ANALYSIS OF PRACTICE EFFECTS ACROSS COGNITIVE DOMAINS IN MILD COGNITIVE IMPAIRMENT

APPROVED BY SUPERVISORY COMMITTEE

Laura Lacritz, Ph.D. ABPP

Munro Cullum, Ph.D. ABPP

Kristin Martin-Cook, M.S.

DEDICATION

I would like to thank the members of my Graduate Committee for their assistance, my Committee Chair who has had faith in my abilities and provided undeniable emotional and academic support to see the light at the end of the tunnel, my significant other for motivation, and my friends and family for their encouragement.

ANALYSIS OF PRACTICE EFFECTS ACROSS COGNITIVE DOMAINS IN MILD COGNITIVE IMPAIRMENT

by

SHRUTHI RAO

THESIS

Presented to the Faculty of the School of Health Professions

The University of Texas Southwestern Medical Center

Dallas, Texas

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF REHABILITATION COUNSELING

Abstract

Serial assessments provide clinically useful information about progression of a disease. Since individuals with mild cognitive impairment are less likely to show decline in cognitive areas other than memory, it is important to analyze which domains are more susceptible to practice effects than others.

The appearance of practice effects in serial assessments is a common challenge for clinicians interpreting neuropsychological tests. Detecting true change can be altered by factors such as test intervals, standardization procedures, alternate forms, respondent characteristics, and cognitive domains impaired in a clinical population. Some cognitive domains such as learning, memory, and executive functioning are known to be more susceptible to practice effects than others such as processing speed, attention, and language. Normal adults are also shown to have greater practice effects over multiple exposures than disease populations.

The review supports the claim that healthy adult individuals are more likely to improve and show greater practice effects during serial assessments than clinical populations. In patients with mild cognitive impairment (MCI), domains that rely on recall and learning test rules like learning, memory, and executive functioning tend to be more susceptible to practice effects than crystallized and skill - based domains such as language, processing speed, and attention.

Future research should focus on developing reliable change indices for each cognitive domain and possibly each neuropsychological test to help provide a comparison and detect true change in test scores. Studies should also focus on developing alternate equivalent forms, which would help minimize practice effects across populations.

Keywords: Mild Cognitive Impairment, practice effects, cognitive domains, serial assessments

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AD– Alzheimer's disease

ADC- Alzheimer's Disease Center

BNT- Boston Naming Test

MCI– Mild Cognitive Impairment

RCI– Reliable Change Index

TMT – Trail Making Test

UDS- Uniform Data Set

WAIS – Wechsler Adult Intelligence Scale

WCST- Wisconsin Card Sorting Test

WMS– Wechsler Memory Scale

CHAPTER ONE

Statement of the Problem

Practice effects are an unavoidable phenomenon of serial testing in longitudinal clinical and research settings. Cognitively intact individuals are more susceptible to practice effects on tests that involve problem solving and have similar content or procedures to recall at each testing (McCaffrey et al., 2000). Various studies show that practice effects on some types of tasks are nearly absent in individuals with dementia or Alzheimer's disease (AD) (Cooper et al., 2001). However, less is known about practice effects in individuals with mild cognitive impairment (MCI), an intermediate stage between the decline of normal aging and dementia where individuals commonly present memory deficits. Since individuals with mild cognitive impairment are less likely to show decline in cognitive areas other than memory, they may have the ability to adapt their performance and retain the rules and items seen on repeated testing. Some studies show that varying forms and testing intervals can significantly decrease the likelihood that practice effects will influence results. However, ways to diminish practice effects within an MCI population have not been clearly explored in previous studies.

There is growing emphasis on clinical outcomes when test data serves as predictors of individual clinical outcomes (Chelune & Najm, 2001). The study of practice effects in individuals with mild cognitive impairment also holds significant clinical value. Practice effects could provide clinically useful information about progression of a disease in populations that are sensitive to test fluctuations, such as MCI. Interventions for clinical populations depend on the rate of disease progression. Determining an accurate diagnosis at each assessment is critical to developing an intervention program. This is crucial so that individuals who are affected by a family member or friend with a progressive clinical diagnosis have the proper information regarding the care of their loved ones. The need to properly assess and diagnose the clinical status of cognitively declining individuals leads to providing the proper treatment. In light of this, cognitive assessments are used as outcome endpoints in the evaluation of the efficacy of certain treatments and medications. These assessments should be reliable in providing accurate evaluation scores during serial testing.

Given that there are differences in the literature about the extent to which practice effects have been seen in the MCI patient population and normal controls, a review of current literature is conducted to compare tests of various cognitive domains administered over multiple exposures in a neuropsychological testing battery among this particular clinical population to determine if practice effects are more common in certain cognitive domains. The review seeks to compare cognitively intact individuals to those with mild cognitive impairment and evaluate the changes seen over the course of multiple test periods. Additionally, the review seeks to evaluate the best procedure methods to minimize practice effects in repeated testing and determine a course of best practice in serial assessments. The following chapter looks at comparing cognitive domains of MCI patients over multiple exposures, as this has not been previously done in a less studied clinical population.

CHAPTER TWO

Review of the Literature

Introduction to the Literature

Detecting cognitive changes in individuals with known or suspected neurological disorders is an important goal of neuropsychological testing. Repeated cognitive testing is useful for assessing change over time as a part of disease progression or in response to treatment. This may be particularly important when testing is used for early detection of a degenerative disease process such as mild cognitive impairment (MCI) or Alzheimer's disease (AD).

Serial assessments are used in a clinical setting for multiple reasons. Evaluators may use serial assessments to track disease progression of an affected population. Assessments may also be used to evaluate the effectiveness of an intervention over the course of months to years, or document recovery after brain injury. The use of repeated testing can help clinicians track changes over time. However, serial neuropsychological assessments bring about interpretive challenges. Complicating the interpretation of scores from repeat testing is determining whether or not differences are due to real change in the individual or by external factors unrelated to the disease, since no test is perfectly reliable. While people naturally fluctuate in their performance to some degree, as reflected by a test's standard error of measurement, some tests may be more sensitive to such fluctuations than others. Clinicians must determine when score differences represent actual change, considering effects of extraneous factors on performance.

