



July 2015

Performance-based Tools for Assessing Functional Performance in Individuals with Mild Cognitive Impairment

Patricia da Cunha Belchior
McGill University - Canada, patricia.belchior@mcgill.ca

Melanie Holmes
MAB-Mackay Rehabilitaiton Center - Canada, melanie.holmes@utoronto.ca


Nathalie Bier
University of Montreal - Canada, nathalie.bier@umontreal.ca

Carolina Bottari
University of Montreal - Canada, carolina.bottari@umontreal.ca

Barbara Mazer
McGill University - Canada, barbara.mazer@mcgill.ca

See next page for additional authors

Follow this and additional works at: <https://scholarworks.wmich.edu/ojot>

 Part of the Geriatrics Commons, and the Occupational Therapy Commons

Recommended Citation

Belchior, P. d., Holmes, M., Bier, N., Bottari, C., Mazer, B., Robert, A., & Kaur, N. (2015). Performance-based Tools for Assessing Functional Performance in Individuals with Mild Cognitive Impairment. *The Open Journal of Occupational Therapy, 3*(3). <https://doi.org/10.15453/2168-6408.1173>

This document has been accepted for inclusion in The Open Journal of Occupational Therapy by the editors. Free, open access is provided by ScholarWorks at WMU. For more information, please contact wmu-scholarworks@wmich.edu.

Performance-based Tools for Assessing Functional Performance in Individuals with Mild Cognitive Impairment

Abstract

Background: It is now recognized that individuals with mild cognitive impairment (MCI) face subtle functional declines that can compromise performance in everyday tasks. However, it is still not clear how to capture these declines in the clinical setting. Thus, the goal of this study was to conduct a scoping review to identify performance-based tools for which the psychometric properties have been evaluated with the MCI population.

Methods: A scoping review of the scientific literature was performed with the guidance of a health science librarian in searching the MEDLINE, PsychINFO, CINAHL, and EMBASE databases from their inception until May 2014.

Results: Nine performance-based tools assessing functional performance in individuals with MCI have been identified in the literature. While construct and content validity have been extensively reported, only two tools provided data on reliability.

Conclusion: Considering that functional decline is part of the normal aging process, it might be challenging to differentiate normal from pathological functional decline in this population. Functional measurement tools might be very sensitive to capture these subtle changes. Although no recommendations can be proposed at this point on a specific tool to assess functional performance in MCI, research in this area is beginning to identify the elements that should be taken into consideration when choosing a tool.

Keywords

instrumental activities of daily living, mild cognitive impairment, assessments

Cover Page Footnote

The Richard and Edith Strauss Canada Foundation funded this study.

Complete Author List

Patricia da Cunha Belchior, Melanie Holmes, Nathalie Bier, Carolina Bottari, Barbara Mazer, Alexandra Robert, and Navaldeep Kaur

Credentials Display

Patrícia Belchior, Ph.D.; Melanie Holmes, M.Sc., OT; Nathalie Bier, Ph.D.; Carolina Bottari, Ph.D.; Barbara Mazer, Ph.D.; Alexandra Robert, M.Sc., OT; Navaldeep Kaur, PhD student

Copyright transfer agreements are not obtained by The Open Journal of Occupational Therapy (OJOT). Reprint permission for this Applied Research should be obtained from the corresponding author(s). Click here to view our open access statement regarding user rights and distribution of this Applied Research.

DOI: 10.15453/2168-6408.1173

Researchers define mild cognitive impairment (MCI) as a transition stage between normal aging and dementia. Yet, studies have shown that not all individuals with MCI will convert to dementia; some individuals remain stable while others improve (Albert et al., 2011; Winblad et al., 2004). MCI is generally classified into subtypes: amnesic MCI, when memory concerns are most prominent and non-amnesic MCI, when other cognitive deficits are more evident (i.e., attention). Further classifications of MCI include either single domain, when major decline occurs in one cognitive skill, or multiple domains, when major decline occurs in multiple cognitive skills (Albert et al., 2011; Winblad et al., 2004). Current diagnostic criteria recognize that in addition to cognitive impairment, individuals with MCI face declines in functional performance, particularly in the performance of instrumental activities of daily living (IADL). Although individuals with MCI are typically independent in performing basic everyday activities (e.g., self-care), they exhibit less efficiency in the execution of more complex activities, for example, taking more time and making more errors during task completion (Albert et al., 2011; Petersen et al., 2014).

A number of studies investigating the differences in IADL performance between individuals with MCI and normal controls have shown that the MCI group performed more poorly (Bangen et al., 2010; Binengar, Hynan, Lacritz, Weiner, & Cullum, 2009; Gomar, Harvey, Bobes-Bascaran, Davies, & Goldberg, 2011; Griffith et al., 2003; Kounti, Tsolaki, & Kiosseoglou, 2006;

Pereira, Yassuda, Oliveira, & Forlenza, 2008; Schmitter-Edgecombe, McAlister, & Weakley, 2012; Wadley, Okonkwo, Crowe, & Ross-Meadows, 2008). However, the nature of this change in performance is not well understood, as few studies have investigated the characteristics of such decline (e.g., specific types of errors during performance).

The question of the best way to capture these subtle but important changes in IADL performance is important, as these declines are typically difficult to detect. Also, because IADL decline is part of the normal aging process, it is difficult to know when a decline becomes pathological. To date, there is no clear operational definition to capture these changes, leaving clinicians with little guidance on how to assess IADL performance in individuals with MCI.

A few studies have started to provide some insight in this regard. For example, Giovannetti, Bettcher, Brennan, Libon, Kessler, et al. (2008) investigated the patterns of functional decline in the MCI population by observing individuals performing simple everyday tasks (e.g., prepare toast with jelly, prepare coffee with cream and sugar). Although results showed that both the normal controls and the individuals with MCI could complete the tasks independently, the individuals with MCI made more errors during task completion. For instance, the execution of the task was not efficient (e.g., pouring too much cream into the coffee), the sequence of the task steps was poor (e.g., applying butter on bread before toasting the

bread), or object selection was not accurate (e.g., using a spoon to spread butter).

In another study, De Vriendt et al. (2012) used qualitative interviews to investigate the process of functional decline in individuals with MCI. Results showed that the execution of activities demanded more energy and that these individuals had diminished performance skills (e.g., difficulty in monitoring the steps of a task, making plans, initiating new tasks). In addition, participants reported difficulties adapting to new situations and were less flexible when reacting to unexpected events. Lastly, Rosenberg, Kottorp, Winblad, and Nygård (2009) found that individuals with MCI have an increased perception of difficulty in using everyday technology, such as remote controls, cell phones, and microwave ovens. This was related to intrapersonal capacities, including the capacity to manage stress, pay attention and focus, recall necessary information, and respond to environmental demands, such as technology design (Malinowsky, Almkvist, Nygård, & Kottorp, 2011).

