID: mild cognitive impairment incipient dementia; MCIND: mild cognitive impairment no dementia; Mem: memory; RPR: Rivermead Paragraph Recall Test; TEA: Test of Everyday Attention; TPT: The Placement Test, TTT: The Token Test; Recog: recognition; VF: verbal fluency; WLL: word list learning; WLM: word list memory

CHAPTER 5

DISCUSSION

This study reviewed 38 articles supporting assessment protocols to discriminate between normal cognitive ability and mild cognitive impairment (MCI). Systematic review of these articles yielded several groupings. The methodologic design for each study was rated either “well-covered,” “adequate,” or “poor.” The test(s) selected to discriminate between normal cognitive ability and MCI were grouped as follows: formal test batteries, screening protocols, equipment-based, computer-based and miscellaneous. From inspection of these groupings, patterns and conclusions emerge as to how speech-language pathologists may employ such diagnostic tools, within their scope of practice, to identify people with mild cognitive impairment. Evidence-based recommendations of specific tests were determined based on the quality assessment of the article as well as the statistical outcomes of the results, specifically focusing on sensitivity, specificity and area under the curve measures, due to the ability of these measures to determine diagnostic power.

Patterns in Quality Assessment

The systematic review yielded 38 total articles. Thirty-seven articles were Level 2 studies (i.e. case-controlled, cohort, studies) and went through the quality assessment process while one study was a Level 3 (i.e. case study or feasibility study) and could not be further rated for methodology. Out of the 37 rated, 3 studies were rated as (++) , the

highest rating given, suggesting that the methodology was well-controlled. Twenty-three of the 37 rated were given the (+) rating, suggesting that the methodology was fairly well controlled and any factors not controlled for most likely did not affect the results. Ten articles were then rated as (-), suggesting that methodological factors that could have affected the results were not well controlled. No Level 1 studies (i.e. randomized control trials) were found in the literature search but the methodological design of diagnostic research hinders the use of randomization of participants since a priori diagnoses of participants need to be made for later comparison. This systematic review considered Level 2 the most powerful type of study considering the diagnostic nature of the review.

Three studies out of 37 rated for quality assessment in the SR were given an overall rating of ++, suggesting that the majority of the criteria used to evaluate the methodology were met. These studies, Woodard, et al. (2005), Knox (2004) and Lopez (2004) were further analyzed as to determine any similarities in methodological factors that contributed to the overall rating. All studies included over 100 participants with 2 of the studies including over 200 participants. When utilizing statistics to determine significance of the findings, appropriate numbers need to be included and small participant groups lower the power of the study. Statistics are based on overall generalizations of the entire group, so large groups of participants need to be used to obtain an accurate representation of the population. These studies used suitable participant numbers to obtain accurate results. Another factor seen in all studies that received the highest methodological rating was the inclusion of reliability and blinding of the investigators. These studies were well controlled in other areas such as the testing environment and similarity between participant groups other than cognitive status (i.e.

gender, age, etc.). These studies also employed the use of multiple, validated reference standards for a priori diagnoses of participants as well as for comparison of results of the test in question. All three studies also used ROC curve analyses to determine the power of the tests in distinguishing between groups. This method of determining sensitivity and specificity measures has been found to be the most appropriate in diagnostic research because it gives a true picture of how well the test identifies individuals who are normal and abnormal. Two studies, Woodard, et al. (2005) and Knox (2004), support the use of diagnostic algorithms to discriminate individuals with MCI versus controls. These tools are objective and thorough and may include a multitude of tests or tasks. Two studies, Woodard, et al. (2005) and Lopez (2004) determined that 2 tasks: List Recall/Delayed Recall and Semantic Fluency/Verbal Fluency are the most sensitive measures to use when identifying subtle decline in cognition. Although good sensitivity measures were found (80-90%) for the tests suggested by Woodard, et al. (2005) and Lopez (2004), the specificities were very low, suggesting that these tests have a large number of false positives, or individuals deemed as “impaired” who are actually “normal.” Knox (2004), had a good balance of both measures (to be discussed later in the evidence-based recommendations). Two out of three studies were dissertations which may explain the amount of detail included in the methodologies and therefore the overall rating of (++). This brings up the point that articles need to be concise but also do not need to leave out methodological details that may leave a reader or evaluator wondering or assuming how the study was carried out.

Twenty-three studies received a rating of (+), suggesting that many test options are available for the discrimination of MCI from normal cognitive ability. The studies

rated as (+) had some methodological detail that was either left out or not a well-controlled for, but these areas most likely did not affect the outcome and results of the studies. Also, these studies only had 1-2 areas that may have been of concern, whereas the poorly judged group had more concerns that did impact the results. Nine of these studies only lacked in external validity such as only including participants that spoke a foreign language, having an abnormally low or high education level or socioeconomic status or having a large difference in gender in participants. Although these factors impact the generalizability of the findings to the greater population, this was most often the only area of concern and therefore still given a rating of (+). Fifteen of the studies lacked any detail about blinding of administrators to participant diagnosis. This once again, is a methodological concern but most likely did not greatly impact the outcome of the study. Overall, these studies did exercise adequate methodological control to ensure that the results were accurate and spanned a wide range of assessment protocols, to be discussed further in the next section.

Eleven out of the 37 articles rated were rated methodologically poor (-) overall for a multitude of factors that were then further examined for patterns. A critical factor when carrying out diagnostic research is what reference standard is used for a priori diagnosis and comparison. After gathering the literature and evaluating diagnostic research, the reference standard has been found to be a key factor in the methodology and accuracy of the results. If the reference standard that is being used to compare the results of the test being evaluated has not been validated or has been proven to be unable to diagnosis the population being studied, the results of the current test are not sound and accurate. Three studies (Standish, et al., 2007; von Gunten, et al., 2006; Grunwald, et al., 2002)

completely relied on the *MMSE* or reported it as the only reference standard. Although the *MMSE* is a widely-used test, several studies determined it was not powerful enough to detect MCI (to be discussed further in the following paragraph). Two studies reported using the *MMSE* along with other test batteries that were not specified in the article (Eibenstein, et al., 2005; Gualtieri & Johnson, 2005). Three out of the 38 articles reviewed did not report specific tests used for reference standards (Inoue, et al., 2005; Mejia, et al., 2004; Artero & Ritchie, 2003). This lack of detail leaves the results of the studies in question to the investigator of the current SR. Other studies such as Nordlund, et al. (2005) used reference standards that were novel and did not include validation or justification as to why they were included. Another factor involved with the poor ratings of these studies is the small participant numbers as well as insignificant findings. If significance was found in a study, the small participant group numbers is not as much of a concern, because the numbers were large enough to find differences in performance. If a study had small participant numbers and the results are determined insignificant, the small group numbers are of more concern because significant differences may have been found with larger numbers. Four of these studies rated as (-) may have found significant results if larger participant group numbers were employed (von Gunten, et al., 2006; Eibenstein, et al., 2005; Karrash, et al., 2005; Barbeau, et al., 2004). Overall lack of detail about reliability, blinding, testing conditions and participant information was observed to be a pattern with the studies rated as (-). The studies rated as poor spanned all test groups and did not show any pattern as to a particular type of test or task. Diagnostic research can be difficult to control for all outlying factors such as blinding, randomization and

participant selection; however the inclusion of methodological control helps to ensure accurate results.