While there are standardized statistical methods for comparing groups in terms of neuropsychological change on various tests, techniques for detecting significant change over multiple exposures are less developed. Chelune and Najm (2001) showed that test data serve as a predictor for clinical status and tracking disease progression in a quantitative way. Therefore, one method for determining accurate change scores would be to evaluate the scores over a period of time among patients who have been exposed to a test on a regular basis (e.g., annually). By determining average change scores on specific tests for a certain population, clinicians can evaluate the progression of a disease and provide appropriate interventions. The resulting change that could influence scores should be considered for patient populations that are susceptible to disease progression, such as Alzheimer's disease (Duff et al., 2007). Although Alzheimer's disease patients have the tendency to consistently deteriorate (Small et. al., 2000), it would be prudent to evaluate the change scores seen in patients that have been diagnosed with mild cognitive impairment. Studies (Peterson & Negash, 2008) have shown that not all patients diagnosed with MCI progress to dementia, however, and some cognitively improve on retest. Furthermore, some patients have received MCI diagnoses that were described as "incidental" which were derived from low baseline scores (Rotrou et al., 2005). Therefore, it is important to determine expected change in this particular disease population at retest. This way, the probable decline would be anticipated and monitored more closely when observing multiple testing periods.

Factors that Affect Change Scores

Changes in test scores over multiple testing sessions can be influenced by various factors. Practice effects are defined as improvements in cognitive test performance due to repeated evaluation with the same test materials, and have traditionally been viewed as sources of error (McCaffrey et al., 2000). The improvement in performance on tests that are due to a previous exposure to the same measure can bring about elevated scores that are not due to the true change in the individuals' cognitive ability (Kaufman, 1990) may alter diagnosis and resulting interventions. The following sections review various factors that could impact cognitive test scores.

Learned Performance

An individual's baseline performance may influence future test scores. Studies conducted by Rapport and Axelrod (1997) and Rapport, Brines, and Theisen (1997) determined that higher IQ scores initially among individuals resulted in greater susceptibility to practice effects upon repeated testing. This could be due to the fact that high functioning individuals may be better able to retain performance rules and test items that could enhance test scores after the first administration. Such advantages may be enhanced by the use of the same test form at each administration. When testing the effects of repeated administration of verbal and non-verbal memory tests over the course of 8 weeks, researchers found that healthy participants greatly benefited from using the same test form at each testing period (Benedict & Zgaljardic, 1998).

Testing Intervals

Based on previous literature, it is common to assume that practice effect declines with longer testing intervals. Theisen, Rapport, Axelrod, and Brines (1998) determined that giving a two-week test-retest interval between each testing session for normal control participants increased the likelihood of higher scores on following testing sessions on verbal memory tests (i.e., WMS Logical Memory and Verbal Paired Associates) in the first retest session, whereas less significant changes were seen in the third and four sessions. The magnitude of increased scores was the greatest between the first and second testing session and slightly smaller (and not significant) between the third and the fourth sessions. The study claimed that retest scores are assumed to be the highest during shorter intervals and should decrease with time. However, Darby et al (2002) reported an absence of practice effects on computerized cognitive tasks involving memory and executive functioning tests that were repeated within in a 24-hour period among patients with MCI. On the other hand, some studies (Van der Elst et al., 2008; Ronnlund et al., 2005) show that practice effects persist even after 3 or 5 years in normal controls. However, with longer intervals, it becomes harder to determine the reliability of change scores because variations in performance or external factors affect these scores in all populations. Possible conclusions point to focusing on the sensitivity of various testing intervals and determining which retest interval would prove to be the most effective in achieving reliable scores.

It is important to consider the test-retest reliability of various neuropsychological tests. Measuring the reliability of each test allows clinicians to provide repeat testing while minimizing factors that add to elevated scores on subsequent tests. The use of alternate forms, providing efficient test intervals, controlling for score elevations with statistical approaches, as well as using tests with high reliability allows clinicians to factor out these influences to explain for unexpected elevated scores at retest. Salinsky et al. (2000) discuss and reference multiple studies evaluating the test-retest reliability of various neuropsychological tests such as the WAIS, Halstead-Reitan battery, Wechsler Memory Scale (WMS), Trail Making Test (TMT), Boston Naming Test (BNT), Stroop Color-World, and many more. Reliability estimates (Pearson's correlations) ranged from .46 to .94 and most practice effects were attributed to shorter test intervals and higher initial test scores (Salinksy et al., 2000).

It would be useful to space out testing periods so that individuals have a novel approach to the test at each administration. It would also be critical for understanding adaptive learning to evaluate the number of exposures individuals had to a specific test prior to administration. Watson et al. (1994) found that practice effects did not reach a plateau until 4-5 administrations mainly in tests of memory and motor speed. However, findings from Benedict and Zgaljardic (1998) claim that practice effects exist over multiple exposures, and significant practice effects were seen after a minimum of four sessions for tests that examined memory recall ability with the WMS. One question that this issue raises is whether or not it is effective to continue annualized testing to achieve accurate scores after a fourth administration since performance may be significantly affected by practice effects and eventually plateau. What seems to be the most variable factor that affects performance on repeated cognitive testing is testing intervals between sessions and the number of administrations.

Respondent Characteristics

Characteristics of the examinee that also may alter performance on repeated testing include age, clinical population, baseline performance, and education level. Studies show that smaller practice effects are present with increasing age. One study using a battery of neuropsychological tests showed that older participants scored lower on follow-up testing compared to younger adults (Temkin et al., 1999). However, a more recent study showed that age was unrelated to practice effects on repeated testing. Age-related differences are present during initial testing but not when comparing within-individual retesting (Salthouse, 2011). Rapport, Brines, Axelrod, & Theisen (1997) showed that higher IQ is more likely to result in higher practice effects in normal controls on multiple testing sessions. Similarly, Stuss, Stethem, & Poirier (1987) showed that greater education is also positively correlated with practice effects.