Despite growing evidence suggesting that individuals with MCI already face subtle but important functional declines, it is still not clear how to capture these IADL performance declines in the clinical setting. Performance-based measurement tools with which evaluators can observe individuals executing a task in a real-world environment might be more sensitive to detect these changes than questionnaires. Yet, to date, there are no guidelines available to clinicians regarding an optimal IADL performance measurement tool that captures this mild change in functional

performance. The goal of this study was to conduct a scoping review of the literature to identify performance-based IADL measurement tools for which the psychometric properties have been evaluated with the MCI population.

Methods

Search Strategy

The authors performed a scoping review using a structured approach to gather the data (Armstrong, Hall, Doyle, & Waters, 2011). A health science librarian provided guidance during the search. MEDLINE, PsychINFO, CINAHL, and EMBASE databases were searched from their inception until May 2014 to identify performance-based measurement tools that have been used to assess functional performance in individuals with MCI. The search included the following words, both as MeSH terms (*in italics*) and as keywords, to identify potentially relevant primary studies: *mild cognitive impairment* (MeSH), or *cognition disorder* (MeSH) AND *psychometrics* (MeSH), or *reliability or validity* AND *outcome assessments (health care)* (MeSH), or *measure** or *assess** or *evaluat** AND *activities of daily living* (MeSH), or *activit** AND *ecological* or “*real life*” or *function**. The authors also searched the titles of the tools. Textbooks reviewing the psychometric properties of functional measurement tools as well as the Google and Google scholar search engines were used to acquire additional information on the clinical utility of the identified tools (e.g., price of the tool, cost, and ordering information).

Inclusion/Exclusion Criteria

In order to determine the appropriate studies for the review, the authors found the tools used to assess MCI, and then looked for studies on the psychometric properties of these tools. Eligible studies on the psychometric properties of the tools met the following criteria: (a) published in English; (b) peer-reviewed; (c) described a measurement tool that is available in English; (d) described an ecologically valid performance-based functional tool (where assessments were performed in a real-world, simulated real-world, or lab-based environment) that has been used to evaluate functional performance in individuals with MCI; and (e) presented the tool's psychometric properties with the MCI population (one or more of the MCI subtypes).

Description of the Tools

Once the authors completed the search and selected the eligible articles, the information regarding each measurement tool was classified according to: (a) study population (i.e., MCI subtype: amnesic, non-amnesic, single, or multiple domain); (b) assessment environment (i.e., lab-based, simulated real-world, or actual real-world environment); (c) psychometric properties specific to the MCI population (i.e., reliability, validity, and responsiveness to change; scoring system adapted from Poulin, Korner-Bitensky, & Dawson, 2013); and (d) clinical utility (i.e., testing situation, time, therapist training, cost, and scoring; classification system adapted from Law, Baum, & Dunn, 2005).

Psychometric Properties

The authors adapted the classification criteria used to rate the measurement tools and the definitions of the psychometric properties from previous studies conducted with a stroke population (Poulin et al., 2013; see Appendix B). These evaluation criteria quantify each psychometric property using a recommended standard and provide guidance in the interpretation of the ratings. We looked at the specific properties below.

Reliability. Reliability is the extent to which a measure is stable over time and produces a consistent outcome under a given condition (test-retest). It also refers to the ability of the examiner to produce the same results across trials (intra-rater) or the ability of different raters to produce the same outcome with the same group of subjects (inter-rater). Internal consistency refers to the extent to which items measure various aspects of the same construct (Portney & Watkins, 2009; Streiner & Norman, 2003).

Validity. Validity is generally understood as the ability of an instrument to measure what it intends to measure. The most frequently reported types of validity include: content validity, or the extent to which the measure adequately covers the domain under investigation, and construct validity, which is sub-classified into: (a) known-groups or divergent validity, which is the ability of an instrument to discriminate between individuals with or without a certain trait, and (b) convergent validity, which indicates that two tools measuring the same underlying phenomenon should produce the same results. Lastly, criterion validity is the

correlation of the measure of interest with some other measure of the same trait, ideally a “gold standard.” Concurrent validity refers to the relationship between test scores and criterion measurement made at the time the test was given (Crocker & Algina, 1986; Portney & Watkins, 2009; Streiner & Norman, 2003), while predictive validity implies that the criterion measure occurs at a future point in time.

Responsiveness to change.

Responsiveness is the ability of a measure to detect changes longitudinally (Portney & Watkins, 2009; Streiner & Norman, 2003). Considering that MCI is a risk factor for dementia, assessing the longitudinal changes in functional performance is an essential element to be considered with this population.

Results

The initial database search retrieved 564 studies: 282 results from MEDLINE, 227 from PsychINFO, 30 from CINAHL, and 25 from EMBASE. After excluding duplicates, 516 articles remained. Six articles met the inclusion criteria in that they included performance-based functional measurement tools for which the psychometric properties had been studied for MCI assessment. A review of these articles’ reference lists yielded another three articles. An occupational therapist as well as a trained research staff reviewed the results yielded by the search strategies and verified that the content was relevant to the objective of this review.

Appendix A includes detailed information on the nine performance-based measurement tools identified in this review: the Direct Assessment of

Functional Status Revised (DAFS-R; McDougall, Becker, Vaughan, Acee, & Delville, 2009); the Day-Out Task (DOT; Schmitter-Edgecombe et al., 2012); the Financial Capacity Instrument (FCI; Griffith et al., 2003); the Functional Cognitive Assessment Scale (FUCAS; Kounti et al., 2006); the Independent Living Scale (ILS; Bangen et al., 2010); the Naturalistic Action Test (NAT; Giovannetti, Bettcher, Brennan, Libon, Kessler et al., 2008); the Texas Functional Living Scale (TFLS; Binegar et al., 2009); the Timed Instrumental Activities of Daily Living (TIADL; Wadley et al., 2008); and the University of California, San Diego Performance-Based Skills Assessment (UPSA) - Short version (Gomar et al., 2011).

Study Population

The study population included participants with different MCI subtypes. Of the nine studies, one did not specify the MCI subtype (Kounti et al., 2006); four recruited individuals with amnesic and non-amnesic single and multiple domain MCI (Bangen et al., 2010; Giovannetti, Bettcher, Brennan, Libon, Kessler et al., 2008; Pereira et al., 2008; Schmitter-Edgecombe et al., 2012); two included participants with amnesic single and multiple domain MCI (Gomar et al., 2011; Binegar et al., 2009); and the final two investigated only participants with the amnesic MCI subtype (Griffith et al., 2003; Wadley et al., 2008).

Environment Context

Of the nine measurement tools, none were administered in a real-world environment. All of the tools were either used in a laboratory context

using real-life materials or in a simulated real-world environment.

Psychometric Properties

Reliability. The reliability ratings are only available for two of the measurement tools included in this review (see Appendix A). For these, different types of reliability have been reported: (a) Inter-rater reliability was reported for the DOT (Schmitter-Edgecombe et al., 2012), a tool in which participants are required to multi-task in a simulated real-world environment, and the FUCAS (Kounti et al., 2006), a lab-based tool of observation of everyday performance; and (b) Internal consistency was reported for the FUCAS. Results indicate overall adequate to excellent evidence for different types of reliability for these tools based upon the evaluation criteria used in this study.