Speaking specifically to diagnostic research, validated reference standards need to be employed to ensure accurate comparisons of the diagnostic test in question. The *MMSE* was found throughout the literature as a reference standard in evaluation but an interesting finding is that 5 studies, 3 of which were rated as (+) concluded that this quick test is not powerful enough to detect the subtle changes in MCI (Standish, et al., 2007; Karrash, et al., 2005; Mejia, et al., 2005; Chapman, et al., 2002; Pasqualetti, et al., 2002). Although these studies determined that this is not an appropriate tool, the *MMSE* has been used for a long period of time by a multitude of professionals and therefore continues to be recognized as a useful test in cognitive evaluation. The benefit of the *MMSE* is the ease of administration as well as the ability to give and score quickly; however, the preliminary results from this systematic review reveal that it may not be powerful enough to detect the subtle changes with MCI. Many tests and screening protocols have been developed and included in this SR that target specific areas that may be affected in MCI and therefore may be short in administration, yet probing more susceptible areas of early change than the *MMSE*.

One out of the 38 studies was Level 3 study and therefore could not receive quality assessment ratings. The study was an exploratory, feasibility design as the first step towards a larger, more controlled study. This study did undergo data extraction and was found to have useful results and fairly controlled methodological design. It was also on the area of olfaction, which was also found in one other study in the systematic review that was rated as poor. The Level 3 study could not be rated but the results were

promising that olfaction screenings are appropriate for discrimination of MCI versus normal cognitive ability. It is recommended that more studies evaluate olfaction as an avenue of early evaluation of this population.

Patterns in Types of Tests

Patterns of tests were seen during the review of the literature and were divided as follows: 14 articles reporting on formal test batteries, 9 reporting on screening protocols, 4 reporting on computer-based tests, 1 reporting on equipment-based tests and 10 tests in a miscellaneous assessments group. Various tests were found throughout the literature, once again supporting the idea that MCI is a heterogeneous disorder. Tests are available that evaluate different cognitive areas and functions, from tests involving immediate and delayed memory, discourse, list learning, visuospatial tasks, reaction time, attention and affect to name a few. Also many tests combine a number of these tasks in a full battery while others provide short screening protocols that focus on a particular area.

Fourteen articles were included in the group of formal test batteries, which included tests such as the *Consortium to Establish a Registry for Alzheimer's disease (CERAD)*, *RAVLT*, *CVLT*, *Repeatable Battery for the Assessment of Neuropsychological Status (RBANS)*, *Boston Naming Test (BNT)* and *Hopkins Verbal Learning Test (HVLT)*. Eleven of the 14 articles were rated as (+) or higher, with all 3 studies rated as (++) included in this formal test group. Protocols such as the *CERAD* and a diagnostic algorithm of multiple tests are thorough with multiple opportunities given for testing specific areas of cognition, whereas a screening tool may only give one opportunity. It has been determined that MCI is a heterogeneous disorder, and so an exhaustive diagnostic tool may be the most powerful and accurate way to distinguish MCI from

normal cognitive ability. Although powerful support from the literature is given to formal test batteries, some of these are “exhaustive” in nature in reference to time of administration and financial cost and cannot be given to every individual. This is where screening measures can be used as a way to determine if individuals are in need of further evaluation or if they are considered normal and not in need of further evaluation. Although this will possibly trade off accuracy of diagnosis for time and ease of administration, screening measures also have a place in early identification.

The screening protocol group did not include any (++) studies whereas the formal test group included 3. Adequate evidence was found, however to support the use of screening tools with the MCI population with 6 out of the 9 studies obtaining a rating of (+). These screening protocols included tasks which ranged from such as testing olfaction or giving a questionnaire. One article was a Level 3 (case-control study) and unable to rate and the other 2 studies in the group were rated as (-). The purpose of a screening tool is to determine who is normal and does not need any further evaluation and who is not normal and needs to receive additional assessment. Therefore, screening tools often have an increase in false positives (specificity) than a full test battery. Overall, the *ABCS* (Standish, et al., 2007), *Mini-Cog* (Borson, et al., 2006), *Geriatric Depression Scale* (Cummings, et al., 2006), *BICQ* (Giaquinto & Parnetti, 2006), *NPI* (Geda, et al., 2004) and 5-item questionnaire including subjective memory complaint (Lam, et al., 2005) were well-represented in the literature and show promising results as to the early identification and discrimination between MCI and individuals with normal cognitive ability.

Seven out of the 10 studies in the miscellaneous test group were also rated as (+), showing that there is a variety of appropriate ways to identify this heterogeneous

population of individuals with MCI. Two of these methods looked at how a certain task was done (i.e. not the actual performance on a task, but patterns of how it was done or how it varied over time). These included measures of intra-individual variability (McCoy, 2004) and inspection time (Bonney, et al., 2004). Four other studies included in this group investigated newly-developed batteries: the *Topographical Recognition Memory Test* (Ritter, et al., 2006), *DMS48* (Barbeau, et al., 2004) *Battery of Lisbon for the Assessment of Dementia* (Riberio, et al., 2006) *Protocole d'Examen Cognitif de la Personne Agee-Lausanne* (von Gunten, et al., 2006). Three of these newly-developed tests were compared to validated reference standards and found to have sensitive diagnostic ability of MCI. Overall, the miscellaneous group of tests was comprised of 10 studies, 7 of which were rated as (+) determining that there are multiple avenues to explore when evaluating the MCI population. The review of the literature shows that MCI is a heterogeneous population and many tests are available that evaluate various areas of cognition. Seven studies in the miscellaneous test group were supported by the literature and are options to use if other areas of cognition need further evaluation.

The equipment-based group only included one study evaluating the ability of EEG theta-power to identify MCI (Grunwald, et al., 2002). This study was rated as (-) and the results suggest that EEG theta-power measures are not the most powerful to discriminate MCI from controls. This study used cognitive haptic task while measuring the EEG theta-power, although the haptic task was also evaluated apart from the EEG measures. There was no diagnostic ability found for EEG measures, but interestingly significant statistical differences were found in performance between the two groups on the haptic task alone. This study supports the use of these tasks and recommended further

research on EEG measures with this population. Once again, the haptic tasks are more clinically relevant and practical than EEG measures, especially in terms of interdisciplinary evaluation.

The computer-based group was composed of 4 articles, 1 of which was rated as (+). The study by Dongier, et al. (2006) was the one study in the group rated as (+) methodologically, but both studies (Doniger, et al., 2006; Dwoltzky, et al., 2004) support the use of this computerized cognitive battery, *Mindstreams* and describe the benefits to include ease of administration, instant data analysis and ability to adjust to the patient needs as testing is occurring. As seen from this SR, computerized test batteries are becoming more apparent in the literature, although the majority of the literature continues to contain pencil-and-paper tests. Using a computer as a tool in evaluation with the older population may not be as comfortable of an option as regular tests because the older population is generally not as familiar with computers as are the younger population. A change may be seen in the near future as more and more individuals in the aging population are familiar and comfortable with computers and therefore are more apt to be accurately tested on a computer. More well-controlled research is needed to determine the ability of the computerized tests in identification of MCI.