Another consideration is that practice effects in one clinical population may not necessarily be the same as in other clinical populations. The degree of change between two different groups can result in ineffective comparisons if one group is already known to show larger practice-related gains (Shatz, 1981). Alternately, some clinical populations with milder cognitive deficits, such as those with MCI (Rotrou et al., 2005), have shown similar practice effects to normal populations (Heaton et al., 2001). Therefore, performance results on retesting should be interpreted with caution given the variability that can be seen.

Standardization

Benedict and Zgaljardic (1998) found that practice effects were present when there is procedural learning, familiarity, or reduced anxiety during testing. Administering tests to patients repeatedly would help increase their familiarity with test items and possibly reduce anxiety. Other factors that must be considered when testing patients is the physical environment. Factors such as the setting of the room, examiner rapport, and patient motivation at each testing period can affect performance (McCaffrey et al., 1992). This consistent environment, along with familiar test content at each testing period, may contribute to score elevations. When proposing a study to look at practice effects, it is imperative to consider all factors that may alter test performance. In research trials, these factors must be controlled as best as possible so that the change seen in tests are due to the patient's performance rather than external factors. Therefore, it is critical to provide standardization among tests in order to reduce the effect of confounding factors.

Even small changes in procedural administration can make a difference in test performance. For example, great variability was shown on a three-word recall test among healthy older subjects depending on the words used (Cullum et al., 1993), while Lacritz et al. (1998) showed that there good recall was seen when explicit prompts to remember the words were provided. This shows that small changes in procedural administration of testing material can change the outcome of test scores, especially if varied instructions are given at each testing session.

Practice Effects in Clinical and Non-clinical Populations

In regards to practice effects, a patient may simply develop better test-taking strategies after each administration. This of course would be under the assumption that the ability to learn a task after repeated exposure was intact for cognitively declining individuals. Test - specific practice results in patients remembering items presented for longer than the testing period. On the second exposure to the test, patients may actually remember items from the first administration, recalling them from long-term memory.

McCaffrey (2000) and Beglinger (2005b) found that normal adults are greatly susceptible to practice effects. Due to their higher cognitive functioning, specifically in terms of memory, normal adults are much more capable to detect similarities between tests and are able to adapt to changes at each testing period than patients with cognitive deficits. Compared to various disease populations such as brain trauma, dementia, HIV, and schizophrenia, studies such as Heaton et al. (1991), Temkin et al. (1998), and Temkin et al. (1999) show that normal controls show significantly greater practice effects than clinical populations. Bartels et al. (2010) studied practice effects in healthy individuals over a one-year period in six cognitive domains and concluded that greater practice effects occurred early-on and stabilized after three months. The study also showed that healthy individuals have greater practice effects in cognitive domains of executive functioning and learning/memory. Therefore, looking at practice effects in repeated testing of disease populations with minimal cognitive deficits is crucial in order to accurately evaluate their diagnosis and prognosis.

Domains Most Susceptible to Practice Effects

When examining practice effects in normal adults and various disease populations, researchers have observed that certain domains are more susceptible to practice effects than others (Collie et al., 2002). Different areas of the brain are utilized when performing different cognitive tasks. Over time, with repeated exposures, learned material presented in one testing session is carried over to another (Basso, Bronstein, & Lang, 1999). Once an individual learns the rules of a test, they are not easily forgotten. When an individual has developed procedural learning or developed effective strategies, these will positively carryover to the next testing session and may positively impact scores at retest. Studies such as ones conducted by Benedict & Zgaljardic (1998) and McCaffrey et al (2000) have sought to determine the extent to which practice effects appear differently on tests of various cognitive domains, specifically with learning, memory, and executive functioning. By using alternate forms for healthy individuals, the studies shows that exposure to the same content across learning and memory domains increase practice effect gains while using varied administration with alternate forms showed smaller gains. However, some studies also hypothesized that practice effects are equal among cognitive domains (Mitrushina & Satz, 1991). This study concluded that among 122 normal elderly adults, practice effects were minimal with the WAIS-R and Wechsler Memory Scale over three annual testing. The most common domains that comprise cognitive assessments include learning, memory, executive functioning, language, processing speed, and attention. In a cross-sectional study, differences were mainly seen in episodic memory and processing abilities (reduced storage capacity, impaired processing efficiency, and diminished ability to coordinate simultaneous activities), working memory and speed rather than crystallized verbal abilities seen in language tests (Salthouse, 1990).

Executive Functioning

Tests that involve learning and applying a rule are more likely to produce higher practice effects than other tests. This would be the case especially if the same test is administered during both sessions. Studies show that tests requiring complex cognitive processing, or formulation of strategies (e.g. Wisconsin Card Sorting Test (WCST), Stroop Test), aids greater performance effects than tests that measure simple cognitive tasks (Basso et al., 1999). Presenting these higher-level cognitive tests repeatedly results in higher practice effects seen in subsequent testing of neuropsychological assessments. Tests that require the learning of rules and procedures are easier to retain over multiple sessions than tests that require recall of a list of items. In the Basso study, fifty male participants completed executive functioning and fluency tests. Reassessment after 12 months showed that tests that involved executive functioning were more susceptible to practice effects than tests that required initiation and generation such as Ruff Figural Fluency Test and F-A-S Verbal Fluency. Similarly, a study conducted by Bartels et al. (2010) examined the practice effects among various cognitive domains in healthy individuals. They showed that individuals continued to show practice effects over four sessions (over the course of a year) in executive functioning tests such as the WCST, TAP Flexibility, and Letter Number Sequencing.

This greatly differed from the cognitive domains such as language and visuospatial functions that showed minimal practice effects over one year.

Language

Language tasks are the least susceptible to practice effects because they relate to crystallized ability. Language development occurs throughout an individual's life and tends to be stable in adulthood. This is not the case in longitudinal studies done over high points of language acquisition, such as adolescence. Patients presenting with symptoms of dementia are reported to decline much faster in areas of language as well as other cognitive domains than individuals with mild cognitive impairment (Rozzini et al., 2007). A study done by Duff et al. (2008) showed that individuals with either amnestic or non-amnestic mild cognitive impairment are capable of retaining language ability over multiple assessments. This shows that MCI patients are similar to healthy individuals regarding domains of language. Therefore, practice effects should be minimal in this area for the MCI population. Language assessments serve as a well-needed control in examining practice effects in various populations since the practice effects should be minimal at retesting.