Validity. For all of the tools, some evidence of validity has been shown (see Appendix A). The most frequently reported types of validity are content and construct validity, with the latter being classified into known-groups and convergent validity. Content validity was reported for all tools and ranged from minimal to adequate evidence. Some type of construct validity was also reported for all tools. While the known-groups validity was assessed in all instruments with minimal to adequate values, convergent validity was only reported in two studies—the DAFS-R and the ILS. Criterion validity has only been measured in relation to four instruments (i.e., FUCAS, TFLS, UPSA, and TIADL). Overall, the DAFS-R and the ILS are the instruments with the most types of validity reported with adequate ratings. However, it

is important to note that the DAFS-R was validated with a Brazilian sample and future research should be conducted to establish its validation with the North American population. Finally, ceiling and floor effects have only been reported in the UPSA (Gomar et al., 2011).

Responsiveness to change. Two measurement tools have reported data on responsiveness: the DAFS-R, an observational measure of functional status (Pereira et al., 2008), and the FCI (Griffith et al., 2003), an instrument specifically designed to assess functional abilities. Minimal information has been reported on the responsiveness of these instruments; therefore, the evidence is poor.

Clinical utility. While most of the tools include observational tasks that are easy to carry out and that require minimal equipment, the DOT requires a more naturalistic setting and should be administered in an environment that is familiar to the participant. The time required for the administration of the tools varies among the instruments, but generally ranges from 15 minutes to over 1 hour. The tools with the shortest administration time are the TFLS, the UPSA, and the TIADL (approximately 15 minutes). Most of the tools require little to no formal training. Only the UPSA requires formal training lasting several hours; it can then be administered by trained non clinicians (Gomar et al., 2011). Overall, different scoring systems were used. However, most of the tests use a point scale (i.e., levels of difficulty or levels of independence) to indicate an individual's ability to perform the given task. The NAT is the

only instrument that factors in errors during task performance.

Discussion

This scoping review identified nine performance-based IADL measurement tools that have been studied with the MCI clientele. First, the authors note that the types of activities evaluated in each measure vary greatly. For instance, some measurement tools focus on financial management (i.e., the FCI), while others focus on a variety of activities, such as taking medication and planning a trip (e.g., the DOT). It is not clear from our review if clinicians should focus on assessing the magnitude of the functional decline of a specific task (e.g., finance management) or if the focus should be on a more general functional decline. In order to understand how to best assess functional performance in individuals with MCI and delineate normal from pathological decline, researchers should address this issue in future studies.

In terms of environments, all of the tools were performed in a laboratory context using real-life material or in a simulated real-world environment. However, it is now recognized that performance observed in the client's home and familiar community environment better reflect real-life abilities compared to clinical settings. Provencher, Demers, Gagnon, and Gélinas (2012) found that evaluations completed in home settings compared to clinical settings are preferable for a more accurate assessment of cooking abilities in frail, older adults with cognitive deficits. Participants were shown to perform better in their homes. In fact, real-world assessments are

considered the optimal manner in which to document the interplay between individuals' cognitive deficits and the environment requirements of their daily activities for a better appreciation of everyday functioning (Schwartz, Segal, Veramonti, Ferraro, & Buxbaum, 2002).

As for psychometric properties, only two of the measurement tools reported evidence on reliability with the MCI population—the DOT (Schmitter-Edgecombe et al., 2012) and the FUCAS (Kounti et al., 2006). This is worrisome given that the subtlety of the types of errors experienced by these individuals requires clinicians to depend on reliable measures to identify errors accurately. Different psychometric properties have been studied for the tools included in this review. No one measure has presented adequate values for both reliability and validity measures.

Proposing specific recommendations to the clinical community about the best IADL performance-based test to use with the MCI clientele poses a challenge. Nonetheless, the information provided in this review can help clinicians make informed decisions when selecting a measurement tool. Researchers are investigating the significant elements to consider when assessing functional performance in individuals with MCI. For instance, it is becoming increasingly apparent that applying error analysis to a performance-based tool might be more sensitive to capture the subtle changes in MCI (Giovannetti, Bettcher, Brennan, Libon, Burke, et al., 2008). One of the instruments included in this review used such an error analysis: the NAT. The error score in the NAT tracks

different types of errors, thus allowing for a more refined analysis of performance than global scores may provide. Furthermore, quantifying and naming errors in different aspects of performance could enable clinicians to identify more specific areas of functional decline in MCI. For example, if a clinician assesses functional performance in cooking, the measurement tool can capture very specific issues based on the new criterion proposed by Albert et al. (2011): “errors/efficiency/time” (e.g., using the wrong ingredients, forgetting the meal is in the oven, taking an extended period of time to read and understand instructions).

Errors could also be analyzed in an even more refined way by using the concept of micro-errors put forward by Seligman, Giovannetti, Sestito, and Libon (2013). Micro-errors are defined as “inefficient but not overtly erroneous execution of task steps” that may include “extra actions, imperfect sequencing not meeting commission error criteria, or microslips” (p. 100). Seligman et al. define microslips as the initiation of an overt error that is not completed. This classification of errors may be much more sensitive to capture the subtle difficulties in functional performances in this population. But this approach should be further validated with MCI. This more sensitive classification of the types of errors would also assist clinicians to identify better the MCI subtypes and the concomitant executive or memory functions that are affected (e.g., errors in memory, such as forgetting the food in the oven, or errors in planning and sequencing, such as organizing each dish and when it is to be prepared).

Although this refined analysis of the patterns of errors is very promising in assessing MCI, most current studies on error analysis only consider task execution. However, it is recognized that performing an activity in everyday life involves four principle cognitive operations (Bottari, Dassa, Rainville, & Dutil, 2009): formulating a goal (e.g., preparing food), planning a solution to attain the goal (e.g., choosing to prepare spaghetti), carrying out the activity (i.e., executing all steps required to prepare the spaghetti), and verifying the attainment of the goal (i.e., verify that the meal was prepared as planned). To observe what the person can really do in everyday living and the types of errors that can occur, it is important that the evaluator consider all four of these cognitive operations and not complete requisite cognitive operations for the person being assessed. For example, the evaluator must not specify the tasks to be performed (i.e., formulate the goal for the person), give a detailed plan of the task (i.e., planning for the person), or mention the equipment to be used (i.e., elements for planning and execution of the task). Thus, the evaluation should use an unstructured approach by providing as little guidance as possible to allow for the observation of the impact of the disease on all aspects of IADL performance. Additionally, these four cognitive operations form a set of sequences that individuals can follow to manage complex or novel task completion (Bottari et al., 2009), which is identified as the most difficult for persons with MCI (Albert et al., 2011). Therefore, considering all aspects of executive functions in complex IADL tasks could be a very promising evaluation strategy

in MCI assessment. At the same time, future research is needed to examine the specific cognitive components that should be emphasized in a measure of functional performance for the MCI population, and should consider not only the contribution of executive functions but also of memory deficits.