The above section once again illustrates the various types of assessments available for the identification of MCI. The next section will discuss further the scope of practice of an SLP and then draw conclusions as to the specific tests that are supported by the literature and area available for an SLP to use.

Evidence-based Suggestions

Many of the studies included in the systematic review judged to have good methodological rigor support various different tests for the discrimination between MCI and normal cognitive ability. When looking across each study's outcomes, the ROC curve measures including sensitivity, specificity and area under the curve that are detailed in Table 11 were determined to be the most revealing of diagnostic ability. Therefore, evidence-based recommendations were determined as the studies with the most methodological control with the best balance of sensitivity and specificity measures. One of the three studies judged as 2++ in the systematic review had a balanced sensitivity and specificity. Knox (2004) determined that a diagnostic algorithm including the *CVLT*, *Dementia Rating Scale*, *BNT*, *WMS-R*, *Trail-making test B* and verbal fluency had a sensitivity of 99.2% and 87.3% for determining normal versus impaired with a large area under the curve (effect size) of 0.98. Since ROC curve analyses are binary in nature, a second analysis was done in the study for determining from the group who was not normal, how sensitive and specific was the diagnostic algorithm to MCI versus Alzheimer's disease. A good balance was found as well, with the algorithm having 89.2% sensitivity and 100% specificity with a large area under the curve (effect size) of 0.95. These results were found from a well-controlled study, making the findings a strong recommendation from an evidence-based standpoint. Diagnostic algorithms such as the one used in this study are becoming increasingly popular in medicine and health-related fields because they bring objectivity to the diagnostic process. It should be noted that although this is a recommendation, the original author suggested this as an addition to the clinical evaluation and not as a replacement for clinical judgment.

The other test recommended from the findings of this systematic review is the Alzheimer's disease Cooperative Study scale for ADL in MCI (*ADCS-MCI-ADL*) (Pernecky, et al., 2006), which had both a controlled methodology, judged as a 2+ in the systematic review as well as balanced sensitivity and specificity. The *ADCS-MCI-ADL* assesses 18 areas of everyday tasks and abilities through an interview to determine if there is any impairment. The results found a balanced of 89% sensitivity and 97% sensitivity with a large area under the curve (effect size) of 0.97. This combination of findings as well as well-controlled methodology led to the evidence-based recommendation of the *ADCS-MCI-ADL* to help discriminate between individuals with MCI from normal cognitive ability. This is an interesting finding and suggestion since the original Mayo clinic criteria for MCI included intact ADL skills. Since this scale evaluates multiple areas of living, it may be more powerful to detecting subtle deficits and is therefore recommended for evaluation and identification of this population.

Although many studies were judged to have a controlled methodology with findings suggesting various tests to discriminate between MCI and normal cognitive ability, the *ADCS-MCI-ADL* and diagnostic algorithm of the *WMS-R*, *CVLT*, *DRS*, *Trail-Making Test B* and verbal fluency are the most highly recommended given the current literature. As MCI continues to have increasing awareness with early identification in health-related fields, more tests may prove to be powerful to the subtle cognitive declines.

Speech-Language Pathology Scope of Practice

As stated in the justification for this systematic review, ASHA (2005b) developed a position statement for the professional statement of the field of speech-language pathology on the role of a speech-language pathologist (SLP) with individuals with dementia. The role of an SLP includes identification of individuals at risk for dementia as well as evaluation of cognitive-communication disturbances. Early identification of dementia includes MCI, and screening measures are the first step in this process. SLP's in the healthcare setting come in contact with aging individuals often, and it would be helpful for the SLP to be actively involved in the identification of this progression, beginning with screening and then determining who may need possible referrals for further evaluation. ASHA (2007) also states that evaluating cognitive domains such as memory, sequencing, executive function and sequencing is considered to be within the scope of practice of an SLP.

When comparing the evidence-based recommendations of the *ADCS-MCI-ADL* with the SLP scope of practice of early identification of dementia, this interview-based scale assessing 18 areas of everyday living is an evaluation protocol that can be utilized by a speech-language pathologist. Interviewing a patient is usually a part of the evaluative process for a speech-language pathologist and implementing this protocol that has been determined to be powerful to detect subtle cognitive decline may help an SLP discriminate between individuals with MCI and normal cognitive ability.

The other evidence-based recommendation from this systematic review is the diagnostic algorithm consisting of the *CVLT*, *WMS-R*, *Trail Making Test B*, *DRS* and a verbal fluency task. Speech-language pathologists often administer verbal fluency tasks,

often described as “categorical fluency” or asking a patient to name as many animals as he/she can in a minute. Brief verbal learning tests such as the *CVLT* and short cognitive rating scales such as the *DRS* are also considered to be appropriate for a speech-language pathologist to administer. The *Trail Making Test B* is a short test given in 5-10 minutes that requires an individual to match alternating letters and numbers in a particular order as quickly as possible. It is quick way to evaluate executive function and planning as well as time taken to complete the task accurately. This is a task that an SLP could be given to a patient to help discriminate individuals with MCI from normal cognitive ability. The *WMS-R* is a test generally given by neuropsychologists. To complete this diagnostic algorithm to accurately distinguish MCI from normal cognitive ability, an SLP may require neuropsychological input or referral for this test, although the rest of the algorithm can be completed by an SLP. Overall, these evidence-based recommendations are helpful to SLP’s to include in their cognitive protocol as well as understand the need for interdisciplinary evaluation.

An emerging role for SLP’s working with this population is the forestalling, postponing or perhaps prevention of loss of cognitive/linguistic abilities in the aging population. One potential risk pool is in that population of persons with MCI who are at a higher risk of developing dementia over time. ASHA not only endorses early identification of progressive disorders but also prevention measures to help individuals while the disorder is in the early stages (ASHA, 1988). Research is ongoing as to how progression of dementia can be forestalled or prevented. The baby-boomer generation as a whole is a group of people who are informed, assertive and active participants in their healthcare. As they begin to enter into this “aging population” speech-language

pathologists can be used as resources to help maintain communicative stability as the risk of cognitive decline increases.

As seen in this systematic review, the majority of the literature on identification of MCI has come from the field of neuropsychology. The interesting finding is that when the SLP scope of practice is compared, the two areas overlap in many ways. Early identification of this subtle decline in cognition should be a team approach from the early screening to the full evaluation and diagnosis. A speech-language pathologist should be aware of the deficits that are present with this population and have a multitude of tests and tasks available to help discriminate early decline in MCI versus normal cognitive ability in the aging population. This systematic review revealed many tests in varied areas to help in this endeavor. Many tests are validated and supported by the literature and are in the scope of practice of a speech-language pathologist. Overall, MCI is a heterogeneous population and should be identified early by an interdisciplinary team of healthcare individuals, including the speech-language pathologist.

Directions for Future Research

Much research on MCI has been accomplished in the area of neuropsychology, but after extensive searching through the literature and professional contacts made, no research has yet been done on MCI in speech-language pathology. Conclusions and recommendations can and have been made by drawing from other fields' research and comparing the results to an SLP's scope of practice, but research in speech-language pathology is still needed. Research in speech-language pathology on how these subtle cognitive deficits affect language function would be helpful in early identification. An important area of research that is greatly needed is on how early identification and

prevention measures affect the progression to dementia in individuals with MCI. This is an emerging role for an SLP, but may have a great impact on the individuals and also the future caseloads of professionals if proven that prevention measures are helpful with this population. It is also recommended that replication of studies be done in the future to continue researching all of the cognitive deficits associated with MCI as well as powerful tests to use for identification. The results of this systematic review will serve as a guideline for the tests available to discriminate MCI from normal cognitive ability for clinical use and further research in this area.