Learning and Memory

Domains of learning and memory are susceptible to practice effects because patients are more likely to remember certain aspects of testing material; such as names, places, and events (as seen in WMS Logical Memory). Kaszniak, Wilson, Fox, and Stebbins (1986) reported that elderly patients gained almost 2 points on both immediate and delayed recall of Logical Memory after a one-year retesting period. A more recent study done by Bartels et al, (2010) showed that individuals gained a significant 15% from baseline in tests of learning and memory after a one-year interval. Both normal adults and certain disease populations with minimal memory impairment have been shown to have high practice effects because the content is repeated at each administration (Benedict & Zgaljardic, 1998). McCaffrey, Ortega, Orsillo, and Nelles (1992) showed that there was a practice effect of .4-.9 standard deviations on Logical Memory over a 7-10 day interval. Also seen in the study was that by session 4, the practice effects were significant among all normal adult participants. Although Logical Memory seems to have the highest practice effects among the Wechsler Memory tests, others such as Verbal Paired Associates and Visual Reproduction are impacted as well. Theisen, Rapport, Axelrod, & Brines (1998) looked at practice effects of the Wechsler Memory Scale among a group of normal adults over four sessions with 2-week intervals. Their results showed that significant practice effects were present in immediate and delayed recall of Logical Memory and delayed Verbal Paired Associates, relatively smaller gains were seen with immediate Verbal Paired Associates and both immediate and delayed Visual Reproduction.

Other factors that determine the likelihood that patients will achieve higher scores on retesting is the complexity of the item materials. Donovan and Rodosevich (1999) studied the difference between acquisitions of the item content to the complexity of the task presented. The study defined a test-retest period to be at 24 hours. Measuring the retention performance after a 24-hour period revealed that the greatest effect sizes were present in tasks that involved low mental ability and average overall complexity. Verbal memory and free recall related tasks would be directly classified in this category.

Attention and Processing Speed

Neuropsychological tests involving attention and processing speed are more likely to stabilize with repeated testing. Studies such as Basso, Bornstein, & Lang (2010), Duff et al. (2008), and Heaton et al. (2001) show that initially, there may be an initial gain in performance once the individual has familiarized themselves with testing rules and operation. However, abilities of this domain tend to stabilize with repeated testing over longer test intervals. Similarly, in just over six weeks, normal controls in Berglinger (2005b) showed significant improvements in all tests except for letter-number sequencing. Similar results were seen in Bartels et al. (2010) where individuals had elevated scores after a three month interval at the second test and plateaued at sessions three and four at six and twelve months. Similarly, four repeated testing sessions occurring a week apart showed significant practice effects for both parts of TMT. This effect diminished after testing again in three months (McCaffrey ,Ortega, and Haase, 1993).

Visuospatial Functioning

Although studies show that tests of learning and memory have greater practice effects than other cognitive domains, visuospatial functioning assessments show more mixed results. Most studies that look at visuospatial functioning, like the Bartels et al. (2010) using RBANS Lines and Figure Copy, show minimal practice effects over multiple testing sessions. However, studies such as Duff et al. (2008) looking at healthy individuals showed dramatic improvement after a one-week interval in this domain. Visuospatial functions show age-related influences, as seen in Mitrushina and Satz (1991). Those who were younger tested higher than those who were older. Additionally, at three retests, the younger group retained the same level of performance while the older group showed some decline.

Methods that Correct for Practice Effects

Repeated testing may be unavoidable in some cases because testing is required to monitor disease progression. Therefore, testing procedures can be changed in order to alleviate some of the uncertainty that practice effects have skewed scores. Most studies seek to moderate practice effects by studying various time frames in between testing exposures. This can ensure that there has been enough distraction between one exposure and the next; that patients have not adapted to the questions shown and learned from past failures and missed responses. Related to practice effects are novelty effects and how that relates to patient performance on neuropsychological testing. In the beginning stages of administration, test items are novel to the patient, but during repeated testing there is a higher chance that the novelty is minimal, if not absent. The novelty of the test (as secured by alternate forms at retest) can determine how the individual presents initially without bias and how their clinical status is a clear representation of their performance. Methods such as determining appropriate control groups, administration of alternate forms, developing confidence intervals, and using appropriate interval periods reduce the potential impact of practice effects.

Control Groups

In research, control groups are utilized in order to compare group mean changes of a disease population to the normal comparative population. General comparisons can be made that accurately show whether or not there is an obvious change present in a study sample compared to a control sample. When comparing an entire sample, group means represent the scatter of scores across variable performance levels from a subset of people being evaluated. Therefore, many studies suggest the use of a control group as a way to control for practice effects (Slade, Sanchez, Townes, & Aldea, 2001). In order to monitor an individual's change over time, comparison to the normal group may not always show an accurate change. This method assumes that practice effects of a comparison group are similar to practice effects in a clinical group (Wilson et al., 2000). Without control groups, external factors, demographics, and testing biases like practice effects may skew results and are unnoticed. Furthermore, this helps the clinician who needs to determine if an individual case is significantly different from the normal population or has changes that are due to individual performance fluctuations. In order to develop a reliable change in a disease population, reliable change scores should also be developed for normal control groups to understand the variables in repeat testing in a non-clinical population that is not expected to change in clinical status.

Alternate Forms

For some tests that use the same item content at each administration, it is particularly challenging to avoid errors in which a patient is able to recall the items by memory prior to administration. Therefore, alternate forms are beneficial in correcting this potential error. Longitudinal studies have shown that the use of alternative forms do not eliminate the presence of practice effects in elderly patients (Dikmen et al., 1999). This would suggest that there are other factors that contribute to practice effects, such as improved test taking skills, learning test rules and procedures, and reduced anxiety. Furthermore, the use of alternative forms adds an additional source of variability that minimizes the occurrence of elevated scores.