To date, no tool found in the literature and used for patients with MCI meets all of these requirements: applied error analysis, the consideration of all operations related to executive functions, use of an unstructured approach, testing of complex IADL, and administered in a real-world setting. For example, even the error analyses and micro-error approach of the NAT from Giovannetti, Bettcher, Brennan, Libon, Kessler, et al. (2008) mainly concerns errors of execution. Also, the NAT uses a structured approach and all necessary objects are generally in sight, therefore guiding the participants into the different steps of the tasks (goal formulation and planning). Finally, this test is not performed in the person's home or community environment and the tasks used may not be significant for a person with MCI.

The IADL Profile is a potentially interesting tool for this population. Although it has only been extensively validated with the traumatic brain injury population (Bottari et al., 2009; Bottari, Dassa, Rainville, & Dutil, 2010), a preliminary result in dementia shows great promise of the IADL Profile in the older population (Bier et al., 2012). This ecological performance-based measure of independence is administered in a person's home and community environments. It aims to establish whether the person's main difficulties in IADL

pertain to one of four task-related operations (formulate goal, plan, carry out, or verify attainment of goal) that particularly consider executive function deficits. It also considers many macro- and micro- indicators of performance, such as time of completion and types of errors, which allows for the identification of difficulties that might be related to other cognitive deficits (Bier et al., 2012). This enables it to achieve the intended goal of creating an evaluation scenario that comes close to the requirements of a complex everyday life situation that explicitly taps into executive functions. This tool framed the evaluation context so as to require the simultaneous planning of the full series of embedded tasks necessary to attain the ultimate goal of hosting a meal for unexpected guests: dressing to go outdoors, going to the grocery store, shopping for food, preparing a hot meal, having a meal with guests, and cleaning up after the meal (Bottari et al., 2010). Two other more structured tasks are also evaluated: making a budget and obtaining information. One of the unique challenges of the IADL Profile—what makes this tool distinct from others—is its non-structured approach. To this end, specific instructions are kept at a minimum and unsolicited assistance is not given unless it is judged necessary. When participants are unable to pursue the tasks, they are given graded assistance. In this manner, the performance is graded on a continuum of independence (e.g., totally independent vs. assistance required to complete the task) and independence scores provide information on elements such as the person's response to cues. Further studies are needed to validate this

measurement tool with the MCI clientele.

Another interesting performance-based tool to assess functional performance in MCI is the Management of Everyday Technology (META). This tool has been validated with MCI but has not yet been translated into English. It was developed by occupational therapists and it assesses the ability to manage technology in everyday life. The META consists of 10 items assessing observable performance skills that are essential to the ability to manage everyday technology (Nygård & Starkhammar, 2007). Although this test is structured, it does consider complex and difficult tasks and thus takes into consideration some operations of executive functions.

Although with the current available evidence no specific recommendations can be proposed to clinicians regarding a specific tool to be used to assess functional performance in MCI, research in this area is beginning to identify the elements that should be taken into consideration when assessing this clientele (i.e., applying error analysis during task performance rather than level of impairment, and considering all cognitive operations necessary for independent living).

Limitations

Every effort was made to ensure that our search encompassed all of the functional measurement tools that have been validated with the MCI population. Yet, it is possible that our search missed some instruments or studies on psychometric properties. The conclusions drawn

from this review are limited to the tools studied.

This review did not consider the quality of the content development process or psychometric evaluation while compiling the details on each tool. No extensive search was carried out to identify any unpublished studies, suggesting this scoping review may be affected by publication bias. In addition, the MCI subtypes recruited for each study also varied. This is an important consideration because different cognitive deficits may impact on functional performance in different ways. Therefore, the results of this study should be interpreted with caution and specific consideration should be given to the MCI subtypes included in each study.

Conclusion and Future Directions

Considering that occupational therapists are involved in assessing clients' functional performance in real-life situations, it is important that they incorporate functional measurement tools with the MCI clientele. While a specific measurement tool cannot be recommended at this time, researchers are pointing out the important components necessary for a measure with this population. Future research should establish operationalization criteria for functional impairment in MCI as well as rates of functional decline in MCI, norms of instruments, and cutoff points. By being able to differentiate pathological decline from the decline seen in normal aging, we can better identify those in need of intervention.

References

- Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., . . . Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's and Dementia*, 7(3), 270-279. <http://dx.doi.org/10.1016/j.jalz.2011.03.008>
- Armstrong, R., Hall, B. J., Doyle, J., & Waters, E. (2011). Cochrane Update 'Scoping the scope' of a Cochrane review. *Journal of Public Health*, 33(1), 147-150. <http://dx.doi.org/10.1093/pubmed/fdr015>
- Bangen, K. J., Jak, A. J., Schiehser, D. M., Delano-Wood, L., Tuminello, E., Han, S., . . . Bondi, M. W. (2010). Complex activities of daily living vary by mild cognitive impairment subtype. *Journal of the International Neuropsychological Society*, 16(4), 630-639. <http://dx.doi.org/10.1017/S1355617710000330>
- Bier, N., Bottari, C., Hudon, C., Joubert, S., Paquette, G., & Macoir, J. (2012). The impact of semantic dementia on everyday actions: Evidence from an ecological study. *Journal of International Neuropsychology*, 19(2), 162-172. <http://dx.doi.org/10.1017/S1355617712001105>
- Binegar, D. L., Hynan, L. S., Lacritz, L. H., Weiner, M. F., & Cullum, M. (2009). Can a direct IADL measure detect deficits in persons with MCI? *Current Alzheimer Research*, 6(1), 48-51. <http://dx.doi.org/10.2174/156720509787313880>
- Bottari, C., Dassa, C., Rainville, C., & Dutil, E. (2009). The factorial validity and internal consistency of the Instrumental Activities of Daily Living Profile in individuals with a traumatic brain injury. *Neuropsychological Rehabilitation*, 19(2), 177-207. <http://dx.doi.org/10.1080/09602010802188435>
- Bottari, C. L., Dassa, C., Rainville, C. M., & Dutil, E. (2010). The IADL profile: Development, content validity, intra- and interrater agreement. *Canadian Journal of Occupational Therapy*, 77(2), 90-100. <http://dx.doi.org/10.2182/cjot.2010.77.2.5>
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York, NY: Holt, Rinehart, and Winston.
- Cullum, C. M., Saine, K., Chan, L. D., Martin-Cook, K., Gray, K. F., & Weiner, M. F. (2001). Performance-based instrument to assess functional capacity in dementia: The Texas