REFERENCES

References marked with an asterisk indicate studies included in the systematic review.

Alzheimer's Association. (2007, March 20). Alzheimer's Disease Prevalence Rates Rise to More than Five Million in the United States. In *2007 Alzheimer's Disease Facts and Figures*. Retrieved April 16, 2007 from www.alz.org/news_and_events_rates_rise.asp.

American Psychological Association. (2007). *About PsycInfo*. Retrieved on April 5, 2007, from <http://www.apa.org/psycinfo/about/>.

American Speech-Language Hearing Association. (1988). *Prevention of Communication Disorders* (position statement). Available from <http://www.asha.org/policy>.

American Speech-Language Hearing Association. (2005a). Knowledge and skills needed by speech-language pathologists providing services to individuals with cognitive-communication disorders. *ASHA Supplement, 25*, in press.

American Speech-Language Hearing Association. (2005b). *The roles of speech-language pathologists working with individuals with dementia position statement* (position statement). Available from <http://www.asha.org/policy>.

American Speech-Language Hearing Association. (2007). *Scope of Practice in Speech-Language Pathology* (Scope of Practice). Available from <http://www.asha.org/policy>.

- American Speech-Language Hearing Association. (n.d.) Steps in the Process of Evidence-Based Practice. Retrieved on January 22, 2007 from <http://asha.org/members/ebp/framing.htm?print=1>.
- *Artero, S., & Ritchie, K. (2003). The detection of mild cognitive impairment in the general practice setting. *Aging and Mental Health*, 7(4), 251-258.
- Au, R., & Bowles, N. (1991). Memory influences on language in normal aging. In D.N. Ripich (Ed.), *Handbook of geriatric communication disorders* (pp. 293-303). Austin, TX: Pro-Ed.
- *Barbeau, E., Didic, M., Tramon, E., Felician, O., Joubert, S., Sontheimer, A., Ceccaldi, M., & Poncet, M. (2004). Evaluation of visual recognition memory in MCI patients. *Neurology*, 62, 1317-1322.
- Bayles, K.A., & Kaszniak, A.W. (1987). *Communication and cognition in normal aging and dementia*. Boston, MA: College-Hill Press.
- Bayles, K.A., & Kaszniak, A.W. (1987). *Communication and cognition in normal aging and dementia*. Boston, MA: College-Hill Press. p. 101.
- Baylor, C.R., & Yorkston, K.M. (n.d.). Using Systematic Reviews and Practice Guidelines: A How-To Guide for Clinicians. Retrieved April 12, 2007 from ASHA Division 2, Neurophysiology and Neurogenic Speech and Language Disorders Members Only Area Web site: www.asha.org/about/membership-certification/divs/div2member/.
- Benjamin, B.J. (1988). Changes in speech production and linguistic behavior with aging. In B.B. Shadden (Ed.), *Communication behavior and aging: A sourcebook for clinicians* (pp. 172-176). Baltimore, MD: Williams & Wilkins.

- *Bonney, K., Almeida, O., Flicker, L., Davies, S., Clarnette, R., Anderson, M., & Lautenschlager, N. (2006). Inspection time in non-demented older adults with mild cognitive impairment. *Neuropsychologia*, *44*, 1452-1456.
- *Borson, S., Scanlan, J., Brush, M., Vitaliano, P. & Dokmak, A. (2000). The Mini-Cog: A cognitive 'vital signs' measure for dementia screening in multi-lingual elders. *International Journal of Geriatric Psychiatry*, *15*, 1021-1027.
- *Borson, S., Scanlan, J., Watanabe, J., Tu, S., & Lessig, M. (2006). Improving identification of cognitive impairment in primary care. *International Journal of Geriatric Psychiatry*, *21*, 349-355.
- Bourgeois, M. (1992). Evaluating memory wallets in conversations with persons with dementia. *Journal of Speech and Hearing Research*, *35*, 1344-1357.
- Bourgeois, M. (2002). "Where is my wife and when am I going home?" The challenge of communicating with persons with dementia. *Alzheimer's Care Quarterly*, *3(2)* 132-144.
- Bourgeois, M. (2002). "Where is my wife and when am I going home?" The challenge of communicating with persons with dementia. *Alzheimer's Care Quarterly*, *3(2)*, 132.
- *Chandler, M., Lacritz, L., Hynan, L., Barnard, H., Allen, G., Deschner, M., Weiner, M., & Cullum, C. (2005). A total score for the CERAD neuropsychological battery. *Neurology*, *65*, 102-106.
- *Chapman, S., Zientz, J., Weiner, M., Rosenberg, R., Frawley, W., & Burns, M. (2002). Discourse changes in early Alzheimer disease, mild cognitive impairment, and normal aging. *Alzheimer Disease and Associated Disorders*, *16(3)*, 177-186.

- Chapman, S., Ulatowska, H., King, K., Johnson, J., & McIntire, D. (1995). Discourse in early Alzheimer's disease versus normal advanced aging. *American Journal of Speech-Language Pathology, 4(4)*, 124-129
- Civil, R., & Whitehouse. P. (1991). Neurobiology of the aging communication system. In D.N. Ripich (Ed.), *Handbook of geriatric communication disorders* (pp. 7-13). Austin, TX: Pro-Ed.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.
- Cox, R. (2005). Evidence-Based Practice in Provision of Amplification. *Journal of American Academy of Audiology, 16*, 419-438.
- *Cummings, J., Raman, R., Ernstrom, K., Salmon, D., & Ferris, S. (2006). ADCS prevention instrument project: Behavioral measures in primary prevention trials. *Alzheimer's Disease and Associated Disorders, 20(3)*, S147-S151.
- *de Jager, C., Hogervorst, M., Combrinck, M., & Budge, M. (2003). Sensitivity and specificity of neuropsychological tests for mild cognitive impairment, vascular cognitive impairment and Alzheimer's disease. *Psychological Medicine, (33)*, 1039-1050.
- *Dixon, R., Lentz, T., Garrett, D., MacDonald, S., Strauss, E., & Hultsch, D. (2007). Neurocognitive markers of cognitive impairment: Exploring the roles of speed and inconsistency. *Neuropsychology, 21(3)*, 381-399.
- *Doniger, G., Dwolatzky, T., Zucker, D., Chertkow, H., Crystal, H., Schweiger, A., and Simon, E. (2006). Computerized cognitive testing battery identifies mild cognitive impairment and mild dementia even in the presence of depressive symptoms. *American Journal of Alzheimer's Disease and Other Dementias, 21(1)*, 28-36.