There is also a significant difference in the intervals in which repeated testing is administered. Practice effects are most likely to be present after two sessions without the use of alternate forms (Collie, Maruff, Darby, and McStephen, 2003). Similarly, Falleti et al. (2007)

administered a neuropsychological battery to 55 adults (age range: 18-40) twice within a 10-minute test-retest interval and a third time at one month. The brief period between the first two tests "corrected" for immediate practice effects and served as a dual baseline while the third administration looked for learned performance after a longer test interval. The results of this study showed that there were no significant practice effects observed at the third administration. The greatest amount of learning, even with alternate forms occurred in the first three sessions. This concluded that performance was stabilized between the first and second administration regardless of the use of alternate forms. The studies suggest that positive performance occurs after a patient has been given the test two or three times.

Statistical Approaches

There have been many inferences on how to best adjust for changes seen on scores from repeated testing. There have also been suggestions on how to control the presence and magnitude of practice effects by using statistical methods that control for the change in test performance. One such method would be to define appropriate confidence intervals when reporting scores from an assessment. This method is utilized in order to provide clinicians an accurate range within which the change scores should fall. Subsequently, scores that fall outside of this range may bring into question if the performance was altered by other factors. The subtraction of pre- from post-test performance is a formulated method (taking standard error into consideration) in which we are able to observe a change with better reliability (Benedict & Zgaljardic, 1998).

Methods Used to Assess Reliable Change

In most cases with disease populations, the risk of making a Type 1 error (concluding a change when there is none) may be greater than that of making a Type II error (failing to identify true change). Although it is common to see change in performance between testing periods, it is critical to recognize what amount of the performance is due to real change. Methods have been put into place to compare tests between various testing periods and set a normative standard of change that is typically seen. These standards are derived from factoring in practice effects and the standard error of difference between test scores (Heaton et al., 2001).

Reliable Change Index

Statistical methods have been developed for clinicians to help determine if a clinically meaningful change has occurred. Researchers developed a method to derive cutoff scores that reflect changes for individual tests that are rare in the general population (Jacobson & Traux, 1991). The reliable change index (RCI), a more conservative approach, is similar to standard deviation. It involves inputting a time 1 and time 2 score above the standard error of difference. The reliable change formula can be used to compare changes that would occur between a disease population and normal controls to determine if the change is above what would be expected in a given population for that particular test. The advantages of using RCI are that it is a more accurate estimate of relative change and it controls for test reliability. The RCI indicates if a possible practice effect influenced scores by how much it is different from normal change scores. The next step that would help clinicians is to determine what items are most likely to produce unlikely positive scores in a declining patient population. Chelune et al., (1993) modifies this method so that the mean practice effect is subtracted from the difference of scores of time 1 and time 2. By correcting this phenomenon, clinicians can use predetermined practice effect means for tests that are most likely to result in higher scores on retesting.

Affects of Practice Effects on MCI

Retesting could provide clinically useful information about the progression of an illness (Duff et al., 2007). Understanding the general practice effects present for specific tests will allow clinicians to rule out extraneous results that do not match with clinical presentation. Progression of disease is influenced by and measure by many factors of genetics, neuroimaging, biomarkers, and clinical characteristics. In a study done by Rozzini and colleagues (2007), 119 patients who met criteria for early stages of dementia were examined and observed for one year. The study shows that patients were significantly worsen in areas of language, executive functioning, and memory after one year when compared to those who were normal. On the other hand, a study done by Maioli and colleagues (2007) determined that not all patients who present with mild cognitive impairment will worsen to dementia. Out of a sample of 100, 23.8% can be seen to worsen over six years, 53.8% remained the same, and 17.3% reverted back to normal. The progression of MCI varies greatly from person to person and can be confounded by factors that interfere with serial testing.

Rozzini et al. (2007) also showed that those who were initially diagnosed with amnestic MCI (presenting majority of deficits in memory) and converted to dementia or AD at one-year follow up showed worsening scores in domains of executive memory. In Mielke et al. (2013), the incidental MCI group unexpectedly showed practice effects in memory at the second visit and significantly declined thereafter. Moreover, Matthews et al. (2014) used the Uniform Data Set (UDS) neuropsychological battery to show that test-exposed participants showed greater mean scores at retest compared to test-naïve participants in MCI and normal controls. These and other studies as reviewed in the literature indicate that MCI patients can vary greatly (improving, declining, or stabilizing) after being exposed to testing material. These variations in test scores for serial assessments complicate the process of accurately evaluating a patient's clinical status. Future studies must focus on determining how the change rate really presents in these clinical populations and what clinicians can do in order to detect significant change.

CHAPTER THREE

Discussion

The importance of understanding practice effects has clinical value in the diagnosis of MCI and other neurocognitive disorders. In addition, learning more about how varying cognitive abilities are affected in repeated neuropsychological batteries aid in developing effective diagnostic procedures for disorders with the potential to worsen over time. Neuropathological studies already show that there is an intermediate stage between normal aging and early symptoms of dementia and Alzheimer's disease (Petersen & Negash, 2008). This middle ground that is often a gray area in clinical diagnosis can be categorized as a diagnosis of mild cognitive impairment. Therefore, early detection and intervention at these stages are useful in determine treatment and clinical characteristics of how the disease will progress.

Early detection and monitoring of a cognitive disorder starts with providing an accurate diagnosis at each visit. In order to provide effective care and assessment, annual visits are best recommended for these patients with known or suspected progressive conditions. Additionally, the literature shows that practice effects are greater in initial exposures that have intervals of three months or shorter with most neuropsychological tests and are largely absent in longer intervals (Galvin et al., 2005; Schrijnemaekers et al., 2006). Therefore, the best practice would be to provide annual testing to minimize practice effects on a majority of tests. As seen from studies by Duff et al. (1998), Heaton et al. (2001), Berglinger et al. (2005), Salinsky et al. (2000) and others, significant practice effects across major cognitive domains (learning, memory, executive functioning) are less likely at annual follow-ups. Since minimal practice effects are expected at first retest, as seen in multiple studies with smaller test intervals (Bartels et al., 2010;

Theisen, Rapport, Axelrod, and Brines, 1998; Duff et al., 2009), a retest could be administered within a short period of the original. The second test would act as the baseline score for subsequent annual assessments, therefore, minimizing initial practice effects.