- Functional Living Scale. *Neuropsychiatry, Neuropsychology and Behavioral Neurology*, 14(2), 103-108. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11417663>
- De Vriendt, P., Gorus, E., Cornelis, E., Velghe, A., Petrovic, M., & Mets, T. (2012). The process of decline in advanced activities of daily living: A qualitative explorative study in mild cognitive impairment. *International Psychogeriatrics*, 24(6), 974-986. <http://dx.doi.org/10.1017/S1041610211002766>
- Giovannetti, T., Bettcher, B. M., Brennan, L., Libon, D. J., Burke, M., Duey, K., . . . Wambach, D. (2008). Characterization of everyday functioning in mild cognitive impairment: A direct assessment approach. *Dementia and Geriatric Cognitive Disorders*, 25(4), 359-365. <http://dx.doi.org/10.1159/000121005>
- Giovannetti, T., Bettcher, B. M., Brennan, L., Libon, D. J., Kessler, R. K., & Duey, K. (2008). Coffee with jelly or unbuttered toast: Commissions and omissions are dissociable aspects of everyday action impairment in Alzheimer's disease. *Neuropsychology*, 22(2), 235-245. <http://dx.doi.org/10.1037/0894-4105.22.2.235>
- Goldberg, T. E., Koppel, J., Keehlisen, L., Christen, E., Dreses-Werringloer, U., Conejero-Goldberg, C., . . . Davies, P. (2010). Performance-based measures of everyday function in mild cognitive impairment. *American Journal of Psychiatry*, 167(7), 845-853. <http://dx.doi.org/10.1176/appi.ajp.2010.09050692>
- Gomar, J. J., Harvey, P. D., Bobes-Bascaran, M. T., Davies, P., & Goldberg, T. E. (2011). Development and cross-validation of the UPSA Short Form for the performance-based functional assessment of patients with mild cognitive impairment and Alzheimer disease. *American Journal of Geriatric Psychiatry*, 19(11), 915-922. <http://dx.doi.org/10.1097/JGP.0b013e3182011846>
- Griffith, H. R., Belue, K., Sicola, A., Krzywanski, S., Zamrini, E., Harrell, L., & Marson, D. C. (2003). Impaired financial abilities in mild cognitive impairment: A direct assessment approach. *Neurology*, 60(3), 449-457. <http://dx.doi.org/10.1212/WNL.60.3.449>
- Kounti, F., Tsolaki, M., & Kiosseoglou, G. (2006). Functional Cognitive Assessment Scale (FUCAS): A new scale to assess executive cognitive function in daily life activities in patients with dementia and mild cognitive impairment. *Human Psychopharmacology: Clinical and Experimental*, 21(5), 305-311. <http://dx.doi.org/10.1002/hup.772>
- Law, M., Baum, C. M., & Dunn, W. (2005). *Measuring occupational performance: Supporting*

- best practice in occupational therapy* (2nd ed). Thorofare, NJ: Slack.
- Loeb, P. A. (1996). *ILS: Independent Living Scales manual*. San Antonio, TX: Psychological Corp.
- Loewenstein, D. A., Amigo, E., Duara, R., Guterman, A., Hurwitz, D., Berkowitz, N., . . . Eisdorfer, C. (1989). A new scale for the assessment of functional status in Alzheimer's disease and related disorders. *Journal of Gerontology*, *44*(4), 114–121.
<http://dx.doi.org/10.1093/geronj/44.4.P114>
- Malinowsky, C., Almkvist, O., Nygård, L., & Kottorp, A. (2011). Individual variability and environmental characteristics influence older adults' abilities to manage everyday technology. *International Psychogeriatrics*, *24*(3), 484-495.
<http://dx.doi.org/10.1017/S1041610211002092>
- McDougall, G. J., Becker, H., Vaughan, P. W., Acee, T. W., & Delville, C. L. (2009). The Revised Direct Assessment of Functional Status for independent older adults. *The Gerontologist*, *50*(3), 363-370. <http://dx.doi.org/10.1093/geront/gnp139>
- Nygård, L. & Starkhammar, S. (2007). The use of everyday technology by people with dementia living alone: Mapping out the difficulties. *Aging & Mental Health*, *11*(2), 144-155.
- Owsley, C., McGwin, G., Sloane, M. E., Stalvey, B. T., & Wells, J. (2001). Timed instrumental activities of daily living tasks: Relationship to visual function in older adults. *Optometry and Vision Science*, *78*(5), 350-359.
<http://dx.doi.org/10.1097/00006324-200105000-00019>
- Owsley, C., Sloane, M., McGwin, J. G., & Ball, K. (2002). Timed instrumental activities of daily living tasks: relationship to cognitive function and everyday performance assessments in older adults. *Gerontology*, *48*, 254-265. <http://dx.doi.org/10.1159/000058360>
- Patterson, T. L., Goldman, S., McKibbin, C. L., Hughs, T., & Jeste, D. V. (2001). UCSD Performance-Based Skills Assessment: Development of a new measure of everyday functioning for severely mentally ill adults. *Schizophrenia Bulletin*, *27*(2), 235–245.
<http://dx.doi.org/10.1093/oxfordjournals.schbul.a006870>
- Pereira, F. S., Yassuda, M. S., Oliveira, A. M., & Forlenza, O. V. (2008). Executive dysfunction correlates with impaired functional status in older adults with varying degrees of cognitive impairment. *International Psychogeriatrics*, *20*(6), 1104-1115.
<http://dx.doi.org/10.1017/S1041610208007631>

- Petersen, R. C., Caracciolo, B., Brayne, C., Gauthier, S., Jelic, V., & Fratiglioni, L. (2014). Mild cognitive impairment: A concept in evolution. *Journal of Internal Medicine*, 275(3), 214-228. <http://dx.doi.org/10.1111/joim.12190>
- Portney, L. G., & Watkins, M. P. (2009). *Foundations of clinical research: Applications to practice* (3rd edition). Upper Saddle River, NJ: Prentice Hall.
- Poulin, V., Korner-Bitensky, N., & Dawson, D. R. (2013). Stroke-specific executive function assessment: A literature review of performance-based tools. *Australian Occupational Therapy Journal*, 60(1), 3-19. <http://dx.doi.org/10.1111/1440-1630.12024>
- Provencher, V., Demers, L., Gagnon, L., & G elinas, I. (2012). Impact of familiar and unfamiliar settings on cooking task assessments in frail older adults with poor and preserved executive functions. *International Psychogeriatrics*, 24(5), 775-783. <http://dx.doi.org/10.1017/S104161021100216X>
- Rosenberg, L., Kottorp, A., Winblad, B., & Nyg ard, L. (2009). Perceived difficulty in everyday technology use among older adults with or without cognitive deficits. *Scandinavian Journal of Occupational Therapy*, 16(4), 216-226. <http://dx.doi.org/10.3109/11038120802684299>
- Schmitter-Edgecombe, M., McAlister, C., & Weakley, A. (2012). Naturalistic assessment of everyday functioning in individuals with mild cognitive impairment: The day-out task. *Neuropsychology*, 26(5), 631-641. <http://dx.doi.org/10.1037/a0029352.suppl>
- Schmitter-Edgecombe, M., Woo, E., & Greeley, D. R. (2009). Characterizing multiple memory deficits and their relation to everyday functioning in individuals with mild cognitive impairment. *Neuropsychology*, 23(2), 168-177. <http://dx.doi.org/10.1037/a0014186>
- Schwartz, M. F., Segal, M. E., Veramonti, T., Ferraro, M., & Buxbaum, L. J. (2002). The Naturalistic Action Test: A standardized assessment for everyday-action impairment. *Neuropsychological Rehabilitation*, 12(4), 311-339. <http://dx.doi.org/10.1080/09602010244000084>
- Seligman, S. C., Giovannetti, T., Sestito, J., & Libon, D. J. (2013). A new approach to the characterization of subtle errors in everyday action: Implications for mild cognitive impairment. *The Clinical Neuropsychologist*, 28(1), 97-115. <http://dx.doi.org/10.1080/13854046.2013.852624>