- *Dwolatzky, T., Whitehead, V., Doniger, G., Simon, E., Schweiger, A., Jaffe, D., & Chertkow, H. (2004). Validity of the *Mindstreams* computerized cognitive battery for mild cognitive impairment. *Journal of Molecular Neuroscience*, *24*, 33-44.
- *Eibenstein, A., Fioretti, A., Simaskou, M., Sucapane, P., Mearelli, S., Mina, C., Amabile, G., & Fusetti, M. (2005). Olfactory screening test in mild cognitive impairment. *Neurological Scientific*, *26*, 156-160.
- *Estevez-Gonzalez, A., Kulisevsky, J., Boltes, A., Otermin, P., & Garcia-Sanchez, C. (2003). Rey Verbal Learning Test is a useful tool for differential diagnosis in the preclinical phase of Alzheimer's disease: Comparison with mild cognitive impairment and normal aging. *International Journal of Geriatric Psychiatry*, *18*, 1021-1028.
- Galluzzi, S., Cimaschi, L., Ferrucci, L., & Frisoni, G., (2001). Mild cognitive impairment: Clinical features and review of screening instruments. *Aging Clinical Experimental Research*, *13*, 183-202.
- *Geda, Y., Smith, G., Knopman, D., Boeve, B., Tangalos, E., Ivnik, R., Mrazek, D., Edland, S., & Petersen, R. (2004). De novo genesis of neuropsychiatric symptoms in mild cognitive impairment (MCI). *International Psychogeriatrics*, *16(1)*, 51-60.
- *Giaquinto, S., & Parnetti, L. (2006). Early detection of dementia in clinical practice. *Mechanisms of Aging and Development*, *127*, 123-128.
- *Greenaway, M., Lacritz, L., Binegar, D., Weiner, M., Lipton, A., & Cullum, M. (2006). Patterns of verbal memory performance in mild cognitive impairment, Alzheimer disease, and normal aging. *Cognitive Behavioral Neurology*, *19(2)*, 79-84.

- Griffer, M., Hargrove, P., & Lund, B. (2005). A Guideline for Applying Systematic Reviews to Child Language Intervention. *Communication Disorders Quarterly*, 26(4), 226-235.
- *Grunwald, M., Busse, F., Hensel, A., Riedel-Heller, S., Kruggel, F., Arendt, T., Wolf, H., and Gertz, H. (2002). Theta-power differences in patients with mild cognitive impairment under rest condition and during haptic tasks. *Alzheimer Disease and Associated Disorders*, 16(1), 40-48.
- *Gualtieri, C. T., and Johnson, L.G. (2005). Neurocognitive testing supports a broader concept of mild cognitive impairment. *American Journal of Alzheimer's Disease and Other Dementias*, 20(6), 359-366.
- *Inoue, M., Urakami, K., Taniguchi, M., Kimura, Y., Saito, J., and Nakashima, K. (2005). Evaluation of a computerized test system to screen for mild cognitive impairment. *Psychogeriatrics*, 5, 36-41.
- Jefferson, A., Wong, S., Bolen, E., Ozonoff, A., Green, R., and Stern, R. (2006). Cognitive correlates of HVOT performance differ between individuals with mild cognitive impairment and normal controls. *Archives of clinical neuropsychology*, 21, 405-412.
- *Karrasch, M., Sinerva, E., Gronholm, P., Rinne, J., and Laine, M. (2005). CERAD test performances in amnesic mild cognitive impairment and Alzheimer's disease. *Acta Neurologica Scandinavica*, 111, 172-179.
- *Kirkpatrick, M., Combest, W., Newton, M., Teske, Y., Cavendish, J., McGee, R., and Przychodzin, D. (2006). Combining olfaction and cognition measures to screen

for mild cognitive impairment. *Neuropsychiatric Disease and Treatment*, 2(4), 565-570.

Kitchenham, B. (2004). *Procedures for performing systematic review*. (Keele University Technical Report TR/SE-0401 and NICTA Technical Report 0400011T.1). Keele, Staffs, UK: Keele University Software Engineering Group, Department of Computer Science and Eversleigh, Australia: Empirical Software Engineering, National ICT Australia Ltd.

Knox, M.R. (2004). Diagnostic accuracy of neuropsychological algorithms in normal aging, mild cognitive impairment, and Alzheimer's disease. (Doctoral dissertation, The University of Texas Southwestern Medical Center at Dallas, 2004). *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 65(2-B), 1031.

*Lam, L., Lui, V., Tam, C., and Chiu, H. (2005). Subjective memory complaints in Chinese subjects with mild cognitive impairment and early Alzheimer's disease. *International Journal of Geriatric Psychiatry*, 20, 876-882.

*Leritz, E.C. (2004). Associative priming and explicit memory in aging and mild cognitive impairment. (Doctoral dissertation, University of Florida, 2004). *Dissertation Abstracts International*, 65, 4293.

*Lopez, M. (2004). Mild cognitive impairment: Diagnostic screening, aspects of memory, and characteristics of affect. (Doctoral dissertation, Rosiland Franklin University, 2004). *Dissertation Abstracts International*, 66, 3416.

- Lu, Y.Y., Haase, J.E., & Farran, C.J. (2007). Persepectives of persons with mild cognitive impairment: Sense of being able. *Alzheimer's Care Quarterly*, 8(1), 75-86.
- Luis, C.A., Loewenstein, D.A., Acevedo, A., Barker, W.W., & Duara, R. (2003). Mild cognitive impairment: Directions for future reference. *Neurology*, 61, 438-444.
- *McCoy, K.J.M. (2004). Understanding the transition from normal cognitive aging to mild cognitive impairment: Comparing the intraindividual variability in cognitive function. (Doctoral dissertation, University of Florida, 2004). *Dissertation Abstracts International*, 65, 6663.
- *Mejia, S., Gutierrez, L.M., Villa, A.R., & Ostrosky-Solis, F. (2004). Cognition, functional status, education, and the diagnosis of dementia and mild cognitive impairment in Spanish-speaking elderly. *Applied Neuropsychology*, 11(4), 194-201.
- Meyer, J., Xu, G., Thornby, J., Chowdhury, M., and Quach, M. (2002). Longitudinal analysis of abnormal domains comprising mild cognitive impairment (MCI) during aging. *Journal of the Neurological Sciences*, 201, 19-25.
- *Mioshi, E., Dawson, K., Mitchell, J., Arnold, R., and Hodges, J. (2006). The Addenbrooke's Cognitive Examination Revised (ACE-R): A brief cognitive test battery for dementia screening. *International Journal of Geriatric Psychiatry*, 21, 1078-1085.
- Molloy, D., Standish, T., Lewis, D. (2005). Screening for mild cognitive impairment: Comparing the SMMSE and the ABCS. *Canadian Journal of Psychiatry*, 50(1), 52-58.

