12-2015

Memory Deficits in Older Adults: Evaluating Spaced Retrieval with Multiple Probe Techniques

Christopher Walmsley
Western Michigan University, qualmsley@gmail.com

MEMORY DEFICITS IN OLDER ADULTS: EVALUATING SPACED RETRIEVAL WITH MULTIPLE PROBE TECHNIQUES

by

Christopher Walmsley

A dissertation submitted to the Graduate College
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
Psychology
Western Michigan University
December 2015

Doctoral Committee:

R. Wayne Fuqua, Ph.D., Chair
Alan Poling, Ph.D.
Scott Gaynor, Ph.D.
Jonathan Baker, Ph.D.

MEMORY DEFICITS IN OLDER ADULTS: EVALUATING SPACED RETRIEVAL WITH MULTIPLE PROBE TECHNIQUES

Christopher Walmsley, Ph.D.

Western Michigan University, 2015

Difficulties in recognizing and remembering the names of individuals are a common behavioral symptom of major neurocognitive disorder. A number of behavioral strategies have been proposed to improve memory deficits, including spaced retrieval, an intervention that emphasizes delayed recall of target information. Unfortunately, many of the studies that report beneficial effects of spaced retrieval use a very limited range of outcome measures, thus calling into question the magnitude and generality of any reported memory improvement. This study reports on the impact of spaced retrieval using four older adults with cognitive impairment living in an assisted living facility. All participants demonstrated difficulty naming and recalling names of staff members at the facility who provided care. A multiple baseline across participants design with embedded probes was used to evaluate the effects of spaced retrieval on the acquisition, delayed recall, and generalization of naming a target staff member. Probe measures included a non-identity matching-to-sample task, naming in the presence of untrained photographs of the target staff member, and naming during brief video presentations of the target staff member. Non-example presentations were also interspersed. Spaced retrieval resulted in within-session increases in delayed recall for all participants, and also resulted in minimal evidence of generalization across probe measures, including live-

person probes. Modified spaced retrieval interventions were then employed to address stimulus control deficits. These modified conditions resulted in further gains in delayed recall performance, as well as improved generalization across probes. These results are discussed in the context of previous research with spaced retrieval. Future directions are also discussed.

ACKNOWLEDGMENTS

I thank Scott Gaynor for his insight into conceptual matters related to my dissertation, and to Jonathan Baker for his helpful knowledge pertaining to the unique challenges older adults face suffering from neurocognitive disorder. I also thank Alan Poling for his continued willingness to support my research endeavors throughout the years. These individuals have provided me with a solid foundation in psychological theory, research, and practice.

I would also like to thank my parents, Rhonda and Stephen Walmsley, for their unconditional support and love over the course of my graduate school career. Neither of them attended college, let alone graduate school, but nevertheless they were always willing to listen to me speak about my experiences and offer words of encouragement. I am ever grateful to them.

Finally, I thank my graduate mentor, Wayne Fuqua, for sharing his perspectives in critical thinking, professional, and ethical issues, and in helping me to navigate academic waters. I am proud to have been a member of his lab, and am excited to move forward, knowing that I will carry his teachings with me into my next career endeavor.

Christopher Walmsley

TABLE OF CONTENTS

ACKNOWLEDGMENTS ii

LIST OF TABLES iv

LIST OF FIGURES v

INTRODUCTION 1

- Dementia Types, Symptoms, and Terminological Changes 2
- Behavioral Excesses and Deficits 4
- Prevalence Of And Projections For Dementia 4
- Costs of NCD 5
- Pharmacological Treatment Models 6
- Behavioral Gerontology 7
- Memory Interventions 9
- Spaced Retrieval 10
- Purpose 17

METHOD 18

- Participants 18
- Screeners 19
- Recruitment 22
- Settings and Materials 23
  - Target (S+) Stimulus Class 24

Table of Contents—continued

Video Probes 24

Textual Prompts 25

Response Definitions 25

Measures of Acquisition 25

Measures of Recall 26

Measures of Generalization 26

Visual Matching-to-sample 26

Control Measures 27

Inter-observer Agreement (IOA) on Dependent Measures 27

Independent Variables 28

Research Design 28

Research Procedures 29

General Procedures 29

Baseline Assessment 29

Independent Variables 31

Treatment Integrity 35

RESULTS 35

Barry 38

George 43

Susan 49

Harriet 54

DISCUSSION 59

Measures Of Recall 59

Comparisons to Previous SR Studies Targeting Name-Face Relations 60

Error Measures 62

Generalization Probes 63

Limitations 68

Future Directions 70

Conclusions 72

REFERENCES 74

APPENDICES

A. Example Set of Target Staff Member 83

B. Example Spaced Retrieval Array 86

C. Spaced Retrieval Data Sheet 88

D. Probe Data Sheet 90

E. Treatment Integrity Checklist 92

F. Human Subjects Review Board Approval 95

LIST OF TABLES

1. Participant Characteristics 19

2. Intervention Targets and Participants' Performance on Baseline Recall and Spaced Retrieval Control Probes 36

LIST OF FIGURES

1. The staggered introduction of conditions across concurrent baselines for Barry and Harriet 37

2. The staggered conditions for Susan and George's concurrent baselines 38

3. Barry: Recall and Recall Error Data 40

4. Barry: Generalization Probes Measures 42

5. George: Recall and Recall Error Data 44

6. George: Generalization Probes Measures 47

7. Susan: Recall and Recall Error Data 49

8. Susan: Generalization Probes Measures 52

9. Harriet: Recall and Recall Error Data 55

10. Harriet: Generalization Probes Measures 57

INTRODUCTION

The number of older adults in the United States continues to grow larger every year. The U.S. Census Bureau reports that between 2000 and 2010, the US population segment aged 65 years and older have increased by 15.1%, notably larger than the 9.1% growth for the total U.S. population (Werner, 2011). Older adults make up 13% of the total population, translating to a figure of just over 40 million individuals (Werner, 2011).

The growth of the older adult population segment is projected to continue in the future, and at a substantially greater rate. Ortman, Velkoff, and Hogan (2014) report that the number of older adults will be 83.7 million by 2050, which is almost double that of the current estimated population of 43.1 million in 2012. Within the older adult demographics, the oldest old (i.e., 85 and above) will increase at the fastest rate, from just fewer than six million in 2012 to approximately 18 million in 2050 (Ortman et al., 2014). Thus, this projection estimates a tripling of this subgroup of older adults. These data suggest that the changing demographics in the United States, termed the "Graying of America", will persist with accelerated growth as more Americans turn 65 and older.

Much of the reason for this population change is due to the first of the baby boom generation turning 65 in 2011. The baby boomer generation refers to those individuals who were born between the years of 1946 and 1964. It is expected that the increased life expectancy will carry the baby boom generation well into the 2030's, as they begin to turn 85 in 2031.

Older adults are at an increased risk of having comorbid health conditions, and in general require careful medical, community, and mental health considerations. In order,

heart disease, cancer, chronic respiratory disease, stroke, and Alzheimer's disease are the five most common causes of death among older adults (Center for Disease Control, 2010). While the first four causes are either staying relatively stable or are decreasing across time, the number of deaths from Alzheimer's disease and other dementias is dramatically rising. Between 2009 and 2010, age-adjusted death due to Alzheimer's disease rose 3.7% (Murphy, Xu, & Kochanek, 2013).

Dementia Types, Symptoms, and Terminological Changes

Although a commonly held belief about dementia is that it is a memory-loss condition, in actuality it impacts a wide range of functioning, and is not identified by only one symptom. With the publication of the DSM-5 in 2013, some noteworthy changes to dementia diagnostic categories occurred. The term "dementia" has now been replaced with the diagnostic classification "neurocognitive disorders" (NCD). The condition that previously was called dementia is now "major neurocognitive disorder". The diagnosis "mild neurocognitive disorder" is a novel addition to the diagnostic nomenclature. In addition, etiology criteria have slightly changed such that the subtypes (e.g., vascular, frontotemporal) are no longer simply text descriptions, but carry with them their own specific criteria. Blazer (2013) argues that by including mild NCD in the DSM-5, older adults who may be exhibiting the beginning signs of dementia, but who would have fallen at sub diagnostic levels of severity under DSM-IV criteria, can still receive a diagnosis and gain access to treatment. Although there is no "cure" for NCD, early detection and intervention is critical to slow the rate of symptom progression and to maintain health and quality of life.

Broadly speaking, NCD is a clinical syndrome defined by the symptoms associated with neuropathological changes. Defining symptoms include multiple deficits in cognitive skills, a notable decline from previous skill levels, and a significant disruption of daily routines and common activities (McCurry & Drossel, 2011). These changes must be progressive and irreversible in nature in order to differentially rule out treatable conditions with an abrupt onset, such as serious injury, medical illness, medication side effects, or a vitamin deficiency.

For a diagnosis of dementia, DSM-5 (American Psychiatric Association, 2013) diagnostic criteria require evidence of multiple cognitive impairments, manifested by both memory impairment and at least one language disturbance. When memory issues begin to surface early on in the progression of the disease, an individual may be able to recall events long past, but be unable to state what they had for lunch that same day. Other types of information, such as where the keys were placed, telephone information, addresses, and shopping lists are reportedly forgotten, thus impairing every day functioning. The individual may be unable to remember the names of loved ones, or confuse one person with someone else, thus presenting challenges for ongoing social and family relationships. Additionally, the older adult may find difficulty recognizing and labeling common objects, such as a chair, dish, or plant. The inability to produce names and labels despite intact visual acuity is typical of agnosia. Apraxia is observed when the older adult has physiologically intact motor functions, but is unable to perform simple tasks (e.g., snapping their fingers, washing the dishes), even after demonstrating the ability to repeat instructions just heard pertaining to the task. Language disturbances can include aphasia. Over the course of the disease, the older adult’s speech may be limited

to only one or to favorite topics, or a repetitive series of questions. Speech may eventually stop completely, rendering the older adult mute although no structural damage exists to the speech apparatus.