- Streiner, D. L., & Norman, G. R. (2003). *Health measurement scales: a practical guide to their development and use* (3rd edition). Toronto, Canada: Oxford University Press.
- Triebel, K. L., Martin, R., Griffith, H. R., Marceaux, J., Okonkwo, O. C., Harrell, L., . . . Marson, D. C. (2009). Declining financial capacity in mild cognitive impairment: A 1-year longitudinal study. *Neurology*, 73(12), 928-934.
<http://dx.doi.org/10.1212/WNL.0b013e3181b87971>
- Wadley, V. G., Okonkwo, O., Crowe, M., & Ross-Meadows, L. A. (2008). Mild cognitive impairment and everyday function: Evidence of reduced speed in performing instrumental activities of daily living. *American Journal of Geriatric Psychiatry*, 16(5), 416-424. <http://dx.doi.org/10.1097/01.JGP.0000310780.04465.13>
- Winblad, B., Palmer, K., Kivipelto, M., Jelic, V., Fratiglioni, L., Wahlund, L.-O., . . . Petersen, R. C. (2004). Mild cognitive impairment—beyond controversies, towards a consensus: Report of the International Working Group on Mild Cognitive Impairment. *Journal of Internal Medicine*, 256(3), 240–246. <http://dx.doi.org/10.1111/j.1365-2796.2004.01380.x>

Appendix A

Summary of Performance-Based Tools

Assessment and description	Study population Environment	Reliability	Validity	Clinical utility
<p>Direct Assessment of Functional Status Revised (DAFS-R) (McDougall et al., 2009) Assesses functional status in older adults using tasks that simulate four everyday tasks: communication, financial skills, medication management*, and shopping. *Not studied with MCI population</p>	<p>Amnesic and nonamnesic (multiple domain) Lab-based using real-world materials</p>	<p>Test-retest No evidence with MCI Inter-rater No evidence in MCI Internal consistency No evidence in MCI</p>	<p>Content validity Adequate: Established by literature review, testing, and consultation with geriatricians (Loewenstein et al., 1989). Construct validity Adequate evidence with MCI <u>Known groups</u> Controls showed higher performance than MCI and AD ($p = 0.009$ and $p < 0.001$, respectively), and MCI higher than AD ($p < 0.001$) (Pereira et al., 2008). <u>Convergent validity</u> Moderate correlation with EXIT25 ($r = -0.513$; $p < 0.001$) (Pereira et al., 2008). Criterion validity No evidence in MCI Responsiveness Minimal evidence: Sensitive to functional decline after 1 year and was useful to establish longitudinal patterns of deterioration. Floor and ceiling effects No evidence with MCI</p>	<p>Testing situation Seated behind a table and moving around the room. Time 30-35 minutes (Loewenstein et al., 1989). Therapist training Administrator should be familiar with administration of standardized assessments. Cost and ordering information See Loewenstein et al. (1989). Scoring For each subtask, a score of 1 is given for a correct answer or 0 for an incorrect answer.</p>
<p>Day-Out Task (DOT) (Schmitter-Edgecombe et al., 2012) Assesses naturalistic tasks that require multitasking in a real-world setting (i.e., gather correct change from organizer on dining room table for the bus ride, take motion sickness medication “Dramamine” located in kitchen cupboard just prior to leaving the house, plan bus route).</p>	<p>Amnesic and nonamnesic (both single and multiple domain) Simulated real world</p>	<p>Test-retest No evidence with MCI Inter-rater Excellent 96.92% for subtask accuracy scores and 99.27% for task sequencing. Internal Consistency No evidence with MCI</p>	<p>Content validity Minimal evidence: The criteria for choosing the specific sub tasks have not been clearly mentioned—the assessment was developed primarily for this study (Schmitter-Edgecombe et al., 2012). The emphasis is on multi-tasking and interweaving during tasks so the assessment can be done efficiently. Construct validity Minimal evidence with MCI <u>Known groups</u> The MCI group required more time to complete the DOT relative to normal controls ($p = 0.01$). They also demonstrated an overall poorer task accuracy relative to controls ($p < 0.01$), performing more subtasks incompletely and inaccurately (Schmitter-Edgecombe et al., 2012). <u>Convergent validity</u> No evidence with MCI Criterion validity No evidence with MCI Responsiveness No evidence with MCI Floor and ceiling effects</p>	<p>Testing situation Walking around the room. Time Depending upon participant. Therapist training Not reported Cost and ordering information See Schmitter-Edgecombe, Woo, and Greeley (2009) and its supplement. Scoring For each subtask, the following scores are given: 1 – complete/efficient, 2 – complete/inefficient, 3 – incomplete/inaccurate, and 4 – never attempted.</p>

<p>Financial Capacity Instrument (FCI) (Griffith et al., 2003) Assesses nine tasks of financial capacity: basic monetary skills, financial conceptual knowledge, cash transaction, check book management, bank statement management, financial judgment, bill payment, knowledge of personal assets/estate arrangement (requires a collateral report)*, and investment decision making. *Not studied with MCI population</p>	<p>Amnestic Lab-based using real-world material</p>	<p>Test-retest No evidence with MCI Inter-rater No evidence with MCI Internal consistency No evidence with MCI</p>	<p>No evidence with MCI Content validity Adequate: Based on conceptual model of the financial capacity construct-revised model presented in Griffith et al. (2003). Addition of a new domain has been made on Investment Decision Making, which was initially a part of Domain 6 on Financial Judgment. Reconceptualization was completed as the original construct of financial judgment was not reflective of one's ability to recognize and avoid different financial frauds. Construct validity Minimal evidence with MCI <u>Known groups</u> Adjusting for group differences based on age and prior financial experience, the control participants performed significantly better than the MCI groups on all domains and total scores although Domain 8 was not analyzed (Triebel et al., 2009). MCI participants demonstrated impairments in FCI domains of conceptual knowledge, bank statement management, and bill payment, as well as overall financial capacity. The control and MCI groups performed significantly better than patients with AD on most financial capacity and cognitive measures when Domains 1-7 were analyzed (Griffith et al., 2003). <u>Convergent validity</u> No evidence with MCI Criterion validity No evidence with MCI Responsiveness Minimal evidence with MCI At 1-year follow-up, individuals with MCI showed significantly greater decline than controls. MCI patients who converted to AD demonstrated significantly lower scores at baseline than controls and MCI patients who did not convert (Triebel et al., 2009). Domain 8 was not counted in global score. Floor and ceiling effects No evidence with MCI</p>	<p>Testing situation Seated behind a table. Time Not reported with MCI patients. Therapist training Administered by trained staff with experience testing persons with memory disorders and dementia. Cost and ordering information Not reported Scoring Scoring is done according to a standardized scoring system. The test score is the sum of the task scores in each domain. Partial task scores can be given to patients with amnesia or aphasia.</p>
<p>Functional Cognitive Assessment Scale (FUCAS) (Kounti et al., 2006) This is a 13-item scale that requires patients to execute six different tasks of daily life: telephone communication,</p>	<p>Not specified Lab-based using real-world material</p>	<p>Test-retest No evidence with MCI Inter-rater Excellent $r = 0.997$ (reliability of total scores of 30 participants</p>	<p>Content validity Minimal evidence: The source of items has not been made explicit. However, the authors have made some assumptions which underlie the construction of this scale: daily life activities are problem-solving situations involving recognizing the problem, planning, and executing the solution to solve the problem.</p>	<p>Testing situation Seated behind a table. Time Not reported Therapist training Not specified Cost and ordering information Not reported Scoring</p>