- *Nordlund, A., Rolstad, S., Hellstrom, P., Sjogren, M., Hansen, S., and Wallin, A. (2005). The Goteborg MCI study: Mild cognitive impairment is a heterogeneous condition. *Journal of Neurol Neurosurgical Psychiatry*, 76, 1485-1490.
- Obler, L.K., Au, R., & Albert, M.L. (1995). Language and aging. In R.A. Huntley & K.S. Helfer (Eds.), *Communication in later life* (pp. 87-90). Newton, MA: Butterworth-Heinemann.
- Pai, M. (n.d.) Searching and selecting primary studies. Systematic Reviews Group, Division of Epidemiology, University of California, Berkley, Retrieved June 15, 2007 on www.medepi.net/meta/lectur.html.
- Pai, M., McColloch, M., & Colford, J. *Systematic Review: A Road Map Version 2.2*. Systematic Reviews Group, University of California, Berkeley, 2002. Retrieved on May 31, 2007 from www.medepi.org/meta/guidelines/Berkeley_Systematic_Reviews_Road_Map_V2.2.
- *Pasqualetti, P., Moffa, F., Chiovenda, P., Carlesimo, G., Caltagirone, C., and Rossini, P. (2002). Mini-Mental State Examination and Mental Deterioration Battery: Analysis of the relationship and clinical implications. *Journal of American Geriatric Society*, 50, 1577-1581.
- *Pernecky, R., Pohl, C., Sorg, C., Hartmann, J., Komossa, K., Alexopoulos, P., Wagenpefil, S., and Kurz, A. (2006). Complex activities of daily living in mild cognitive impairment: Conceptual and diagnostic issues. *Age and Ageing*, 35, 240-245.

- Petersen, R.C., Doody, R., Kurz, A., Mohs, R., Morris, J.C., Rabins, P.V., Ritchie, K., Rossor, M., Thal, L., & Winblad, B. (2001). Current concepts in mild cognitive impairment. *Archives of Neurology*, *58*, 1985-1992.
- Petersen, R.C., Smith, G.E., Waring, S. C., Ivnik, R.J., Tangalos, E.G., & Kokem, E. (1999). Mild cognitive impairment: Clinical characteristics and outcome. *Archives of Neurology*, *56*(3), 303-308.
- Petersen, R.C., Stevens, J.C., Ganguli, M., Tangalos, E.G., Cummings, J.L., & DeKosky, S.T. (2001). Practice parameter: Early detection of dementia: Mild cognitive impairment (an evidence-based review). *Neurology*, *56*, 1133-1142.
- *Riberio, F., de Mendonca, A., Guerreiro, M. (2006). *Dementia and Geriatric Cognitive Disorders*, *21*, 284-290.
- Riberiro, F., Guerreiro, M., & DeMendonca, A. (2007). Verbal learning and memory deficits in mild cognitive impairment. *Journal of Clinical and Experimental Neuropsychology*, *29*(2), 187-197.
- Riberiro, F., Guerreiro, M., & DeMendonca, A. (2007). Verbal learning and memory deficits in mild cognitive impairment. *Journal of Clinical and Experimental Neuropsychology*, *29*(2), 194.
- Ricci, S., Celani, M.G., & Righetti, E. (2006). Development of clinical guidelines: Methodological and practical issues. *Neuroscience*, *27*, 228-230.
- Ripich, D. (1991). Language and communication in dementia. In D. Ripich (Ed.) *Handbook of geriatric communication disorders*. (pp. 262-263). Austin, TX: Pro-Ed.

- Ritchie, K., Allard, M., Huppert, F., Nargeot, C., Pinek, B. and Ledesert, B. (1993). Computerized Cognitive Examination of the Elderly (ECO): The development of a neuropsychological examination for clinic and population use. *International Journal of Geriatric Psychiatry*, 8(11), 899-914.
- *Ritter, E., Despres, O., Monsch, A., and Manning, L. (2006). Topographical recognition memory sensitive to amnesic mild cognitive impairment but not to depression. *International Journal of Geriatric Psychiatry*, 21, 924-929.
- Sackett, D.L., Straus, S.E., Richardson, W.S., & Haynes, R.B. (2000) *Evidence-based medicine: How to practice and teach EBM*. Edinburgh: Churchill Livingstone.
- Scottish Intercollegiate Guidelines Network. (2001). SIGN 50: A Guideline Developers' Handbook. Retrieved on March 7, 2007 from <http://www.sign.ac.uk/guidelines/>.
- Shadden, B.B. (1997). Discourse Behaviors in Older Adults. *Seminars in Speech and Language*. 18(2), 143-157.
- Shewan, C., & Henderson, V.L. (1988). Analysis of Spontaneous Language in the Older Population. *Journal of Communication Disorders*. 21, 139-154.
- *Standish, T., Molloy, D., Cunje, A., and Lewis, D. (2007). Do the ABCS 135 short cognitive screen and its subtests discriminate between normal cognition, mild cognitive impairment and dementia? *International Journal of Geriatric Psychiatry*, 22, 189-194.
- The CINAHL Database. (n.d.) *Products and Services*. IN CINAHL Information Services (n.d.). Retrieved on April 5, 2007, from <http://www.ovid.com/site/catalog/DataBase/40.jsp?top=2&mid=3&bottom=7&subsection=10>.

US Library of Medicine, National Institute of Health (2002). *Fact Sheet: What is the difference between MEDLINE and PubMed?* Retrieved on April 5, 2007, from http://www.nlm.nih.gov/pubs/factsheets/dif_med_pub.html.

*von Gunten, A., Brull, J., Kung, A., Simms, K., Taillefer, D., Geneau, D., and Duc, R. (2006). The Protocole d'Examen Cognitif de la Personne Agee-Lausanne (PECPA-L): A cognitive assessment tool for the French-speaking elderly in Switzerland. *European Neurology*, 55, 22-30.

*Woodard, J., Dorsett, E., Cooper, J., Hermann, B., and Sager, M. (2005). Development of a brief cognitive screen for mild cognitive impairment and neurocognitive disorder. *Aging, Neuropsychology, and Cognition*, 12, 299-315.

*Wylie, S.A., Ridderinkhof, K.R., Eckerle, M.K., and Manning, C.A. (2007). Inefficient response inhibition in individuals with mild cognitive impairment. *Neuropsychologia*, 45, 1408-1419.

*Xiao, S., Yao, P., Li, X., and Zhang, M. (2002). Neuropsychological testing profiles of patients with Alzheimer's disease and mild cognitive impairment: A case-control study. *Hong Kong Journal of Psychiatry*, 12(4), 2-5.

Zhang, Y., Schuff, N., Jhang, G., Bayne, W., Mori, S., Schad, L., Mueller, S., Du, A., Kramer, J., Yaffe, K., Chui, H., Jagust, W., Miller, B., and Weiner, M., (2007). Diffusion tensor imaging of cingulum fibers in mild cognitive impairment and Alzheimer disease. *Neurology*, 68, 13-19.

APPENDIX A

Quality Assessment Checklist		
Reference (Author, date)		
Database:		
Checklist completed by:		
Checklist criteria		This study's rating
1	The nature of the test being studied is clearly specified.	Well covered Not addressed Adequately addressed Not applicable Poorly addressed
2	The test is compared with an appropriate standard.	Well covered Not addressed Adequately addressed Not applicable Poorly addressed
3	Where no standard exists, a validated reference standard is used as comparator.	Well covered Not addressed Adequately addressed Not applicable Poorly addressed
4	Patients for testing are selected either as a consecutive series or randomly, from a clearly defined study population.	Well covered Not addressed Adequately addressed Not applicable Poorly addressed

5	The test and reference standard are measured independently (blind) of each other.	Well covered Not addressed Adequately addressed Not applicable Poorly addressed
6	The test and reference standard are applied as close together in time as possible.	Well covered Not addressed Adequately addressed Not applicable Poorly addressed
7	Results are reported for all patients that are entered into the study.	Well covered Not addressed Adequately addressed Not applicable Poorly addressed
8	A pre-test diagnosis is made and reported.	Well covered Not addressed Adequately addressed Not applicable Poorly addressed
9	Reliability information was completed and reported in the study.	Well covered Not addressed Adequately addressed Not applicable Poorly addressed
10	Internal validity was rated as...	Well covered Not addressed Adequately addressed Not applicable Poorly addressed
11	External validity (socioeconomic status, language, ethnicity, group numbers, etc.) was rated as...	Well covered Not addressed Adequately addressed Not applicable Poorly addressed

12	How reliable are the conclusions of this study? Code ++, +, or -	
----	---	--

APPENDIX B

The following references are the primary sources for all assessment protocols discussed in this systematic review. The references are listed alphabetically according to the test name.