Behavioral Excesses and Deficits

NCD is characterized by a cluster of observed behavioral deficits and excesses. The most common form of NCD, Alzheimer's, has the common features of memory problems as the first symptom, along with impaired language, and motor and visuospatial difficulties. Buchanan (2006) describes common behaviorial excesses observed in older adults with NCD. Physical aggression tends to be common during activities of daily living, in particular during bathing routines, and in other situations when staff and the older adult are in close proximity. Verbal aggression, or threats of harm to others, is also common. Along with increased caregiver burnout, physical aggression increases the risk of falls when the older adult is engaging in an aggressive episode. These excesses, along with behavioral deficits in memory and speech as described above, can create considerable strain on loved ones, staffing agencies, and communities.

Prevalence of and Projections for Dementia

NCD is prevalent among older adults. In the United States, approximately 6.8% of adults age 60 and older suffer from some form of dementia (World Health Organization and Alzheimer's Disease International, 2012). Total numbers, vary, but it is estimated to affect between 3.9 million and 5 million Americans (National Institute of Aging, 2014; Alzheimer's Association, 2014). Perhaps most concerning with these figures lies in the drastic increases in NCD rates among the older age groups. Turner (2003) reports that for every five years of age after 65, the incidence rate nearly doubles

(see also World Health Organization and Alzheimer's Disease International, 2012). For the oldest old (i.e., age 85 and above), one in every three has NCD (Alzheimer's Association, 2014). This is particularly troubling, as the oldest old are the fastest growing subgroup among older adults.

NCD projections reveal rapidly growing prevalence and incidence rates that mirror the anticipated demographic changes in the United States older adult population. As the number and proportion of older adults grow, so too do the number of NCD cases. The Alzheimer's Association (2014) indicates that as the baby boomer generation ages, the number of older adults with NCD will triple by 2050, from 5 million to about 16 million. Such staggering statistics support the notion that NCD is a growing public health concern.

Costs of NCD

The intensive amount of care provided to older adults with NCD due to the discussed behavioral excesses and deficits exhibited across the progression of the disease is costly. Hurd, Martorell, Delavande, Mullen, and Langa (2013) conducted detailed in-home clinical assessments with a subsample of older adults selected from a larger study. Measures of cost of care included out-of-pocket-spending (e.g., nursing home stays, medical visits, prescription drugs, etc.), spending by Medicare, net nursing home spending as distinguished by number of nights spent in a nursing home and nightly nursing home fees, and formal and informal home care (i.e., number of hours that either paid professionals or a member of the family provided assistance with activities of daily living or instrumental activities of daily living). Forgone wages, or time spent caregiving that could otherwise be spent earning income, was also included in the analysis. Results

indicated that the yearly estimated cost per person with NCD is $41,689. Based on this figure, the total monetary cost of NCD is estimated to be between $157 billion and $215 billion. Medicare pays $11 billion of the total cost of NCD care. Assuming that the prevalence rates continue to increase, these costs will double by 2040.

Pharmacological Treatment Models

Given the increasing number of older adults with NCD and its associated financial and social burden, effective treatments are warranted. The commonly employed first line intervention is psychopharmacological in nature. Brooks and Hoblyn (2007) report that among adult adults residing in a variety of settings, overall frequencies for psychotropic medication use were 28% for antipsychotics, 12% for antidepressants, and 33% for anxiolytics and hypnotics. These data are troubling, considering that many of these medications are prescribed off-label, with unknown efficacies for behavior management with older adults suffering from NCD. Further, the Food and Drug Administration has put black box warnings on such medications, highlighting the increased risks of short-term mortality if used with older adults. Pertaining to pharmacological treatments targeting memory decline, many purported cognitive enhancers have little, if any benefit, in the early stages of NCD (a robust review of psychotropic medications is outside the scope of this paper, but for a detailed review of medications used to treat behavior excesses and deficits in older adults with NCD, see Tsolaki, Pantazi, & Kazis, 2001, and Wang et al., 2006).

To address concerns on the reliance of medications and other restrictive interventions, the Omnibus Budget Reconciliation Act (OBRA, 1987) legislated the use of nonpharmacological interventions as a first course of intervention before turning to

psychotropic drugs. This act also requires that gradual dose reductions occur when older adults living in skilled nursing facilities receive medication, along with behavioral interventions to ultimately discontinue the use of drugs as an attempt to fade out more intrusive behaviors management procedures (Lundervold & Lewin, 1992). The spirit of OBRA constitutes a shift in how behaviors observed with older adults with NCD should be treated, and Behavior Analysis is well-suited to provide behavioral interventions.

Behavioral Gerontology

Properly applied, behavioral interventions have shown to effectively treat a range of behaviors. As defined by Burgio and Burgio (1986), behavioral gerontology is "…the study of how antecedent and consequent environmental events interact with the aging organism to produce behavior" (p. 321). Knowledge, therefore, of the environmental circumstances in which the older adult engages is also imperative. Functional analysis refers to the various assessment methods that share the goal of identifying environmental events that influence a given target behavior. Tools used as part of the assessment can include informant interviews, descriptive observations, and experimental functional analyses. As a best practice in behavioral gerontology, the identified events are then applied systematically in treatment.

Several behavior analysts, in addition to Burgio and Burgio (1986) have spoken to the importance of applying behavioral technologies to the problems older adults with NCD face (Bourgeois, 1991; Fisher & Carstensen, 1990; Buchanan, Christenson, Houlihan, & Otrom, 2011; Fisher & Swingen, 1997; Williams & Ascione, 1983). Behavior analytic research has shown that changes to the environment couched in functional analysis can alleviate various problems. By conducting an experimental

functional analysis where environmental variables are systematically altered across conditions, the function of problem behavior for three older adults in a long-term care setting was identified and used to inform treatment, which proved to be effective (Dwyer-Moore & Dixon, 2007). For an older adult who was aggressive during self-care routines, Baker et al. (2006) trained the nursing assistant to apply noncontingent breaks into the routine after identifying escape as the reinforcer maintaining challenging behavior. This resulted in drastic reductions in aggression, allowing for care to be completed more quickly and comfortably.

Function-based interventions have also been successfully used to address hoarding (Baker, Leblanc, Raetz, & Hilton, 2011), disruptive vocalizations (Buchanan & Fisher, 2002), promote increased independence in ADL routines by training staff to use least-to-most prompting techniques (Engelman, Altus, Mosier, & Mathews, 2003), and re-establish vocal mands (Oleson & Baker, 2014). Methods to increase activity engagement have included structured choice making (Leblanc, Cherup, Feliciano, & Sidener, 2006) and descriptive verbal prompts (Engelman, Altus, & Mathews, 1999), to name a few.

These applied efforts strongly suggest that older adults with NCD are sensitive to changes behavior-environment relations, and that the principles of behavior are operative with this population. Related to persistent outcomes in the face of anticipated decline, Trahan, Donaldson, McNabney, and Kahng (2014) taught three older adults with NCD to use a communication card to gain access to a highly preferred activity. Following training, the participants engaged in the communicative response at high rates. Importantly, the effects of such training persisted across several months without any

booster sessions. These results suggest that older adults with NCD can learn new adaptive communicative responses, and that these operants remain at strength across time. Such interventions demonstrate that communication skills can be relearned, at least in some contexts.

Interventions that specifically target vocal verbal behavior with the goal of improving memory performance is seldom researched in behavior analysis. As of this writing, only one published article targeted vocal verbal behavior. Dixon, Baker, and Sadowski (2011) reported on training collateral responses (i.e., tacts, echoics, and intraverbals) to pictures in the presence of a contextual stimulus (i.e., a colored background). By requiring additional verbal responses in the presence of these stimuli, delayed recall improved slightly as compared to recall with pictures set against a different colored background to which collateral responses were not trained.

More research is needed in this area, as memory deficits are a common deficit observed in older adults with NCD, creating stress to both the individual and family. In a review of behavioral gerontology articles published in the *Journal of Applied Behavior Analysis*, Trahan , Kahng, Fisher, and Hausman (2011) provide similar sentiments by urging behavior analysts to broaden their applied efforts by targeting memory deficits. Clinicians outside of behavior analysis, however, are using various interventions, the natures of which are briefly described below.

Memory Interventions

Changes in memory functioning, as described earlier, causes significant impairment in daily functioning. Stimuli that once occasioned behavior no longer evoke the appropriate response. This is commonly seen, for example, when the older adult with

NCD experiences great difficulty in remembering new information presented, such as where their room is located in the retirement community, or the name of their caregiver. In addition, long-term memory deficits are observed. Memory deficits, in general, can be conceptualized as deficits in stimulus control. Prosthetic memory aids have been researched and have found to be useful in guiding behaviors in various contexts (Bourgeois & Hickey, 2009). External aids can include calendars, exaggerated signs and labels, memory books, and other stimuli that help to remind the older adult of important events, such as when to take medications and how to locate the bathroom. Bourgeois (1993), for example, trained older adults to use a memory wallet that contained various conversational points (e.g., daily schedule, family), and observed increased in on-topic statements made with other residents.

These and other external memory aids have value in the maintenance of independence for as long as possible in the face of progressive decline with NCD. Another challenge faced, however, is that the older adult may forget to use the memory aid itself. Training the use of memory aids is therefore paramount to successful implementation. Several techniques have been evaluated, including the use of vanishing cues and cueing hierarchies to train a particular skill (see Bourgeois & Hickey, 2009, for a review of these and other procedures). Spaced retrieval is another form of memory training that has received attention in the research literature.

Spaced Retrieval

Spaced retrieval (SR) is a procedure developed outside of behavior analysis that has been gaining empirical support for its effectiveness to teach simple associations that the older adult has shown to “forget”. SR procedures are described in the literature as

being characterized by conducting cued-response trials across specifiable periods of time within session, as opposed to conducting such trials in a massed format (Camp, Foss, O'Hanlon, & Stevens, 1996). More specifically, the recall delay value between the presentation of a set of verbal instructions and the contextual cue to engage in a target behavior is systematically adjusted according to prior performance. The verbal instructions presented at the start of the session specify a conditional discrimination task (e.g., "When the buzzer sounds, I would like you to hand me the picture of Sue and say that her name is Sue"). Given correct responding following the presentation of the cue, the recall delay value to the next cue presentation is increased. If an error is made, the researcher restates the verbal instructions and prompts an echoic response with the participant. The following recall delay value is then reduced to the previous delay value where responding was correct. The procedure essentially provides ample practice of responding across time, where sessions are arranged as a social visit to reduce frustration, fatigue, and testing anxiety.