<p>shopping, orientation in place, taking medication, personal hygiene, and clothing.</p>		<p>-10 controls, 10 MCI, 10 with dementia – by 2 raters). For sub-scores, <i>r</i> ranges from 0.983 to 1.000. Internal consistency Excellent Cronbach’s α ranged from 0.89 to 0.92 for all items and sub-scores (<i>n</i> = 75).</p>	<p>Construct validity Minimal evidence with MCI <u>Known groups</u> FUCAS is able to sufficiently discriminate patients with MCI from those with moderate to severe dementia. Two parameters of executive function (working memory and goal maintenance) classified MCI and mild dementia with statistical significance ($p < 0.0001$). Twenty percent of MCI individuals and 37% of dementia patients were correctly identified. <u>Convergent validity</u> No evidence with MCI Criterion validity Adequate evidence with MCI <u>Concurrent validity</u> Total scores correlate ($p < 0.01$) with CAMCOG ($r = 0.784$), MMSE ($r = 0.622$), and FRSSD ($r = 0.781$). Subscales significantly correlate ($p < 0.01$) with corresponding subscales of the CAMCOG, MMSE, and FRSSD at moderate to high levels. Responsiveness No evidence with MCI Floor and ceiling effects No evidence with MCI</p>	<p>A score of 1 indicates no problem with the executive parameter examined in a certain activity, 2 indicates a mild-to-moderate problem, and 3 indicates a severe problem. Sub-scores of performance for each executive parameter which reflects the total patient’s performance in the six activities can be obtained.</p>
<p>Independent Living Scale (ILS)* (Bangen et al., 2010) This measure is comprised of 68 items across 5 subscales* (memory/orientation, managing money, managing home and transportation, health and safety, and social). Items include verbal questions and performance-based tasks (Loeb, 1996). *Only managing money and health and safety have been studied with MCI population</p>	<p>Amnesic and nonamnesic (both single and multiple domain) Lab-based using real-world material</p>	<p>Test-retest No evidence with MCI Inter-rater No evidence with MCI Internal consistency No evidence with MCI</p>	<p>Content validity Adequate: It was established through literature review. ILS was derived from the Community Competence Scale and included only those items that demonstrated adequate psychometric properties. Construct validity Adequate evidence with MCI <u>Known groups</u> The managing money subscale was able to significantly discriminate between amnesic MCI individuals and normal controls. It was seen that the amnesic MCI group had performed significantly worse relative to normal controls ($p < .001$). The health and safety subscale demonstrated a trend toward decreased performance by the nonamnesic MCI group relative to normal controls ($p = .04$) (Bangen et al., 2010). <u>Convergent validity</u> Global cognitive function (measured by the Dementia Rating Scale Total T-score) was significantly associated with two subtasks of ILS: managing money ($r = 0.48$; $p < .001$) and health and safety ($r = 0.29$; $p = 0.002$) subscales. Criterion validity No evidence with MCI Responsiveness No evidence with MCI</p>	<p>Testing situation Seated behind a table. Time Not reported for the two subtasks tested. Therapist training Test administrators should have knowledge or experience with the MCI population and should be familiar with administration of standardized assessments. Cost and ordering information See Bangen et al. (2010) or online at www.pearsonclinical.com. Scoring A raw score is obtained for each subscale. Standard scores are derived from the raw scores by using appropriate tables in the manual.</p>

			<p>Floor and ceiling effects No evidence with MCI</p>	
<p>Naturalistic Action Test (NAT) (Giovannetti, Bettcher, Brennan, Libon, Kessler, et al., 2008) Analyses execution of task steps through accomplishment and error. It includes three items: preparing a toast with butter and jelly and coffee with cream and sugar; wrapping a gift while salient distractor objects (e.g., garden shears, electric tape) are included on the tabletop; and packing a lunchbox with a sandwich, snack, and drink, and a schoolbag with supplies for school, while several necessary objects (e.g., thermos lids) are stored out of view in a drawer containing potentially distracting objects (e.g., spatula, thread, etc.).</p>	<p>Amnestic and nonamnestic (both single and multiple domain)</p> <p>Simulated real world</p>	<p>Test-retest No evidence with MCI Inter-rater No evidence with MCI Internal consistency No evidence with MCI</p>	<p>Content validity Adequate: Items were derived from the Multi-Level Action Test (MLAT). They were chosen according to the level of difficulty. This was followed by identification and testing of three items which varied significantly in terms of standardized error rates in order to create a short version of the assessment scale. Construct validity Minimal evidence with MCI <u>Known groups</u> In terms of overall impairment, controls had significantly better performance than the MCI ($p < 0.01$) and the mild AD groups ($p < 0.01$). Also, individuals with MCI were found to have significantly better performance than the AD group ($p < 0.01$). <u>Convergent validity</u> No evidence with MCI Criterion validity No evidence with MCI Responsiveness No evidence with MCI Floor and ceiling effects No evidence with MCI</p>	<p>Testing situation Seated behind a table or standing at a counter. Time Depending upon the participant. Therapist training Not specified Cost and ordering information See Giovannetti Bettcher, Brennan, Libon, Kessler, et al. (2008). Scoring Individuals are scored on: accomplishment of each subtask (e.g., bread toasted, sandwich made) and error score (i.e., toasts more than one slice of bread). Each item has a particular number of steps to be performed. Thus, the accomplishment score is the percentage of completion of required steps (with or without error).</p>
<p>Texas Functional Living Scale (TFLS) (Binegar et al., 2009) TFLS is a performance-based measure of functional abilities. It is comprised of 21 items organized into 5 subscales: dressing (e.g., put on jacket), time (e.g., state time on clock, set clock), money (e.g., count money, make change), communication (e.g., address envelopes, call home), and memory (e.g., recall payee of checks, recall amount of checks).</p>	<p>Amnestic (both single and multiple domain)</p> <p>Lab-based using real-world material</p>	<p>Test-retest No evidence with MCI Inter-rater No evidence with MCI Internal consistency No evidence with MCI</p>	<p>Content validity Adequate: The scale was formed following a thorough review of existing performed-based measures of instrumental activities of daily living skills. Items evaluating a range of cognitive-behavioral abilities which could be more sensitive to Alzheimer's disease (in early stages) were gathered for the development of this measure (Cullum et al., 2001). Construct validity Minimal evidence with MCI <u>Known groups</u> A cutoff score of 48 represents the highest combination of sensitivity and specificity in predicting diagnosis (sensitivity of 56.7% and a specificity of 63.3%, with an accurate group classification of 60%; Binegar et al., 2009). <u>Convergent validity</u> No evidence in MCI Criterion validity Minimal evidence with MCI <u>Concurrent validity</u> No significant association was found between TFLS and MMSE scores in individuals with MCI ($p = 0.253$). A</p>	<p>Testing situation Seated behind a table. Time 15-20 minutes (Cullum et al., 2001). Therapist training Administrator should be familiar with administration of standardized assessments. Cost and ordering information See Binegar et al. (2009) or online at www.pearsonclinical.com. Scoring The maximum possible score is 52 points, with higher scores indicating better performance. The point values vary across functional tasks. For example, a person who can point out the date correctly on a one-year calendar will gain 3 points. If he or she identifies the correct week but not the correct day as required, they acquire 2</p>