Additionally, reliable change indices, regression to the mean, and other statistical approaches can be used to factor in practice effects that are seen at repeat testing. This is especially the case when comparing a clinical population to healthy individuals during serial assessments. Utilizing normal controls tested at similar intervals is an effective method in helping to determine true change in clinical populations. As reviewed in the literature, normal controls do show greater practice effects at each testing interval compared to clinical populations.

Most MCI patients show deficits in learning and memory earlier than in other cognitive domains. The literature suggests that these domains, along with executive functioning show greater practice effects over multiple exposures without plateauing. This is unlike cognitive domains of language, processing speed, attention, and visuospatial functioning which tend to show smaller changes over multiple exposures (Bartels et al., 2010). Domains of memory and learning could be easier for patients to learn test rules and become familiar with item content (Beglinger et al., 2005). However, other domains, like language, are more crystallized and are more influenced by factors of IQ or education level. Moreover, literature findings show that studying certain cognitive domains that are more susceptible to elevated performance will help detect change in patients.

These practices can be utilized at centers across the country to better evaluate patients with cognitive deficits who vary in their performance over the years. As one example, Alzheimer's Disease Centers (ADC) administer a neuropsychological battery to patients annually. They evaluate both healthy and diseased populations, monitoring progress and assessing their cognitive abilities to help individuals affected by dementia. With MCI individuals, who show milder yet varying degrees of impairment, monitoring disease progression is essential. Knowing that one-year test intervals are best to minimize practice effects in longitudinal studies of progression, the best practice for these centers is to continue providing annual assessments. When possible, centers should also consider utilizing alternate forms for tests every other year to retain the novelty of testing material, especially for tests of learning/memory and executive functioning. However, the drawback of using alternate forms becomes a consideration, as they are not perfectly reliable for assessing the same level of cognitive ability.

It would be practical to evaluate patients who are most susceptible to rapid cognitive decline. Patients with MCI, older than 65 years, have a risk of 10-15% per year to develop dementia in comparison to the healthy population who have a risk of 2% per year (Defrancesco et al., 2010). Although MCI patients may originally show deficits in memory and not in other cognitive domains, the influence of the impairment may quickly spread to other areas over time. Therefore, it is critical to understand how the change in performance over multiple domains presents itself in a clinical setting. This could be misrepresented if a patient shows accelerated performance due to practice effects. By understanding the scores variances, clinicians are able to provide a more accurate diagnosis at each visit and inform family members and caretakers of the cognitive changes seen in patients.

CHAPTER FOUR

Suggestions for Future Research

The current review of the literature poses some directions for future research on the topic of practice effects and MCI. Although methods are put into place to detect practice effects in testing sessions, it would be beneficial to understand which test items contribute to elevated scores and what factors can be altered to minimize its occurrence.

Although various studies (Collie et al., 2002; Duff et al., 2008; Rozzini et al., 2007; Rotrou et al., 2005) focus on practice effects in amnestic and non-amnestic MCI, other subtypes of MCI should be considered in future studies on practice effects (i.e. amnestic, non-amnestic, single-domain, and multiple-domain). MCI is classified as an intermediate stage between the decline of normal aging and dementia with minimal cognitive deficits. Therefore, it is possible that individuals with different MCI diagnoses display different trajectories in their progression. A future study should consider examining practice effects in all subtypes of MCI using the same neuropsychological battery over the same testing intervals.

Expanding on Bartels et al. (2010), future studies should encompass a greater variety of cognitive tests studied in the same sample of patients. Most studies include only a handful of tests in each cognitive domain. However, many more neuropsychological assessments should be considered in one study to provide an accurate idea of how practice effects are present globally across all cognitive domains and tests used. Moreover, using alternate forms of tests every other year would also be beneficial to consider for a future study.

Past studies have examined various test intervals from 24-hours to 3 years (Galvin et al., 2005; Schrijnemaekers et al., 2007; Collie et al., 2003). Limited data are available for longitudinal studies past three test exposures with the MCI population. This may partially be due to individuals with MCI cognitively declining and converting to an AD diagnosis. However, it would still be useful to study the longitudinal path of test results and observe testing performance with multiple previous exposures to the material.

Lastly, developing reliable change indices for each cognitive test would also be important in providing a comparison for clinicians to detect true change. Future research could focus on combining the scores of various neuropsychological assessments by domain and developing global indices to represent normal change within certain populations. Additional research would be required to assess if this could be useful for patients with MCI long-term and how this affects identification of disease progression.

References

Albert, M. S., Jones, K., Savage, C. R., Berkman, L., Seeman, T., Blazer, D., & Rowe, J. W. (1995). Predictors of cognitive change in older persons: MacArthur studies of successful aging. *Psychology and Aging*, *10*(4), 578.

Bartels, C., Wegrzyn, M., Wiedl, A., Ackermann, V., & Ehrenreich, H. (2010). Practice effects in healthy adults: a longitudinal study on frequent repetitive cognitive testing. *BMC Neuroscience*, *11*(1), 118.

Basso, M. R., Bornstein, R. A., & Lang, J. M. (1999). Practice effects on commonly used measures of executive function across twelve months. *The Clinical Neuropsychologist*, *13*(3), 283-292.

Beglinger, L. J., Gaydos, B., Tangphao-Daniels, O., Duff, K., Kareken, D. A., Crawford, J., ... & Siemers, E. R. (2005). Practice effects and the use of alternate forms in serial neuropsychological testing. *Archives of Clinical Neuropsychology*, *20*(4), 517-529.

Benedict, R. H., & Zgaljardic, D. J. (1998). Practice effects during repeated administrations of memory tests with and without alternate forms. *Journal of Clinical and Experimental Neuropsychology*, *20*(3), 339-352.