A number of relations have been trained that persist across significant durations of time. Procedures reported vary according to the intervention target, but the critical manipulation of spaced recalls adjusted based on performance is the common thread. Behaviors successfully trained with SR include compensatory swallowing in an older adult with dysphagia (Brush & Camp, 1998b), naming common household items (Fridriksson, Holland, Beeson, & Morrow, 2005), using external memory aids (Bier et al., 2008), and independent use of walking aids (Creighton, Davison, Ploeg, Camp, & O'Connor, 2015). SR effects have been replicated across NCD types, including Korsakhoff's Syndrome (Camp & Schaller, 1989). Alzheimer's (Cherry, Hawley,

Jackson, & Boudreaux (2009), and vascular NCD (Brush & Camp, 1998a). SR techniques have been shown to be effectively delivered by direct care staff in residential settings as well (Hunter, Ward, & Camp, 2012).

Throughout the SR literature, two classes of behavioral relations have been targeted in session. The first class can be summarized as intraverbal targets, in which the SR cue takes the form of a question (e.g., "What do you do when you want to walk?") and the targeted response is "Use my walker". The second class can be summarized as tact targets, where nonverbal stimuli are presented as the cue (e.g., a beeper, pictures of various household objects) and the target response is naming the item. In both intraverbal and tact relation targeting, the researchers' goal is to also evaluate transfer of the target relation to naturalistic settings (e.g., increased use of the walker when mobile, and naming objects that they had previously demonstrated difficulty with, respectively). The nonverbal response (e.g., actual use of walker) may be required as part of a correct response (e.g., Brush & Camp, 1998a) or not (e.g., Hunter, Ward, & Camp, 2012). This is important to note, as requiring the participants to engage in the nonverbal behavior along with the intraverbal response would presumably influence generalization of the skill to the natural setting.

The results of these studies using SR techniques are promising. Several concerns exist, however, as noted by Creighton, Ploeg, and O'Conner (2013) in a review of SR. The authors report that of the 34 studies reviewed, only 11 included a control condition. This greatly reduces the internal validity of the remaining 23 published articles because the methodology does not allow for maturation effects, or any changes in performance over time in the absence of intervention, to be ruled out. Furthermore, the ecological

validity of teaching name-face and object-name relations is called into question, as some of the target stimuli were either unknown to the participants (e.g., a person's face that the participant had never encountered before), or objects that would be rarely encountered in daily life (e.g., an octopus).

In addition, Creighton et al. (2013) state that the types of filler activities that take place during long recall delays is either poorly described, or varies considerably across studies (e.g., reporting in the methods section that researcher and participant engaged in light conversation, or that tasks such as doing crossword puzzles were introduced during these delays). Events that occur between SR trials are arguably critical variables in and of themselves, as certain activities could facilitate or hinder performance.

Other procedures that are not clearly described in all reports targeting name-face relations include the possible inclusion of non-example presentations, which would provide evidence for any rote responding during SR training (e.g., if the target is naming a person, then pictures of other people are presented at times to test if the older adult has a tendency to name all faces as the target). With SR training with photographs, it is unclear if the stimulus array is hidden in between cue presentations, or is visible to the participant throughout. This is a variable that may confound results. The participant may continue to look at the target photograph and repeat the name, or view the re-arrangement of the array in between trials.

Oren, Willerton, and Small (2014) also conducted a review of SR, and provide additional points to consider. Along with issues of procedural clarity, no measures of treatment integrity or interobserver agreement have been reported in the literature. They recommend that to help address internal validity concerns, within-subject multiple

baseline designs could increase the generality of the findings when the subject pool is small, to improve upon experimental control by having each participant serve as their own control, and to increase the sensitivity of treatment outcomes. Finally, they recommend that when training name-face relations, that live person transfer tests be incorporated into the methodology.

Only four published studies using SR to teach name-face relations embedded live person transfer tests within the protocol. This test involves the introduction of the actual person whose photograph serves as the training stimulus during SR sessions. The person enters the room, gives the experimenter a written message, and then sits down at the testing table. After allowing some time for spontaneous recall of the target person's name, the experimenter provides a verbal prompt (i.e., "This is my friend, do you know her name?"). If responding is still not evoked, a further prompt is delivered (i.e., Her picture is on the board. Would you hand me her picture?"). The final prompting sequence in this test involves the experimenter give the photograph of the target person to the participant and stating, "Take another look at the picture. Now can you tell me her name?" These tasks within the testing condition are scored as pass or fail (Cherry, Walvoord, & Hawley, 2010; Cherry, Hawley, Jackson, & Boudreaux, 2009; Hawley & Cherry, 2004; and Hawley, Cherry, Boudreaux, & Jackson, 2008).

The following numbers reported are in aggregate, as the results of the live person transfer test were only displayed in group format. Out of six participants in total in the Cherry et al. (2009) study, only 2% demonstrated spontaneous recall. With the additional verbal prompt, this percentage increased to 20%. The prompted selection response task further increased correct responding among 66%. Finally, handing the picture to the

participant and prompting the response verbally resulted in 24% of participants accurately stating the target's name. These data are of concern, as they reveal poor transfer of the target skill to more naturalistic settings (e.g., when the target person visits the older adult).

Cherry, Walvoord, & Hawley (2010) and Hawley and Cherry (2004) report similar findings, with no spontaneous recall among participants, some evidence of responding when verbally prompted, and the greatest amount of correct responding when prompted to engage in the selection response. Hawley et al. (2008) report live person transfer findings by participant, with one out of 12 participants engaging in spontaneous recall on at least one out of four tests. As with the previous studies, responding was most consistently evoked during the selection response test. These findings suggest that although correct responding persists both within and across SR training sessions, transfer to the actual person is limited.

An analysis of the stimulus and response conditions of training as compared to the tasks tested in the live person transfer situation can shed light on why reliable responding established during training fails to transfer in the presence of the target person. As initially suggested by Sidman (1971) and further explored by Baker, Leblanc, and Raetz (2008), analyzing the stimuli set up to guide responding and the types of behavior targeted increases the extent of operational specificity by providing rich information on the precise nature of the training and testing environments. This also allows for a thoroughgoing understanding of what relations may be contributing to the poor transfer of control observed.

During SR training, the stimuli arranged that help guide a target response include a) an audible cue (i.e., a buzzer), and b) the presence of the target photograph. The buzzer functions as a conditional discriminative stimulus, in that it alters the evocative function of the target photograph in the stimulus array. The presence of the buzzer and the target photograph then guide several responses: a selection response (i.e., taking the correct photograph from the array and handing it to the experimenter), along with a vocal verbal response (i.e., stating the name of the person in the photograph). This verbal response can be categorized as a tact, as its form is under the functional control of a prior discriminative stimulus (i.e., the photograph), and is maintained by generalized conditioned reinforcement (i.e., affirmative feedback by the experimenter) (Skinner, 1957). Eventually, responding becomes reliably evoked by these conditions.

The stimulus conditions in the live person test, however, are quite different. There is no reliable audible cue, nor is the array of photographs in view. Instead, the stimulus presented is the sight and sounds of the actual person entering the room and briefly speaking with the experimenter. Because the stimulus properties of the actual person do not share common physical dimensions with a two-dimensional photograph of that person, behavior is not evoked. The testing conditions, thus, are too different from the original training sessions to support transfer. Further evidence for this lies in the boosted performance when the photograph array is reintroduced. When the test stimulus conditions begin to take a similar form to the training conditions, transfer is more reliably observed. Thus, overall previous SR research in teaching names has utilized the “train and hope for generalization” method instead of incorporating generalization technologies in training conditions. Brush and Camp (1998a) and Camp and Schaller (1989) report on

two notable exceptions, in which the individual whose name was targeted in spaced retrieval was present throughout all spaced retrieval sessions, allowing for trials to be conducted directly in the presence of the stimulus in which responding is desired. In the context of service provision this may be feasible, but in other situations access to the individual may be difficult (e.g., the family member does not live in town).

Purpose

In a conceptual paper, Stokes and Baer (1977) proposed several behavioral techniques for promoting generalization of a target skill acquired in a training session to other contexts in which the behavior should occur. In their description of training sufficient exemplars, Stokes and Baer suggest that after mastering one exemplar, training continues with a second exemplar, and then potentially a third and so on until extensive generalization effects are observed in untrained stimulus conditions. This description has clear relevance to SR, where only one exemplar is trained. The training of additional stimuli within a stimulus class (e.g., various visual presentations of a target person) may therefore increase the probability of correct responding during live-person transfer tests, thereby increasing the effectiveness of SR and of the social validity of the technique. In order to assess for these stimulus deficits and the effects of exemplar training, probes that test extensions of the trained response across untrained stimuli are also needed.

Since Stokes and Baer's seminal paper, several investigations on developing and evaluating procedures to train exemplars have been done across a wide range of target skills in the treatment of individuals with autism. Marzullo-Kerth, Reeve, Reeve, & Townsend (2011) used multiple exemplar training to establish a generalized repertoire of sharing items, and Brown, Peace, & Parsons (2009) successfully trained generalized

imitation skills through exemplar training. More recently, Wunderlich, Vollmer, Donaldson, & Phillips (2014) targeted tacts using exemplar training and observed stimulus generalization to untrained stimuli. Because such behavioral deficits of "memory" in older adults can be conceptualized as issues in stimulus control, exemplar training is a defensible strategy. Older adults with major NCD may therefore also benefit from these procedures by reestablishing verbal responses with respect to important people in their lives. This can arguably improve the quality of their life by facilitating greater amounts of meaningful social interactions. The present study therefore sought to 1) evaluate the effects produced by SR using multiple probe techniques, and 2) determine the extent to which intervening on stimulus control deficits promotes transfer across generalization probes.

METHOD

Participants

Seven participants were originally recruited for the purposes of this study. Two participants demonstrated difficulty in following simple instructions, as evidenced during the administration of the cognitive screeners, and also exhibited significant vision and hearing impairment. For example, one of the older adults whose participation from the study was terminated engaged in verbal errors when attempting to repeat the name of a caregiver during the delayed recall portion of the behavioral assessment. A third participant who had successfully met inclusionary criteria for the study was terminated from participation after one session of spaced retrieval due to failing health as a result of kidney infection.