AB Cognitive Screen: Molloy, D., Standish, T., Lewis, D. (2005). Screening for Mild Cognitive Impairment: Comparing the SMMSE and ABCS. *Canadian Journal of Psychiatry, 50*, 183-189.

ADAS-cog: Mohs, R., Rosen, W., and Davis, K. (1983). The Alzheimer's disease assessment scale: An instrument for assessing treatment efficacy. *Psychopharmacological Bulletin, 19*, 448-450.

ADCS-MCI-ADL: Galasko, D., Bennett, D., Sano, M., et al. (1997). An inventory to assess activities of daily living for clinical trials in Alzheimer's disease. The Alzheimer's disease cooperative study. *Alzheimer's Disease and Associated Disorders, 11 (Supplement 2)*, 33-39.

Addenbrooke Cognitive Exam: Mathuranath, P., Nestor, P., Berrios, G., et al. (2000). A brief cognitive test battery to differentiate Alzheimer's disease and frontotemporal dementia. *Neurology, 55,11*, 1613-1620.

Animal Naming: Isaacs, B., and Kennie, A. The Set Test as an aid to the detection of dementia in old people. *British Journal of Psychiatry, 123*, 467-470.

- Automated Neuropsychological Assessment Metrics*: Reeves, D., Kane, R., & Winter, K. (1994). *Automated Neuropsychological Assessment Metrics (ANAM): Test Administrator's Guide Version 1.0*. San Diego, CA: National Cognition Recovery Foundation.
- Benton's Visual Retention Test*: Benton, A.L. *Constructional Apraxia: Some Unanswered Questions*. In A.L. Benton, *Contributions to Clinical Neuropsychology*. New York. Aldine. 1969a.
- Boston Naming Test*: Kaplan, E., Goodglass, H. and Weintraub, S. (1978). *The Boston Naming Test*. Boston: Veterans Administration Medical Center.
- Brief Visuospatial Memory Test*: Benedict RHB. (1997). *Brief Visuospatial Memory Test-Revised: Professional Manual*. Odessa, FL, Psychological Assessment Resources.
- California Verbal Learning Test*: Delis, D., Kramer, J., Kaplan, E. and Ober, B. (2000). *The California Verbal Learning Test (2nd ed.)*. New York: The Psychological Corporation.
- Cambridge Examination for Mental Disorders of the Elderly*: Roth, M., Tym, E., Mountjoy, C., Huppert, F., Hendrie, H., Verma, S., et al. (1986). CAMDEX. A standardised instrument for the diagnosis of mental disorders in the elderly with special references to the early detection of dementia. *British Journal of Psychiatry*, 149, 698-709.
- Center for Epidemiologic Studies Depression Scale*: Radloff (1977). The CES-D Scale: A self report depression scale for research in the general population. *Applied Psychological Measurement*, 1, 385-401.

- Clinical Dementia Rating*: Hughes, C., Berg, L., and Danziger, W. (1982) A new clinical scale for the staging of dementia. *British Journal of Psychiatry*, 140, 566-572.
- Clinical Dementia Rating*: Morris, J. (1993). The clinical dementia rating (CDR): Current version and scoring rules. *Neurology*, 43, 2412-2414.
- Clock Drawing Test*: Shulman, K., Gold, D., Cohen, C., and Zuccherro, C. (1993). Clock-drawing and dementia in the community: A longitudinal study. *International Journal of Geriatric Psychiatry*, 8, 487-496.
- CLOX*: Royall, D., Cordes, J., and Polk, M. (1998). CLOX: An executive clock drawing task. *Journal of Neurology: Neurosurgery and Psychiatry*, 64, 588-594.
- Consortium to Establish a Registry for Alzheimer's Disease*: Morris, J., Heyman, A., Mohs, R., et al. (1989). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuro-cognitive assessment of Alzheimer's disease. *Neurology*, 39, 1159-1165.
- Constructional Praxis*: Rosen, W., Mohs, R., and Davis, K. (1984). A new rating scale for Alzheimer's disease. *American Journal of Psychiatry*, 141, 1356-1364.
- Controlled Oral Word Association*: Benton, A.L. Differential behavioral effects in frontal lobe disease. *Neuropsychologia*
- Cornell Scale for Depression in Dementia*: Alexopoulos, G., Abrams, R., Young, R., et al. (1988). Cornell Scale for Depression in Dementia. *Biological Psychiatry*, 23, 271-289.
- ECA*: Richie, K. and Ledesert, B. (1991). The measurement of incapacity in the severely demented elderly: The validation of a behavioural assessment scale. *International Journal of Geriatric Psychiatry*, 6, 217-226.

- ECO*: Ritchie, K., Allard, M., Huppert, F., Nargeot, C., Pinek, B., and Ledesert, B. (1993). Computerized cognitive examination of the elderly (ECO): The development of a neuropsychological examination for clinic and population use. *International Journal of Geriatric Psychiatry*, 8, 899-914.
- Flanker task*: Eriksen, B., Eriksen, C. (1974). Effects of noise letters upon the identification of target letters in a non-search task. *Perception and Psychophysics*, 16, 143-149.
- Free and Cued Recall Selective Reminding*: Grober, E., Buschke, H., Crystal, H., Bang, S., and Dresner, R. (1988). Screening for dementia by memory testing. *Neurology*, 38, 900-903.
- Geriatric Depression Scale*: Hoyle, M., Alessi, C., Harker, J., et al. (1999). Development and testing of a five-item version of the Geriatric Depression scale. *Journal of the American Geriatric Society*, 47, 873-878.
- Geriatric Depression Screening Scale*: Brink, T., Yesavage, F., Lum, O., Heersema, P., Adey, M. and Rose, T. (1982). Screening tests for geriatric depression. *Clinical Gerontology*, 1, 37-43.
- Gist-level response coding*: Delis, D., Kramer, J., and Kaplan, E. (1984). The California Proverb Test. *Unpublished protocol*. Stein, B., and Kirby, J. (1992). The effect of text absent and text present conditions: Summarization and recall of text. *Journal of Reading Behavior*, 24, 217-232.
- Hamilton Depression Rating Scale*: Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology and Neurosurgical Psychiatry*, 23, 56-61.