Chelune, G. J., Naugle, R. I., Lüders, H., Sedlak, J., & Awad, I. A. (1993). Individual change after epilepsy surgery: practice effects and base-rate information. *Neuropsychology*, *7*(1), 41.

Collie, A., Maruff, P., Darby, D. G., & McStephan, M. I. C. H. A. E. L. (2003). The effects of practice on the cognitive test performance of neurologically normal individuals assessed at brief test–retest intervals. *Journal of the International Neuropsychological Society*, *9*(03), 419-428.

Cooper, D. B., Epker, M., Lacritz, L., Weiner, M., Rosenberg, R. N., Honig, L., & Cullum, C. M. (2001). Effects of Practice on Category Fluency in Alzheimer's Disease*. *The Clinical Neuropsychologist*, *15*(1), 125-128.

Cullum, C. M., Thompson, L. L., & Smernoff, E. N. (1993). Three-word recall as a measure of memory. *Journal of Clinical and Experimental Neuropsychology*, *15*(2), 321-329.

Darby, D., Maruff, P., Collie, A., & McStephen, M. (2002). Mild cognitive impairment can be detected by multiple assessments in a single day. *Neurology*, *59*(7), 1042-1046.

Dikmen, S. S., Heaton, R. K., Grant, I., & Temkin, N. R. (1999). Test–retest reliability and practice effects of expanded Halstead–Reitan Neuropsychological Test Battery. *Journal of the International Neuropsychological Society*, *5*(04), 346-356.

Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology*, *84*(5), 795.

Duff, K., Lyketsos, C. G., Beglinger, L. J., Chelune, G., Moser, D. J., Arndt, S., ... & McCaffrey, R. J. (2011). Practice effects predict cognitive outcome in amnestic mild cognitive impairment. *The American Journal of Geriatric Psychiatry*, *19*(11), 932-939

Duff, K., Beglinger, L. J., Schultz, S. K., Moser, D. J., McCaffrey, R. J., Haase, R. F., ... & Paulsen, J. S. (2007). Practice effects in the prediction of long-term cognitive outcome in three patient samples: A novel prognostic index. *Archives of Clinical Neuropsychology*, *22*(1), 15-24.

Duff, K., Beglinger, L. J., Van Der Heiden, S., Moser, D. J., Arndt, S., Schultz, S. K., & Paulsen, J. S. (2008). Short-term practice effects in amnestic mild cognitive impairment: Implications for diagnosis and treatment. *International Psychogeriatric*, *20*(05), 986-999.

Falleti, M. G., Maruff, P., Collie, A., & Darby, D. G. (2006). Practice effects associated with the repeated assessment of cognitive function using the CogState battery at 10-minute, one week and one month test-retest intervals. *Journal of Clinical and Experimental Neuropsychology*, *28*(7), 1095-1112.

Galvin, J. E., Powlishta, K. K., Wilkins, K., McKeel, D. W., Xiong, C., Grant, E., ... & Morris, J. C. (2005). Predictors of preclinical Alzheimer disease and dementia: a clinicopathologic study. *Archives of Neurology*, *62*(5), 758-765.

Heaton, R. K., Temkin, N., Dikmen, S., Avitable, N., Taylor, M. J., Marcotte, T. D., & Grant, I. (2001). Detecting change:: A comparison of three neuropsychological methods, using normal and clinical samples. *Archives of Clinical Neuropsychology*, *16*(1), 75-91.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*(1), 12.

Kaszniak, A. W., Wilson, R. S., Fox, J. H., & Stebbins, G. T. (1986). Cognitive assessment in Alzheimer's disease: cross-sectional and longitudinal perspectives. *The Canadian Journal of Neurological Sciences, 13*(4 Supplement), 420-423.

Kaufman, A. S., & Kaufman, N. L. (1990). *Kaufman Brief Intelligence Test*. New York, NY: John Wiley & Sons, Inc.

Lacritz, L. H., & Cullum, C. M. (1998). The hopkins verbal learning test and CVLT: a preliminary comparison. *Archives of Clinical Neuropsychology*,*13*(7), 623-628.

Lezak, M. D. (Ed.). (2004). *Neuropsychological Assessment*. London, New England: Oxford University Press.

National Institute of Aging (NIH). (2014) *Alzheimer's disease.* Retrieved from http://www.nia.nih.gov/alzheimers/publication/alzheimers-disease-fact-sheet

Maioli, F., Coveri, M., Pagni, P., Chiandetti, C., Marchetti, C., Ciarrocchi, R., ... & Pedone, V. (2007). Conversion of mild cognitive impairment to dementia in elderly subjects: a preliminary study in a memory and cognitive disorder unit. *Archives of Gerontology and Geriatrics*, *44*, 233-241.

Mathews, M., Abner, E., Kryscio, R., Jicha, G., Cooper, G., Smith, C., & Schmitt, F. A. (2014). Diagnostic accuracy and practice effects in the National Alzheimer's Coordinating Center Uniform Data Set neuropsychological battery. *Alzheimer's & Dementia*, 13, 433-439.

McCaffrey, R. J., Ortega, A., & Haase, R. F. (1993). Effects of repeated neuropsychological assessments. *Archives of Clinical Neuropsychology*, *8*(6), 519-524.

McCaffrey, R. J., Ortega, A., Orsillo, S. M., Nelles, W. B., & Haase, R. F. (1992). Practice effects in repeated neuropsychological assessments. *The Clinical Neuropsychologist*, *6*(1), 32-42.

Mielke, M. M., Roberts, R. O., Savica, R., Cha, R., Drubach, D. I., Christianson, T., ... & Petersen, R. C. (2013). Assessing the temporal relationship between cognition and gait: slow gait predicts cognitive decline in the Mayo Clinic Study of Aging. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *68*(8), 929-937.

Mitrushina, M., & Satz, P. (1991). Effect of repeated administration of a neuropsychological battery in the elderly. *Journal of Clinical Psychology*, 47(6), 790-801.

Petersen, R. C., & Negash, S. (2008). Mild cognitive impairment: an overview. *CNS Spectrums*, *13*(1), 45.