Four participants were therefore included in this study (2 males and 2 females). Ages ranged from 78 to 98 years old. All participants lived in an assisted living facility within a larger continuing care retirement community in Southwest Michigan, with either a confirmed physician's diagnoses of NCD as indicated by medical records review, or probable NCD. All participants were Caucasian and held advanced college degrees. Participants' caregivers reported symptoms that included difficulty remembering the names of staff that were routinely involved in their daily care. Exclusionary criteria included: a) the inability to follow simple directions such as those included in the experimental protocol, b) severe visual and auditory impairments that would interfere with the participant's ability to sense the experimental stimuli, and c) a score on the Geriatric Depression Scale (GDS) that indicated evidence of depression. Refer to table 1 for a summary of participants' age, diagnosis, and scores on researcher-administered screeners.

Table 1

Participant Characteristics

Participant	Sex	Age	Diagnosis	MMSE	GDS	CDT
Harriet	Female	78	Vascular Dementia	17	0	3
Barry	Male	97	Vascular Dementia	14	1	3
Susan	Female	91	Vascular Dementia	15	3	4
George	Male	98	Probable Dementia	14	0	3

Note. MMSE = Mini-mental Status Exam. GDS = Geriatric Depression Scale. CDT = Clock-Drawing Test.

Screeners

Three screeners were administered. These included the Mini-Mental Status Exam (MMSE), Clock-Drawing Test (CDT), and Geriatric Depression Scale-Short Form (GDS). These screeners have been extensively used in NCD research and have good

psychometric properties (Folstein, Folstetin, & McHugh, 1975; Lopez et al., 2005; Lourenco & Veras, 2006; Schramm et al., 2002). The MMSE is a 30-point cognitive screener that assesses current functioning by presenting a series of questions and/or tasks relating to orientation, recall, naming objects, copying figures, problem solving, and sentence construction. In terms of a single cutoff score, scores below a 24 suggest abnormal results given a college education. The GDS is a screening instrument for depression that is valid with older adults with cognitive impairment. Fifteen questions are asked in oral form, with scores above 5 suggestive of depression. The CDT provides a brief screening on visuospatial impairments that may indicate NCD. The participant is given a piece of paper with a large circle at the center and is instructed to a draw numbers in the circle as they would appear on the face of a clock, and to make the hands of the clock read ten after eleven. Scores range from 1 (i.e., perfect representation) to 6 (no reasonable representation of a clock), with a cutoff of 3 or above indicative of cognitive impairment. These measures were given to help confirm, to the best they can, of probable NCD, as well as to rule out depression as a cause for any observed cognitive deficits. No diagnosis was made as a result of information obtained from these screeners.

Barry (all names are pseudonyms) was a 97 year-old male residing on the third floor of the assisted living facility. He had a diagnosis of Alzheimer’s as indicated in his medical records. He was mobile with a walker. Barry would engage in perseverative conversational speech during interactions. This took the form of speaking about a toy that he patented during high school, and would occur during unrelated points in the conversation. Barry did not wear glasses, nor did he use hearing aides. Barry’s daily medication regimen included 20 mg of Namenda XR, a cognitive enhancer, which was

held constant throughout the course of his participation in the current study. He scored 14 on the Mini-Mental Status Exam (MMSE), suggesting moderate cognitive impairment. His GDS score was 1, which indicated that he was not depressed at the time of this study. Finally, he scored a 3 on the Clock-Drawing Test (CDT), indicating cognitive impairment.

Harriet was a 78 year-old female who lived on the first floor of the assisted living facility, with a diagnosis of vascular NCD. The first floor served as a special care unit for older adults with NCD. The unit an outdoor garden and patio, and had locked exit doors that could only be opened with the correct keypad code. Harriet was mobile with a walker, and engaged in delusional speech in the form of stating that she was the boss of this place. She would also assert to caregivers and the researchers that her family was made up of doctors. Harriet did not wear glasses or hearing aids. Harriet's MMSE score was 17, suggesting moderate cognitive impairment. Her GDS score was a 0, indicating that she did not endorse any statements suggestive of depression. Lastly, her CDT score was a 3, providing further evidence of cognitive impairment.

Susan was a 91 year-old female who lived on the third floor of the assisted living facility. She had received a diagnosis of vascular NCD as indicated in her medical records, wore glasses, was mobile with a walker, and required assistance with transports to and from sitting positions. Susan exhibited delusional speech by stating on occasion that she had bugs in her apartment. Her daily medication regimen included psychotropic prescriptions. She took 75 mg of Seroquel, an antipsychotic medication, and 20 mg of Celexa, an antidepressant. Throughout her participation, her psychotropic medications remained constant. Susan's MMSE score was 15, indicating moderate cognitive

impairment. She scored a 3 on the GDS, suggesting that she was not depressed, and a score of 4 on the CDT, providing further evidence of cognitive impairment.

George was a 98 year-old male with probable NCD residing on the third floor of the assisted living facility. He wore glasses and hearing aides, and was mobile with a walker. He frequently took walks throughout the hallways before meals, where he would stop halfway at a window to view the scenery outside. As part of his daily medication regimen, George took 25 mg of Trazodone, an anti-depressant, at bedtime. This remained constant throughout the length of his participation. He scored 14 on the MMSE, indicative of moderate cognitive impairment, a score of 0 on the GDS, and a score of 3 on the CDT.

Recruitment

Participants were recruited through a continuing care retirement community located in the Southwest Michigan region. The site employs a home-style level of assisted living, with two caregivers to each wing who assist with leisure activities, activities of daily living, and the delivery of medication regimens. Participants were selected based on staff referral for memory concerns as they relate to difficulties in naming individuals, and direct observation of purported memory concerns.

Informed consent was obtained before moving on to pre-assessment measures. All participants had an individual with durable power of attorney serving as legal guardian. Thus, the informed consent process was initiated with the individual with legal authority over the older adult of interest. To ensure compliance with site and federal privacy policies (e.g., HIPAA), staff first contacted the person with power of attorney to obtain permission for the researcher to call and request a meeting at the recruitment site

to discuss the research project. If meeting on site was not possible for the representative power of attorney, the project was described over the phone and a copy of the informed consent document was mailed to a location of their choosing to sign and return via surface mail. In addition to obtaining informed consent by the legal authority, assent was given verbally by the participant to reduce the possibility of coercion.

Setting and Materials

All sessions took place in the participants' respective apartments within the retirement community. Adequate lighting was ensured by opening window blinds, or by turning on a sufficient number of ceiling and desk lights. The researcher sat facing the participant, separated by a small table. Sessions were between 30 and 50 minutes in length, were held three to five times per week, and occurred between 9:30 AM and 12:30 PM.

A 16.75 x 13.75 in. black foam board was placed on the table, with neon green tape used to create a 3 x 3 matrix. All photographs of staff were taken with a digital camera, printed to 4 x 3 in. dimensions, laminated, and then mounted on 1 x 1 in. pieces of foam board to increase the ease with which participants picked up the photos.

All staff photographs had a white wall as the backdrop. The camera distance was held constant, such that photographs were of the torso and above. Multiple photographs were taken of staff members, which included full face (i.e., the individual looked straight on to the camera), three-fourths view (i.e., the individual turns slightly to one direction), two-thirds view (i.e., the individual turns at an even greater direction), and profile view (i.e., the individual is turned 90 degrees from the front). In addition, staff member

photographs were taken with different clothing arrangements, and, if they wore glasses, photographs were taken with the glasses both on and off.

Photograph arrays during each SR session were of a constant mix of staff members with various hair styles, hair color, attire, age, and race to control for the extent to which extraneous stimulus properties gained influence over performance. Arrays always included at least one photograph of each viewing angle. This was to control for the possibility of irrelevant features of the stimuli controlling responding (e.g., selecting a particular photograph because it was a profile shot, and not because it was the training stimulus). Refer to Appendices A and B for example stimulus sets and SR arrays.

Target (S+) Stimulus Class. The targets for intervention were selected based upon frequency of contact the staff member had with the participant, and availability for live-person probes. A target stimulus class included multiple photographs and videos of the staff member. For Harriet, the resident care coordinator served as the target for intervention, as this staff member was closely involved in overseeing Harriet's care, and had common interactions with her throughout her shifts. For Barry, George, and Susan, the activity director served as the intervention target. Barry, George, and Susan all participated in-group activity programming several times per week, which were either led or assisted by the activity director.

Video Probes. Brief videos were taken with a Canon Vixia HF R10 HD Camcorder of non-targeted staff members, as well as of the target staff member. Videos consisted of various settings, such as a hallway, an office, and dining room, and were of a series of social interactions with the person taking the video footage (e.g., talking about upcoming activities, the weather, asking for directions to the health center, and so on).

These videos were then spliced into 10-s clips using video editing software and shown to participants on a 2009 white Macbook. Both the screen brightness and video volume were turned to max settings to ensure the participant could hear and see the videos.

Textual Prompts. Any text was printed on white strips of paper in bolded Times New Roman and font size 28. Pictures used during the activity during delays were of various landscapes and architecture (e.g., a baseball stadium, a mountain range). Other session materials include paper and pencil data sheets, a beeper, and a stopwatch.

Response Definitions

A correct response during spaced retrieval trials was scored when the participant a) selected the target photograph, and b) named the individual in the photograph within 10-s of the beeper sounding. During probe trials, a correct response was scored when the participant stated the name of the individual in the generalization probe within 10-s in the presence of an example probe trial. A correct response was also considered when the participant stated “I don’t know” or made some form of statement that indicates that they are not familiar with the person’s name during S- (non-example) probe trials. Below, three categories of dependent variables are described, followed by two types of dependent measures that were used to rule out alternative explanations for performance.

Measures of Acquisition. Dependent variables of interest included the percentage of correct responding across trials per session and the percentage of errors made across trials. In addition, errors were analyzed across both selection and naming responses per session.

Measures of Recall. Median delay at which responding was correct per session were calculated for analysis of behavior maintained across session time. This measure served as the primary dependent variable.