- Hamilton Depression Rating Scale*: Hamilton, M. (1979). Development of a scale for primary depressive illness. *British Journal of Social Clinical Psychology*, 6, 278–296.
- Hopkins Verbal Learning Test*: Brandt, J. (1991). The Hopkins Verbal Learning Test: Development of a new memory test with six equivalent forms. *Clinical Neuropsychologist*, 5, 125-142.
- Informant Questionnaire on Cognitive Decline in Elderly*: Jorm, A., Korten, E. (1988). Assessment of cognitive decline in the elderly by informant interview. *British Journal of Psychiatry*, 52, 209-213.
- Instrumental Activities of Daily Living*: Lawton, M., and Brody, E. (1969). Assessment of older people: Self-maintaining and instrumental activities of daily living. *Gerontologist*, 9, 179-186.
- Letter cancellation A/B*: Caplan, B. (1985). Stimulus effects in unilateral neglect? *Cortex*, 21, 69-80.
- Mattis Dementia Rating Scale*: Lucas, J., Ivnik, R., Smith, G., Bohac, D., Tangalos, E., et al. (1998). Normative data for the Mattis Dementia Rating Scale. *Journal of Clinical and Experimental Psychology*, 20, 536-547.
- Memory Complaint Questionnaire*: Crook, T., Feher, E., & Larrabee, G. (1992). Assessment of memory complaints in age-associated memory impairment: The MAC-Q. *International Psychogeriatrics*, 4, 165-176.
- Mental Deterioration Battery*: Carlesimo, G., Caltagirone, C., Gainotti, G., et al. (1996). The Mental Deterioration Battery: Normative data, diagnostic reliability and

qualitative analyses of cognitive impairment. The Group for the Standardization of the Mental Deterioration Battery. *European Neurology*, 36, 378-384.

Mindstreams Global Assessment Battery: Dwolatzky, T., Whitehead, V., Doniger, G. et al. (2003). Validity of a novel computerized cognitive battery for mild cognitive impairment. *BMC Geriatric*, 3, 4. NeuroTrax Corporation (2003). *Mindstreams cognitive health assessment*. <http://www.mindstreamshealth.com/content/mssupmat.pdf>.

Mini Mental State Exam: Folstein, M., Folstein, S., and McHugh, P. (1975). 'Minimal state' A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12, 189-198.

Mini-Cog: Borson, S., Scanlan, J., Brush, M., Vitaliano, P., and Dokmak, A. (2000). The Mini-Cog: A cognitive 'vial signs' measure for dementia screening in multi-lingual elderly. *International Journal of Geriatric Psychiatry*, 15, 1021-1027.

Nelson Adult Reading Test: Nelson, H., & Willison, J. (1991). *The National Adult Reading Test (2nd Edition)*. Windsor: NFER-Nelson.

Nelson Norms of Free Association: Nelson, H.E. (no date). *The Nelson Adult Reading Test (NART) test manual*. Windsor, England: NFER-Nelson.

Neuropsychiatric Inventory: Cummings, J., Mega, M., Gray, K., Rosenberg-Thompson, S., Carusi, D., et al. (1994). The Neuropsychiatric Inventory: Comprehensive assessment of psychopathology in dementia. *Neurology*, 44, 2308-2314.

Pattern comparison and letter comparison: Salthouse, T., and Babcock, R. (1991). Decomposing adult age differences in working memory. *Acta Psychologica*, 79, 155-170.

- Pfeffer Functional Activities Questionnaire*: Pfeffer, R., Kurosaki, T., Harrah, C., Chance, J., & Filos, S. (1982). Measurement of functional activities in older adults in the community. *Journal of Gerontology*, 37, 323-329.
- Protocole d'Examin Cognitif*: von Gunten, A., Brull, J., Kung, A., Simms, K., Taillefer, D., Geneau, D., and Duc, R. (2006). The Protocole d'Examen Cognitif de la Personne Agee-Lausanne (PECPA-L): A cognitive assessment tool for the French-speaking elderly in Switzerland. *European Neurology*, 55, 22-30.
- Repeatable Battery of Neuropsychological Status*: Randolph, C.: RBANS Manual—Repeatable Battery for the Assessment of Neuropsychological Status. San Antonio, Tex, Psychological Corp (Harcourt), 1998
- Rey Auditory Verbal Learning Test*: Rey, A. (1964). L'examen clinique en psychologie. Paris: Presses Universitaires de France.
- Rey-Osterrieth complex figure copy*: Rey, (1941). L'examen psychologique dans les cas d'encephalopathie tramatique. *Archives de Psychologie*, 28, 215-285.
- Rivermead Paragraph Recall*: Wilson, B., Cockburn, J., and Baddeley, A. (1985). *The Rivermead Behavioral Memory Test*. Thames Valley Test Company: Reading.
- Sniffin' Sticks Screening Test*: Kobal, G., Kummel, T., Sekinger, B., Barz, S., Roscher, S., Wolf, S. (1996). "Sniffin' Sticks": Screening of olfactory performance. *Rhinology*, 34, 222-226.
- Stepwise comparative status analysis*: Wallin, A., Edman, A., Blennow, K., Gottfries, C. G., Karlsson, I., Regland, B., & Sjogren, M. (1996). Step-wise comparative status analysis (STEP): A tool for identification of regional brain syndromes in dementia. *Journal of Geriatric Psychiatry and Neurology*, 9, 185–199.

- Stroop Test*: Stroop, J.R. Studies of Interference in Serial Verbal Reaction. *Journal of Experimental Psychology*, 1935, 18, 643-662.
- Test of Everyday Attention*: Robertson, L., Ward, T., Ridgeway, V., and Nimmo-Smith, I. (1994). *The Test of Everyday Attention*. Thames Valley Test Company: Bury St Edmunds.
- The Placement Test*: Milwain, E. (1999). An evaluation of memory loss in old age and Alzheimer's disease. D.Phil, thesis, University of Oxford.
- The Token Test* (shortened version), *Bisecting lines, Spatial rotation, Incomplete letters and drawings*: Lezak, M. (1995). *Neuropsychological Assessment, 3rd edition*. Oxford University Press: New York.
- Topographical Recognition Memory Test*: Warrington, E. (1996). *Topographical Recognition Memory Test*. Psychology Press: hove.
- Trail Making Test A/B*: Reitan, R. (1979). *Trail Making for Adults*. Bern, Gottingen: Hogrefe-Verlag.
- University of Pennsylvania Smell Identification Test*: Doty, R., Shaman, P., and Dann, M. (1984). Development of the University of Pennsylvania Smell Identification Test: A standardized microencapsulated test of olfactory function. *Physiological Behavior*, 32, 489-502.
- Visual object and space perception silhouettes*: Bonello, P. J., Rapport, L. J., & Millis, S. R. (1997). Psychometric properties of the Visual Object and Space Perception Battery in normal older adults. *The Clinical Neuropsychologist*, 11, 436-442.
- Wechsler Adult Intelligence Scale-R*: Wechsler, D. (1981). *WAIS-R manual*. New York: Psychological Corporation

Wechsler Memory Scale-R: Wechsler D. *Wechsler memory scale—revised manual*. San Antonio: Psychological Corporation, 1987.

Wide Range Achievement Test: Wilkinson, G. (1993). *Wide Range Achievement Test-3 (WRAT-3) administration manual*. Wilmington, DE: Wide Range.

Wisconsin Card Sorting Test: Berg, E.A. (1948). A simple objective test for measuring the flexibility of thinking. *Journal of General Psychology*.

World Health Organization Battery of Cognitive Assessment Instruments: World Health Organization. (1996). *World Health Organization Battery of Cognitive Assessment Instrument for elderly*. Geneva: World Health Organization.