Rapport, L. J., Axelrod, B. N., Theisen, M. E., Brines, D. B., Kalechstein, A. D., & Ricker, J. H. (1997). Relationship of IQ to verbal learning and memory: test and retest. *Journal of Clinical and Experimental Neuropsychology*, *19*(5), 655-666.

Rapport, L. J., Brines, D. B., Theisen, M. E., & Axelrod, B. N. (1997). Full scale IQ as mediator of practice effects: The rich get richer. *The Clinical Neuropsychologist*, *11*(4), 375-380.

Rönnlund, M., Nyberg, L., Bäckman, L., & Nilsson, L. G. (2005). Stability, growth, and decline in adult life span development of declarative memory: cross-sectional and longitudinal data from a population-based study. *Psychology and aging*, *20*(1), 3.

Rotrou, J., Wenisch, E., Chausson, C., Dray, F., Faucounau, V., & Rigaud, A. S. (2005). Accidental MCI in healthy subjects: a prospective longitudinal study. *European Journal of Neurology*, *12*(11), 879-885.

Rozzini, L., Vicini Chilovi, B., Conti, M., Delrio, I., Borroni, B., Trabucchi, M., & Padovani, A. (2007). Neuropsychiatric symptoms in amnestic and nonamnestic mild cognitive impairment. *Dementia and Geriatric Cognitive Disorders*, *25*(1), 32-36.

Salinsky, M. C., Storzbach, D., Dodrill, C. B., & Binder, L. M. (2001). Test–retest bias, reliability, and regression equations for neuropsychological measures repeated over a 12–16-week period. *Journal of the International Neuropsychological Society*, *7*(5), 597-605.

Salthouse, T. A. (2011). Effects of age on time-dependent cognitive change. *Psychological Science*, 22(5), 682-688.

Salthouse, T. A. (1990). Working memory as a processing resource in cognitive aging. *Developmental Review*, *10*(1), 101-124.

Schrijnemaekers, A. M. C., de Jager, C. A., Hogervorst, E., & Budge, M. M. (2006). Cases with mild cognitive impairment and Alzheimer's disease fail to benefit from repeated exposure to episodic memory tests as compared with controls. *Journal of Clinical and Experimental Neuropsychology*, *28*(3), 438-455.

Shatz, M. W. (1981). WAIS practice effects in clinical neuropsychology. *Journal of Clinical and Experimental Neuropsychology*, *3*(2), 171-179.

Slade, P., Sanchez, P., Townes, B., & Aldea, G. S. (2001). The use of neurocognitive tests in evaluating the outcome of cardiac surgery: Some methodological considerations. *Journal of Cardiothoracic and Vascular Anesthesia*, *15*(1), 4-8.

Small, B. J., Fratiglioni, L., Viitanen, M., Winblad, B., & Bäckman, L. (2000). The course of cognitive impairment in preclinical Alzheimer disease: three-and 6-year follow-up of a population-based sample. *Archives of Neurology*,*57*(6), 839-844.

Stuss, D. T., Stethem, L. L., & Poirier, C. A. (1987). Comparison of three tests of attention and rapid information processing across six age groups. *The Clinical Neuropsychologist*, *1*(2), 139-152.

Temkin, N. R., Heaton, R. K., Grant, I., & Dikmen, S. S. (1999). Detecting significant change in neuropsychological test performance: a comparison of four models. *Journal of the International Neuropsychological Society*, *5*(04), 357-369.

Theisen, M. E., Rapport, L. J., Axelrod, B. N., & Brines, D. B. (1998). Effects of practice in repeated administrations of the Wechsler memory scale-revised in normal adults. *Assessment*, *5*(1), 85-92.

Van Der Elst, W., Van Boxtel, M. P., Van Breukelen, G. J., & Jolles, J. (2008). Detecting the significance of changes in performance on the Stroop Color-Word Test, Rey's Verbal Learning Test, and the Letter Digit Substitution Test: the regression-based change approach. *Journal of the International Neuropsychological Society*, *14*(01), 71-80.

Watson, F. L., Pasteur, M. A. L., Healy, D. T., & Hughes, E. A. (1994). Nine parallel versions of four memory tests: An assessment of form equivalence and the effects of practice on performance. *Human Psychopharmacology: Clinical and Experimental*, *9*(1), 51-61.

Wilson, B. A., Watson, P. C., Baddeley, A. D., Emslie, H., & Evans, J. J. (2000). Improvement or simply practice? The effects of twenty repeated assessments on people with and without brain injury. *Journal of the International Neuropsychological Society*, *6*(04), 469-479.

BIOGRAPHICAL SKETCH

Shruthi Rao
rao.shruthi.22@gmail.com

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE	YEAR(s)	FIELD OF STUDY
The University of Illinois at Urbana-Champaign Champaign, IL	B.S.	2007-2011	Psychology. Molecular and Cellular Biology
The University of Texas Southwestern School of Health Professions Dallas, TX	M.R.C.	2011-2014	Rehabilitation Counseling

<u>Positions and Employment</u>

2012-2014 Academic Coach- Matthew Housson & Associates
2010-2011 Career Services Paraprofessional- University of Illinois Career Center

<u>Clinical Experience</u>

2013 Intern, UT Southwestern Supported Employment
2012-2013 Intern, Pate Rehabilitation
2012-2013 Research Assistant, Texas Healthy Presbyterian Institute of Environment and Exercise Medicine
2012-2013 Intern, UT Southwestern Rehabilitation Services

<u>Presentations and Publications</u>

2011 Analysis of Dendritic Spines in a Mouse Model of Fragile X Mental Retardation, University of Illinois at Urbana-Champaign, Bachelor's Thesis
2010 Morphological analysis of dendritic spines in a mouse model of Fragile X Mental Retardation using live animal, 2-photon microscopy, University of Illinois at Urbana-Champaign [Society of Neuroscience Annual Conference Abstract Submissions]

<u>Professional Memberships</u>

2013-Present American Psychological Association (Student Affiliate)
2012-Present National Rehabilitation Association
2012-2014 North Texas Rehabilitation Association