A first trial probe was conducted at the start of each session to determine the extent to which participants can correctly select the target photograph and state the target's name after the passage of time between sessions. Data from these probes were transformed into cumulative number of correct cold probes across sessions.

Measures of Generalization. Correct responding was measured across exemplars and non-examples to test for control by the target stimulus class as training progressed. This included responding across untrained exemplar photographs (i.e., naming probes), video probes, and live-person probes.

S+ (Example) probes. These probe types involved the stimulus presentation from a member of the target stimulus class. Such probes measured the extent to which spaced retrieval training transferred across the target stimulus class.

S- (Non-example) probes. These probes involved the stimulus presentation of members from the non-targeted stimulus class that served as distractors in the array (i.e., a staff member who was not the individual to be targeted during spaced retrieval). Such probes measured the extent to which the participant could discriminate when they did not know the name of a staff member, and therefore responses indicating that they did not know the staff member's name were scored as correct.

Visual Matching-to-sample. Along with generalization probes that assessed for naming, correct selection responses were measured during matching-to-sample probes to rule out vision impairment as a possibility for poor performance during spaced retrieval

sessions. Such probes also provided evidence for generalization within such selected stimulus sets (and thus control by a stimulus class).

Control Measures. Both types of control probes were administered in a semi-random fashion during spaced retrieval sessions, with the delay value held constant for a given probe as determined by the delay at which responding was previously correct during spaced retrieval.

Hidden probes. During this probe type, the photograph array was rearranged out of the participant's sight in between spaced retrieval trials. This was done to test for any positional control over performance (i.e., to rule out the possibility that participants were being influenced by seeing the rearrangement of the array occurring in front of them).

New-target probes. During this probe type, a photograph of a different staff member was now targeted in SR training. The new target was always previously used as a distractor stimulus. The researcher would begin by stating the instructions used at the start of each spaced retrieval session and emphasized that a new photograph was being targeted (e.g., "Now whenever you hear the buzzer sound, I would like you to hand me the picture of Linda and tell me that her name is Linda."). The new-target probe provided information on whether or not the participant being familiar with the photograph alone produced correct responding.

Inter-observer Agreement (IOA) on Dependent Measures. A secondary observer concurrently and independently collected session data across 39% of sessions across all conditions. Afterwards, the primary and secondary data collectors' data sheets were compared to determine counts of agreements (i.e., whether a correct or incorrect participant response was scored by both observers) and disagreements (i.e., whether one

observer scored a correct response while the other scored an incorrect response for that trial, and vice versa). Trial-by-trial IOA was calculated by dividing the number of agreements for each trial by the number of agreements plus disagreements per trial. IOA was 96% (range = 73-100%).

Independent Variables

The primary variable to be manipulated is the spaced retrieval training procedure described in the research literature (e.g., Cherry, Walvoord, & Hawley, 2010) , which entailed presenting recall trials at adjusting lengths of delay based on participant performance, with feedback given for errors and descriptive praise provided for correct responses.

Other independent variables introduced were identified based on stimulus control deficits (failure of spaced retrieval training), and include **exemplar training** embedded within the spaced retrieval protocol for Barry and George. Exemplar training entailed training the naming the selection and naming responses in the presence of new photographs of the target person with differing nonessential properties (e.g. the person at different angles, wearing different clothing), in a serial format. **Prompted orienting responses** were added to spaced retrieval trials for Susan, and **textual prompts** embedded within spaced retrieval were used with Harriet.

Research Design

A multiple baseline design across participants with embedded probes was used to evaluate the differential effects of spaced retrieval and modified spaced retrieval protocols on measures of generalization. Harriet and Barry were concurrently run first,

followed by Susan and George. Multiple probes entailed non-identity matching-to-sample, naming, and video probes for all participants.

Research Procedures

General Procedures. Sessions began with the researcher casually talking with the participant in order to build rapport. At any point during the session the participant requests a break, the researcher provided five minutes for the participant to use the bathroom, talk a short walk, or obtain a drink. In addition to requested breaks, noncontingent break opportunities occurred every 15 minutes with the researcher asking the participant if they would like a break for a few minutes. During all types of probes, no programmed consequences were delivered (i.e., no praise statements or feedback of any kind given). The researcher would neutrally say "ok", and move on to the next trial.

Baseline Assessment. ***Naming probes.*** The researcher presented three exemplars of the target staff member one at time to the participant and prompted them to name the person in the photo (e.g., "Tell me who this is"). Three S- (non-examples) were also presented in this format, interspersed across example photographs. The three example photographs were randomly selected from a pool of photographs of the target person, with no single example photograph being re-presented more than two times across sessions.

Recall probes. To probe for the extent to which the participant can successfully repeat the name after a delay in the absence of any intervention, the researcher presented photographs of staff members that were non-targets and presented their name to the participant (e.g., "This is Linda"), followed by a prompt for the participant to repeat the name in the presence of the photograph. The researcher would then remove the

photograph and engage the participant in a distractor task by holding another picture up of a scene (e.g., a sailboat on the lake), and ask a series of standard questions. After 60-s had elapsed, the researcher presented the photograph that was shown earlier and prompted the participant to name the staff member. Only non-targeted staff members were used in this portion of the assessment in order to prevent acquisition of the target name during baseline.

Non-identity matching-to-sample probes. The experimenter presented a given staff member's photograph as the sample, and instructed the participant to point to another photograph of the person that was included in the matrix array with distractor photographs present. The correct comparison stimulus was always of a different photograph of that same person presented as the sample with either a) different attire, or b) at a different viewing angle. Three example and three S- (non-example) trials were presented.

Video probes. Video probes involved presenting a video to the participant (e.g., "I have a video I would like to show you"). If the participant indicated that they had difficulty viewing the video, the researcher repositioned the video screen to reduce any glare from external lighting sources or to improve the viewing angle. The video would then be played once more. At the end of the video, the researcher prompted the participant to name the person in the video. Three example videos of the target staff member were presented, and interspersed among them were three videos of non-targeted staff members (i.e., S- presentations)

Live person probes. These probes would provide evidence for or against responding to transfer to the presence of the actual three-dimensional person being

targeted. While the researcher was sitting with the participant, the target person entered the room, delivered a note to the researcher, turned towards the participant and greeted them. Requiring the person to greet them aided in establishing a naturalistic social environment that provided an opportunity for the participant to state the name. The researcher waited 10-s after the person sat down to allow for correct responding to occur independently.

After 10-s, the researcher sounded the discriminative stimulus used in training. If the stimulus did not evoke a correct response after 10-s, the researcher supplemented it by asking if the participant knew who the person was. This probe constituted a positive probe. Negative probes entailed the same procedures as above, but a non-example entering the room (i.e., a person not targeted during the training). Both the person from the target stimulus class and an unrelated person was used to help determine if responding was rote (i.e., they will state the name of the target regardless of who is present).

Independent Variables. ***Spaced retrieval training.*** Spaced retrieval trials began following at least two sessions of baseline assessment. The matrix board was placed in front of the participant. Photographs were presented individually and then placed on the matrix. The researcher stated the name of the person in the photograph during each presentation. Each array always consisted of eight distractors along with the target photograph.

After all photographs had been placed on the matrix and named by the researcher, the beeper was introduced to the participant. A brief set of instructions was presented to the participant to establish the sound of the beeper as a discriminative stimulus for

scanning the array (e.g., “When the beeper sounds, I want you to hand me the picture of Walter and tell me his name is Walter”). A practice trial would then take place immediately after the instructions on the first session in order to ascertain that the participant could sense the beeper and that it would evoke the target selection and naming response. After every trial, the position of the photographs in the array was rearranged to control for positioning effects. Following the final SR trial of the session, matching-to-sample, naming, and video probes were administered. These probes were carried out the same as in baseline.

The interval schedule used by Cherry and Simmons-D’Gerolamo (1999) was adopted, which started with a 5-s recall delay. If correct responding was evoked by the cue after this initial 5-s delay value, then recall delays were increased to 10, 20, 40, and 60 seconds. After correct responding at a 60-s delay, recall delays were increase by 30 seconds. After a 180-s recall delay with correct responding, delays were increased by 120-s. If an incorrect response occurred, the next recall delay value would return to the length of the preceding delay where correct responding was evoked. For example, if after correctly responding at a 40-s recall delay an error is recorded at the 60-s delay, the next delay would reset to 40-s. If another error is made, then the recall delay was further reduced to 20-s, and so on. After an error, immediate feedback was provided to the participant in the form of stating the correct name, pointing to the correct photograph, and restating the instructions. The researcher would then prompt the participant to repeat the target response. To prevent long periods of time with the participant guessing and potentially becoming frustrated, or if the participant was nonresponsive, a latency

termination criterion of 10-s was used after the presentation of the cue to determine when to provide the feedback.

At the start of all subsequent sessions, a first trial probe was administered to determine if the effects of the spaced retrieval training had carried over from the previous session. The researcher would sound the beeper cue after placing the matrix board with all photographs in front of the participant and wait 10-s for a response. No programmed consequences were given for correct, incorrect, or non-responses by the participant. Performance during these cold probes is graphically displayed at the bottom panel of each participant's recall and recall error displays (refer to figures 3, 5, 7, and 9). The researcher would then remove the photographs, and begin introducing them one by one to initiate spaced retrieval trials.

The events that take place during recall delays of 60 seconds or greater were standardized to increase the extent to which the environment was held constant, allowing for increased control by reducing potential confounding stimuli introduced during such delays. The researcher had on hand a set of pictures that were presented to the participant. A standard set of open-ended questions was then asked in relation to those pictures (e.g., "Does this remind you of anything?") leading up to the next spaced retrieval trial.

Spaced retrieval + exemplar training. Exemplar training embedded within SR began after three consecutive errors at a given delay interval during the SR condition, or after three consecutive sessions in which the longest delay value reached did not differ across more than two sequential delay intervals, and poor transfer across at least one generalization probe measure was obtained. These criteria constituted a "delay limit" and

provided evidence for the greatest delay at which responding stabilized with poor generalization.