			<p>moderate correlation was reported between TFLS and MMSE total scores when both MCI and normal controls groups were combined ($p = 0.019$; Binengar et al., 2009).</p> <p>Responsiveness No evidence with MCI</p> <p>Floor and ceiling effects No evidence with MCI</p>	<p>points. On identification of the correct month only, 1 point is earned, hence the assessment scale captures varying levels of functioning.</p>
<p>Timed Instrumental Activities of Daily Living (TIADL) (Wadley et al., 2008) Assesses speed and accuracy of five functional tasks commonly encountered in everyday life: telephone use, nutrition evaluation, financial abilities, grocery shopping, and medication management.</p>	<p>Amnesic</p> <p>Lab-based using real-world material</p>	<p>Test-retest No evidence with MCI</p> <p>Inter-rater No evidence with MCI</p> <p>Internal consistency No evidence with MCI</p>	<p>Content validity Adequate: Four of the five tasks have been adapted from Owsley, McGwin, Sloane, Stalvey, and Wells (2001). Criteria for selection of above tasks included: (a) functional assessments which are fundamentally required for independent living irrespective of gender, socioeconomic status, and ethnic origin; (b) tasks requiring a strong cognition and decline which could hinder the independence; and (c) tasks which are brief and are amenable to correct timing where the task administration can be standardized.</p> <p>Construct validity Minimal evidence with MCI</p> <p><u>Known groups</u> For overall accuracy scores, unadjusted odd ratios (OR) revealed that MCI individuals were 2.29 times more likely than controls to make errors during the task performance. However, the effect disappeared when depression was taken into account. Still, on examination of specific tasks, there was a significant association between MCI classification and error status only for the grocery shopping task. It was found that the MCI patients were 5.27 times more likely than the controls to commit errors such as locating a distractor item rather than the target item on this task (Wadley et al., 2008). With adjustment of depression, the effect remained statistically significant.</p> <p><u>Convergent validity</u> No evidence with MCI</p> <p>Criterion validity Adequate evidence with MCI</p> <p><u>Concurrent validity</u> In the MCI group, individuals with completion time deficits ($N = 36$) had worse global cognitive function (mean DRS score = 131.58, $SD = 6.70$) than those with no speed deficit (mean DRS score = 136.42, $SD = 5.66$), ($t(46) = 2.241$, $p=0.030$). Similarly, MCI participants with accuracy deficits ($N = 26$) had worse global cognitive functioning (mean DRS score = 130.27, $SD = 5.95$) than those with no errors</p>	<p>Testing situation Seated behind a table and moving around the room.</p> <p>Time Average of 15 minutes.</p> <p>Therapist training Trained interviewer/tester.</p> <p>Cost and ordering information See Owsley, Sloane, McGwin, and Ball (2002).</p> <p>Scoring Accuracy scores: 1 = Completed within the time limit with no errors; 2 = Completed within the time limit with minor errors; 3 = Not completed within the time limit or completed with major errors. Wadley et al. (2008) used dichotomous scores (as only few MCI individuals made major errors) to rate the tasks: 1 = Completed within the time limit with no error and 2 = Completed with errors or not within the time limit. Each task had to be completed within a predetermined time period. If the participant failed to complete the task within the given time frame, the testing for that item stopped.</p>

			(mean DRS score = 135.77, SD = 6.51), ($t(46) = 3.058, p = 0.004$; Wadley et al., 2008). Responsiveness No evidence with MCI Floor and ceiling effects No evidence with MCI	
University of California, San Diego Performance-Based Skills Assessment (UPSA) - Short version (Gomar et al., 2011) Assesses patient performance in 27 items divided into four functional domains: comprehension/planning (e.g., planning a trip to the zoo), financial procedures (e.g., counting money, writing a check), communication (e.g., call directory assistance), and transportation (e.g., taking a bus).	Amnesic (both single and multiple domain) Lab-based using real-world material	Test-retest No evidence with MCI Inter-rater No evidence with MCI Internal consistency No evidence with MCI	Content validity Adequate: Developed with inputs from patients, health care professionals, published reports, and review of previously developed instruments. Construct validity Minimal evidence with MCI <u>Known groups</u> The UPSA differentiates between healthy controls and MCI. The probability that a participant with MCI would have a lower UPSA score was 0.84. At a cutoff of $p = 0.50$, sensitivity for identification of healthy participants was 0.88 and specificity = 0.58. <u>Convergent validity</u> No evidence with MCI Criterion validity Adequate evidence with MCI <u>Concurrent validity</u> Significant correlation of the short version with the Alzheimer's Disease Cooperative Study–Activities of Daily Living Inventory (ADCS-ADL), Spearman's rank order method ($p = 0.63, p = 0.0001$; Goldberg et al., 2010). In addition, the short version was significantly correlated with the full UPSA scale in all the groups examined: 0.86 for healthy controls, 0.87 for MCI, and 0.88 for AD (Gomar et al., 2011). Responsiveness No evidence with MCI Floor and ceiling effects Not generally prone to ceiling effects in healthy participants or to floor effects in AD participants (Goldberg et al., 2010).	Testing situation Seated behind a table. Time 10-15 minutes. Therapist training Administered by trained non-clinicians. Training requires several hours (Patterson, Goldman, McKibbin, Hughs, & Jeste, 2001). Cost and ordering information Available in Patterson et al. (2001). Scoring Depending on the subtask assessed, a score of 1 is given to correct answers and 0 to incorrect; or a score of 2 is given to correct answers and 0 to incorrect. Total scores for each subscale are calculated by transforming raw scores into a 0-to-10 scale, yielding comparable scores on each scale.

Appendix B

Evaluation Criteria for Psychometric Properties (Poulin, Korner-Bitensky, & Dawson, 2013)

Standards for rating reliability:

Internal consistency (split-half or Cronbach's alpha statistic):

Excellent: ≥ 0.80 ; Adequate: 0.70–0.79; Poor: < 0.70

Test-retest and inter-rater (correlation coefficient or kappa statistic):

Excellent: ≥ 0.75 ; Adequate: 0.4–0.74; Poor: < 0.40

Standards for rating validity and responsiveness:

Excellent: Most major forms of testing reported

Adequate: Several types of testing or several studies reported

Poor/Minimal evidence: Minimal information reported and/or evidence from pilot studies

No evidence: No studies and/or no information available

Conflicting: Two or more studies showing different findings