The trial presentation delay value from the last SR session carried over to the first session in the exemplar training condition. Exemplar training was used to train responding across additional photographs of the target person in a sequential fashion. The first session of this phase began by introducing the materials as noted at the beginning of spaced retrieval, except that a different exemplar was used. All other procedures were the same as in the SR condition. Once correct responding stabilized according to the phase change criteria as noted above, a third exemplar was trained.

Spaced retrieval + orienting response requirement. Susan displayed frequent errors during two sessions of SR by selecting the distractor photograph in which the target photograph was located during the previous trial. Thus, an orienting response requirement was added to promote scanning the entire array. At the start of this modified condition, all photographs were named and placed on the matrix board as before. The experimenter then prompted Susan to engage in an orienting response by instructing her to point to a photograph (or photographs) with a particular stimulus property, such as individuals wearing a white shirt, or those who had blonde hair. After she selected the corresponding photographs, the researcher delivered feedback (e.g., "Yes, that's right, there are three people who are blonde!"). Then, Susan was prompted to identify a stimulus property of the target photograph (e.g., "Point to the picture with the woman wearing a black sweater"). Following the selection, the researcher again delivered feedback. The delay to the SR trial was then initiated. Following each SR trial, the photograph array was rearranged and the two orienting responses were prompted.

Spaced retrieval + textual prompt. Harriet did not reach phase change criteria during the embedded exemplar training within SR condition. Thus, a textual prompt was added to the target photograph. The textual prompt was faded by two letters after three consecutive correct SR trials (i.e., from "Kristen" to "Krist", then "Krist" to "Kri").

Treatment Integrity

The researcher developed a checklist consisting of the essential protocol steps (see appendix E). An independent observer using this checklist scored the researcher's correct or incorrect use of protocol across 26% of sessions. Percentage of steps correctly implemented was calculated from the checklist by dividing the number correct by the total number of protocol steps to provide a measure of treatment integrity. Treatment integrity was 98% (range = 92-100%).

RESULTS

A summary is shown in table 2 of each participant's intervention target and percentage of correct responding during baseline recall probes and two types of control probes. All participants exhibited difficulty in restating the names of people after a specified period of time during baseline recall measures. Harriet, Barry, and George's performance during hidden probes were commensurate with errors made when the array was re-arranged in view, providing evidence that positional influences were not present. Susan's poor performance during hidden probes (0%) provided evidence, in conjunction with her error analysis, that her performance during spaced retrieval was heavily influenced by the position of the target photograph.

Table 2

Intervention Targets and Participants' Performance on Baseline Recall and Spaced Retrieval (SR) Control Probes

			SR Control Probes	
Participant	Intervention Target	Baseline 60-s Recall	Hidden Probe	Target Change
Harriet	Kristen, Care Coordinator	0/6	69%	13%
Barry	Kate, Activity Director	0/6	67%	0%
Susan	Kate, Activity Director	3/6	0%	0%
George	Kate, Activity Director	1/6	100%	0%

Note. Baseline recall measured with non-targeted staff photographs. The fraction values refer to the number of correct recalls out of the total number of recall trials. Percentage values refer to the percentage of correct responding during control probes.

Finally, all participants failed to engage in correct responding during probe trials in which the target photograph changed. The one exception was with Harriet, where she correctly selected and named the new target on one occasion. This only occurred at a spaced retrieval delay value of 5-s, however, where at greater delays, her performance did not shift in accord with the new target.

Figure 1 visually depicts the staggered introduction of conditions across concurrent baselines for Barry and Harriet. Spaced retrieval was associated with an increasing trend in median recall delay reached for Barry before stabilizing, and low stable median delays for Harriet. Exemplar training within spaced retrieval was associated with an increasing trend of even greater degree for Barry before stabilizing again, and no change in trend, level, or variability for Harriet. The introduction of a textual prompt during spaced retrieval for Harriet resulted in an increasing trend.

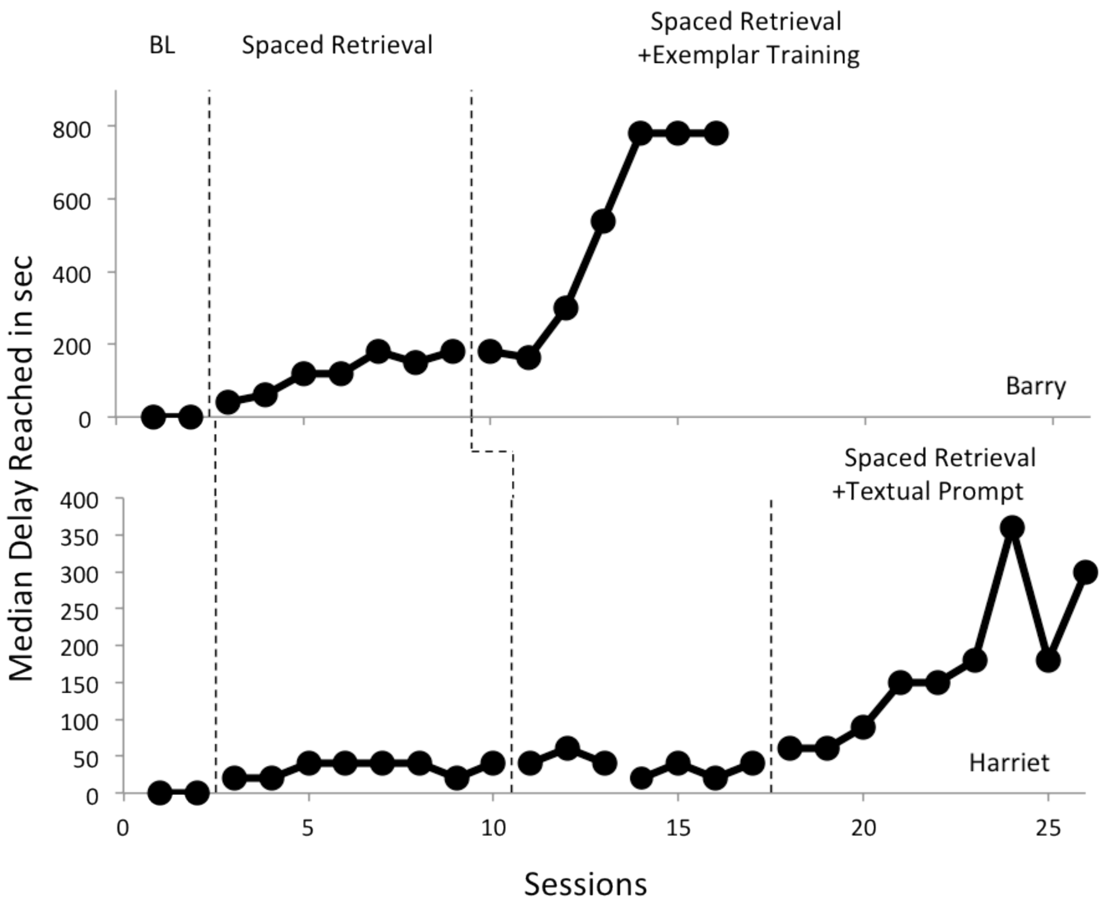

Figure 1. The staggered introduction of conditions in the concurrent multiple baseline design is visually demonstrated for Barry and Harriet.

Figure 2 visually depicts staggered conditions for Susan and George's concurrent baselines. Spaced retrieval resulted in little gains in delay values for Susan, and an increasing trend for George. Tailored interventions based upon stimulus control deficits as assessed by measures to be discussed below were then implemented. Susan's median recall delay performance exhibited an increasing trend before stabilizing, while George's median delay values experienced an additional increasing trend until stabilizing at the highest delay value.

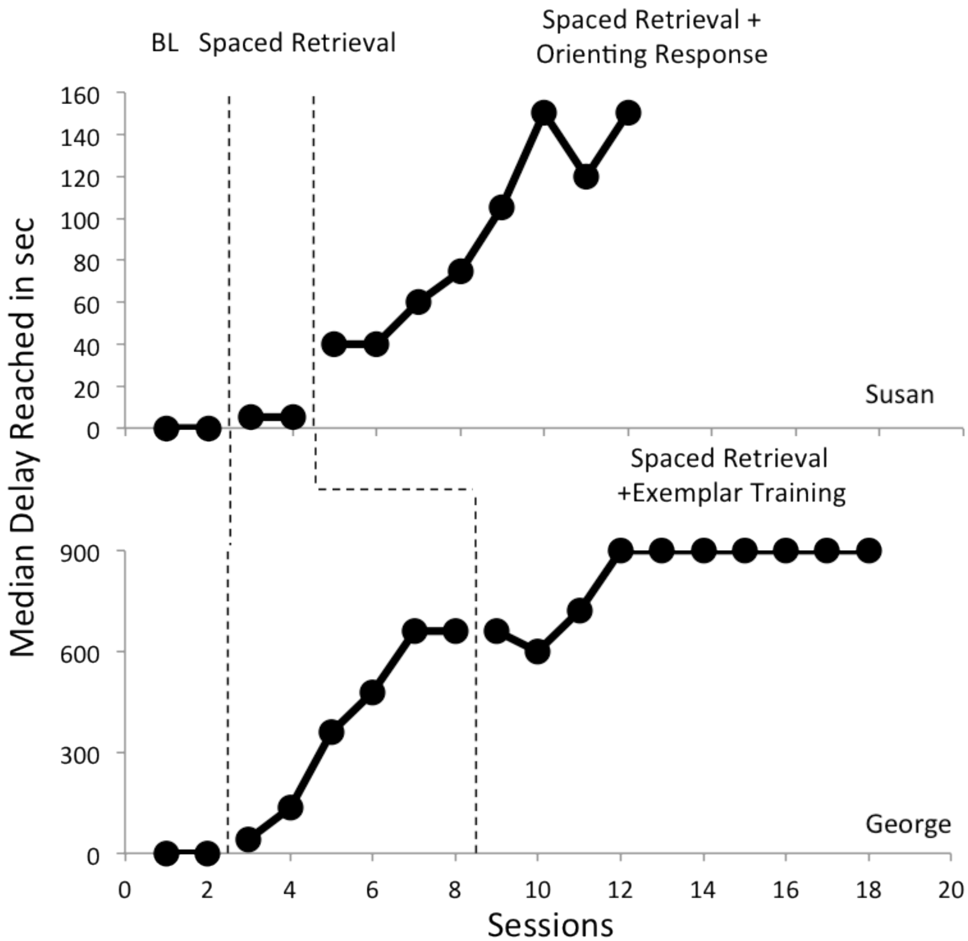

Figure 2. The staggered introduction of conditions in the concurrent multiple baseline design is visually demonstrated for Susan and George.

Because the sequence of assessments and interventions varied across subjects, depending on their response to earlier interventions, subsequent results are presented on a subject by subject basis to highlight the unique assessment and interventions that were used to evaluate effects on memory probes and to demonstrate experimental control for various individualized interventions.

Barry

Figure 3 shows Barry's performance across baseline, spaced retrieval, and spaced retrieval plus exemplar training sessions. The top panel represents the median delay

value reached in seconds on correct trials. Baseline was stable, with median delay values of zero seconds, indicating that his performance did not maintain at any delay value. When the spaced retrieval condition began, a level change was observed, as seen in the increase to a 40-s median delay reached during the first session of spaced retrieval. Across the spaced retrieval training condition, Barry's median recall delay value demonstrated an increasing trend over the following five sessions, from 40-s to 180-s, before stabilizing at the 180-s median recall delay value across the final three sessions of the condition. The introduction of serial exemplar training during spaced retrieval sessions resulted in an increasing trend after two sessions, from 165-s to 780-s. This observed trend occurred across three sessions before stabilizing at 780-s for the final two sessions of this condition.

The middle panel of Figure 3 shows the percentage of naming and selection errors across spaced retrieval and exemplar sessions. Selection errors dropped to zero levels after two sessions (M = 2%, range = 0-5%), while naming errors demonstrated an increasing trend over the entire condition (M = 19%, range = 0-40%). Selection errors remained at zero levels throughout serial exemplar training, while naming errors exhibited a level change from 40% to 17%, and then decreased to zero percent after four sessions in that condition.

The bottom panel of Figure 3 shows the cumulative number of correct selection and naming responses during cold probes conducted at the beginning of every session after an initial spaced retrieval session. No correct responding occurred throughout the entirety of the spaced retrieval condition. After three sessions of exemplar training, Barry engaged in a correct selection response. On session 15, Barry correctly engaged in both

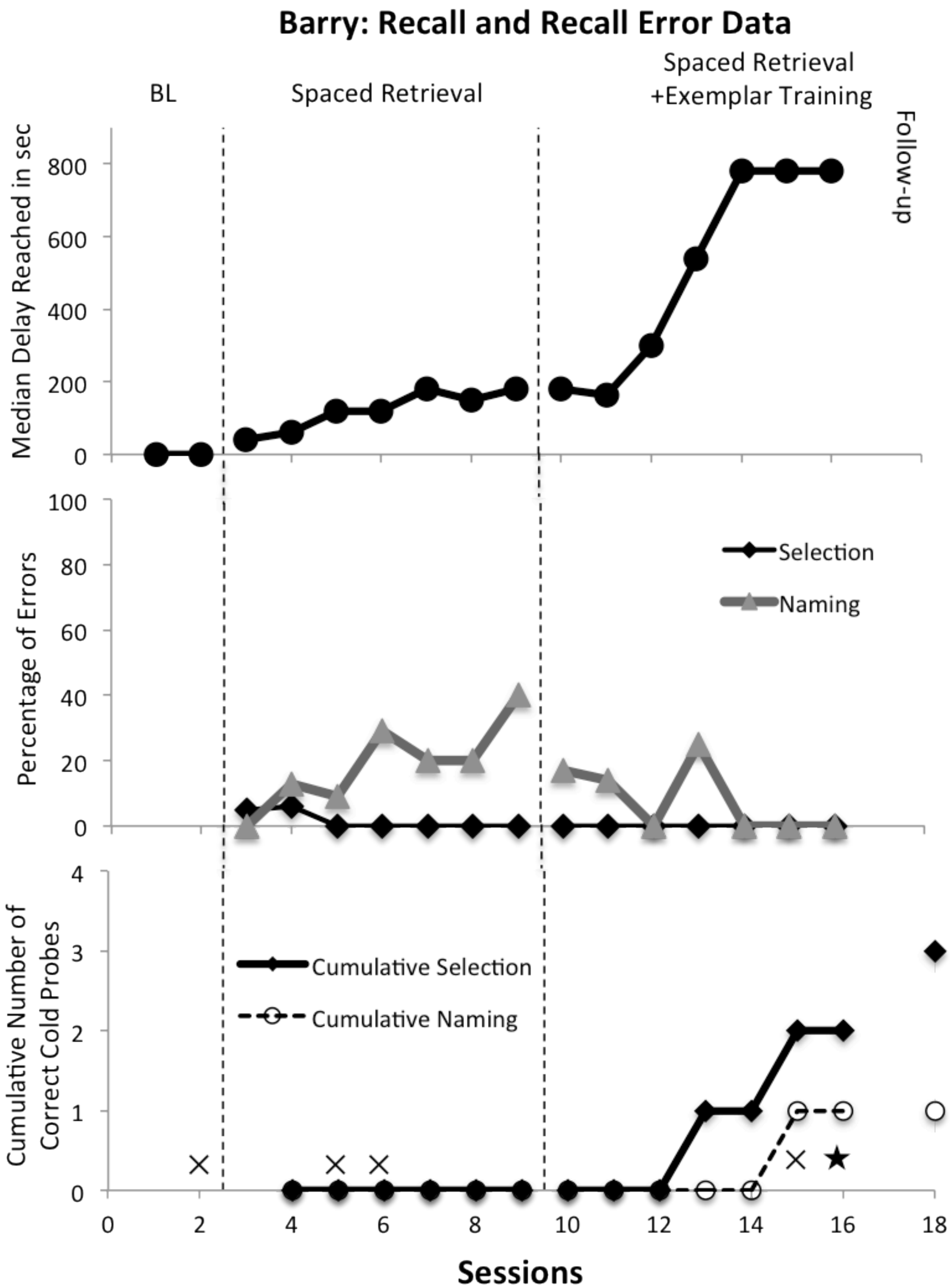

Figure 3. Barry's Recall Performance and Error Analysis. The top panel depicts median delay values reached with correct recall trials across sessions. The middle panel displays Barry's percentage of errors for selection and naming responses during spaced retrieval sessions. The bottom panel depicts Barry's cumulative number of correct selection and naming responses during cold probes administered at the start of each session. Also depicted in the bottom panel are failed and passed live-person probes, as indicated by X's and stars, respectively.

the selection and naming response. Correct responding did not occur, however, on the following and final session, but was observed across selection and naming responses at the one-week follow-up.

The bottom panel also depicts failed and passed live-person probes, as indicated by X's and stars, respectively. On the second session of baseline, Barry did not state the name of the target (i.e., he said "I don't know" when asked if he knew her name). He responded in a similar manner during probes conducted on sessions five and six. After five sessions of exemplar training, Barry again failed the live-person probe. On the following session, however, he passed.

Figure 4 depicts Barry's correct responses across example and non-example probe measures over time. The top panel shows correct responding during selection-based nonidentity matching-to-sample probes. During baseline, he correctly matched different pictures of both non-examples and examples of the target staff across 67% and 100% of probes. Responding on these probes remained stable throughout the remaining conditions.

The middle panel of Figure 4 depicts the percentage of correct naming responses during photo probes. Responding was stable at zero percent across examples and non-examples during baseline. During the spaced retrieval condition, a spike in responding with example photo probes was seen on session six, but then dropped back down to zero levels throughout the remainder of the condition (M = 13%, range = 0-67%). A large level change in correct naming example photo probes was observed during the first session of exemplar training, with an increase from 0% to 100% which remained at this level with the exception of a slight decrease in responding during session 15 before

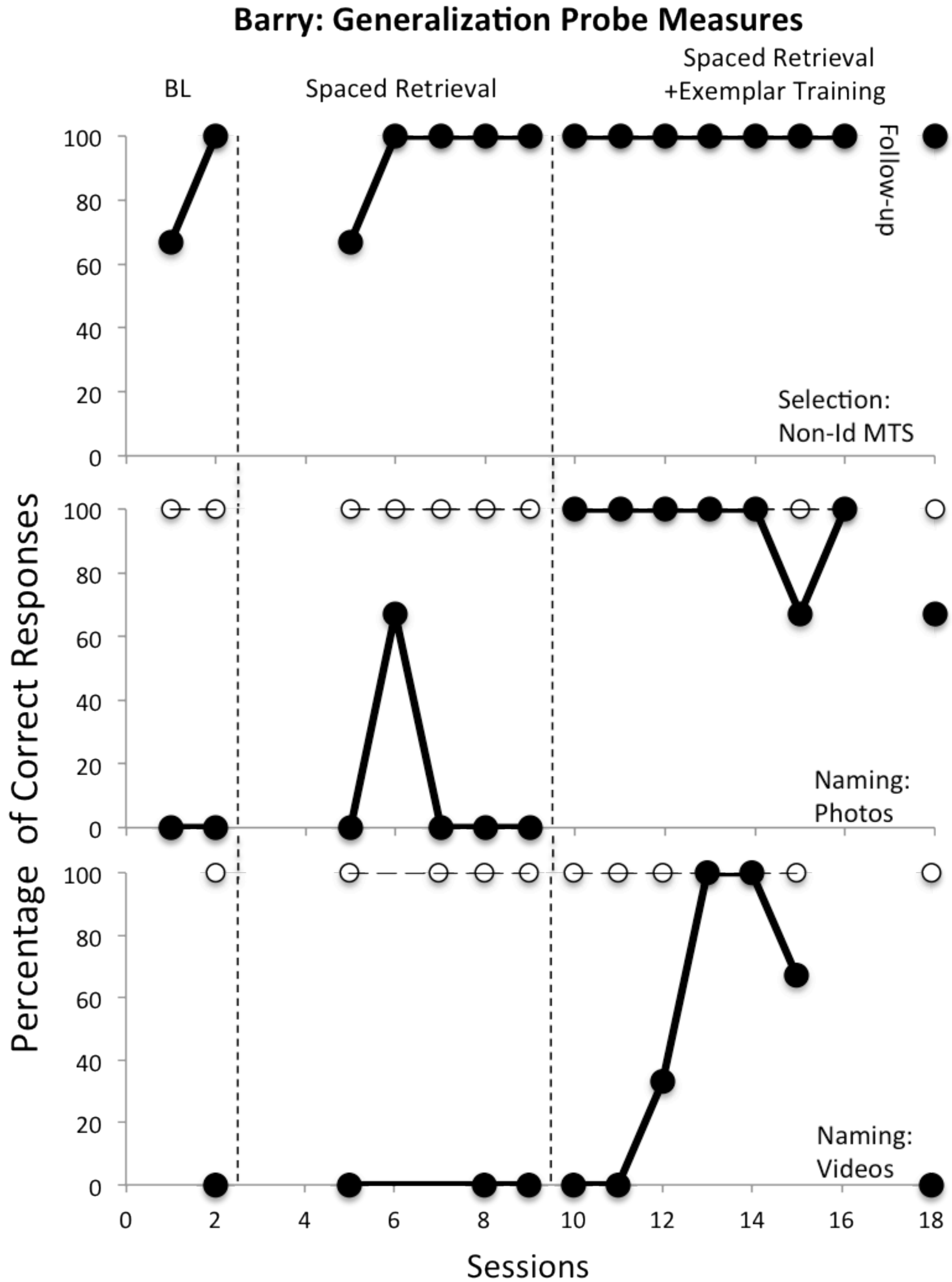

Figure 4. Barry’s performance across selection-based matching-to-sample of non-identity photographs (top panel), topography-based naming of photographs (middle panel), and topography-based naming of videos (bottom panel). Across all probe measures, example probes are displayed on the solid line and closed circle data paths, and non-example probes are displayed on the dashed line with open circles data paths.

returning to previous levels at 100% (M = 95%, range = 67-100%). Responding during non-example photo probes remained stable at 100% across all conditions. During the one-week follow-up, responding during photo probes with examples dropped to 67%.

The bottom panel of Figure 4 demonstrates the percentage of correct naming responses during video probes. Naming was stable at 100% across videos of non-examples during all conditions (i.e., he always stated "I don't know" when presented with videos of staff members not being targeted for intervention). Correct naming during example videos was at zero during baseline, and remained at this level throughout the spaced retrieval condition. Following two sessions of exemplar training, correct responding in the presence of example videos showed an increasing trend until stabilizing for two sessions at 100%. During the final video probe of examples, Barry's responding dropped to 67%. At the one-week follow-up, his naming responding dropped back to baseline levels.

George

Figure 5 depicts George's performance across baseline, spaced retrieval, and spaced retrieval plus exemplar training sessions. The top panel represents the median recall delay value reached in seconds on correct trials. George's baseline was steady, with median delay values of zero seconds, indicating that his performance did not maintain at any delay value. During the initiation of the spaced retrieval condition, a level change was observed, where the median delay value increased to 40-s. An increasing trend also occurred from the first session of spaced retrieval until the second to last session, at which point the median delay value for correct responding stabilized at 660-s. Median delay value remained stable at 660-s for the first session of exemplar

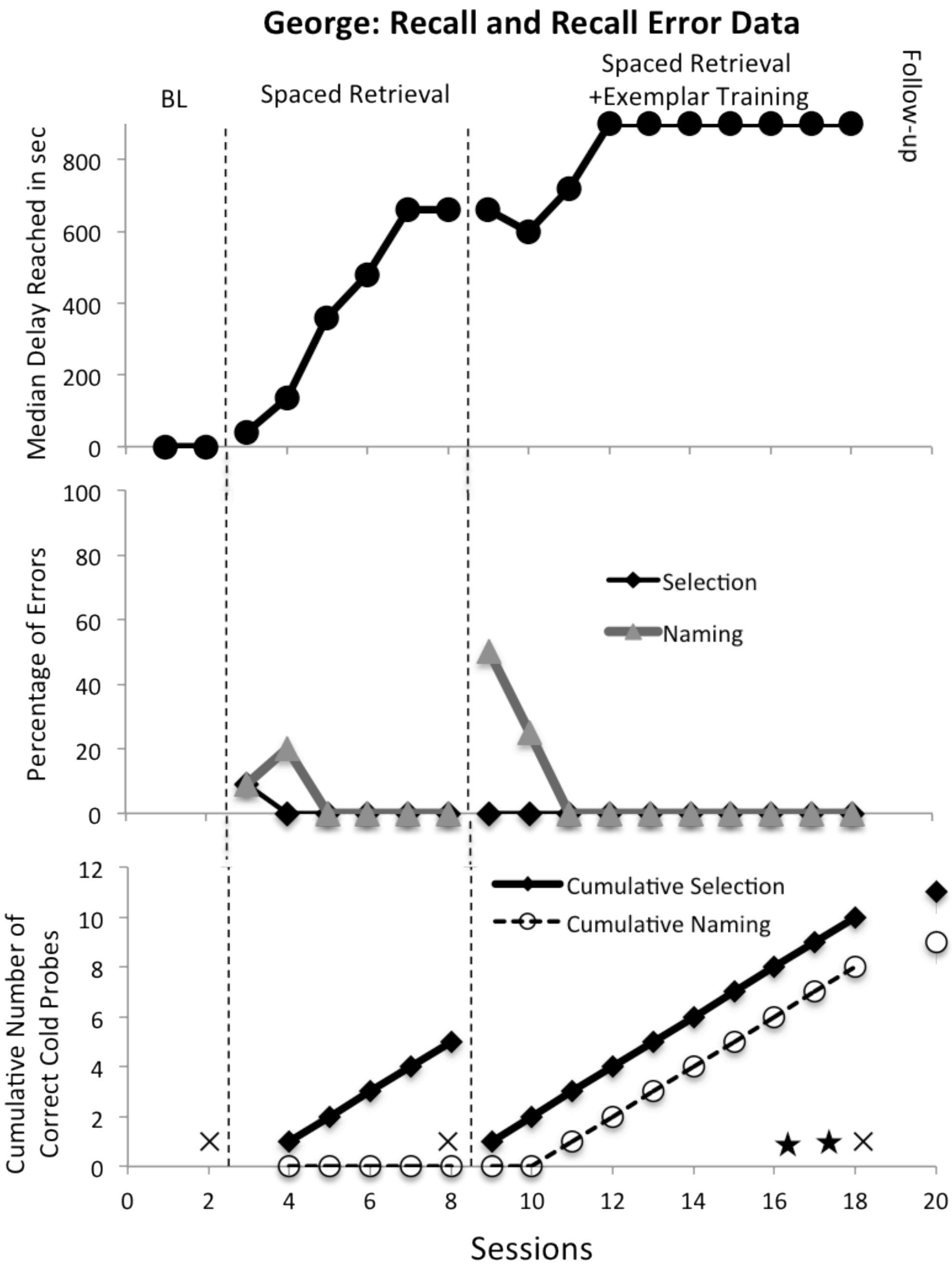

Figure 5. George's Recall Performance and Error Analysis. The top panel depicts median delay values reached with correct recall trials across sessions. The middle panel displays George's percentage of errors for selection and naming responses during spaced retrieval sessions. The bottom panel depicts George's cumulative number of correct selection and naming responses during cold probes administered at the start of each session. Also depicted in the bottom panel are failed and passed live-person probes, as indicated by X's and stars, respectively.

training before slightly dropping to 600-s on the second session of exemplar training. An increasing trend was observed following this slight drop in delay value until stabilizing at the maximum delay value of 900-s for the remainder of the condition.

The middle panel of Figure 5 shows the percentage of selection and naming errors George engaged in across conditions. After George engaged in selection errors during five percent of spaced retrieval trials within the first session, selection errors remained at zero levels across the rest of his participation. George engaged in naming errors at a slightly elevated percentage as compared to his selection errors during the first two sessions of spaced retrieval until stabilizing at zero levels for the rest of the condition. With the introduction of exemplar training to spaced retrieval, George's naming errors increased to 50% before dropping to 25%, and then staying constant at 0%.

The bottom panel of Figure 5 visually demonstrates the cumulative number of correct selection and naming responses during cold probes conducted at the beginning of every session after an initial spaced retrieval session. George engaged in correct selection responses at a steady rate for the entirety of the spaced retrieval condition. No correct naming responses occurred during this condition. A steady rate of selection responses continued to occur during the exemplar training condition. After two sessions of exemplar training, correct naming responses began occurring at a steady rate that persisted throughout the remainder of the condition. During the one-week follow up, this performance maintained.

The bottom panel of figure 5 also shows failed and passed live-person probes, as indicated by X's and stars, respectively. During a live-person probe during baseline and at the end of the spaced retrieval condition, George failed to state the name of the target.

Correct naming during live-person probes occurred within the exemplar training condition at sessions 16 and 17. Both correct instances of naming were unprompted (i.e., when the target greeted George, he said "How's Kate doing?"). During a live-person probe on the very next session, however, George failed to name the target.

Figure 6 visually depicts George's correct responding across example and non-example probe measures over time. The top panel shows correct responding during selection-based nonidentity matching-to-sample probes. George's baseline pattern of matching across examples and non-examples was stable at 100%. The percentage of correct matching with examples of the target continued to be stable at 100% across the spaced retrieval condition. Non-example matching became more variable during spaced retrieval, with a mean of 67% and range of 33-100%. During exemplar training, matching responses for both examples and non-examples exhibited some variability, with means of 90% (range = 67-100%) and 73% (range = 33-100%) respectively. At the one-week follow up, matching performance was commensurate with prior levels.

The middle panel of Figure 6 depicts George's percentage of correct naming responses during photo probes. At baseline, George engaged in zero percent correct naming during example photo probes, and 100% correct naming during non-example photo probes (i.e., he consistently stated "I don't know" across various photographs of the target staff member, as well as saying "I don't know" in the presence of photographs of non-targeted staff. During spaced retrieval, correct naming during example photo probes was highly variable, alternating between zero percent and 67%, then 100% before alternating once more between zero and 67%. George's correct naming during non-example presentations was stable at 100% until dropping to 67% at the final probe

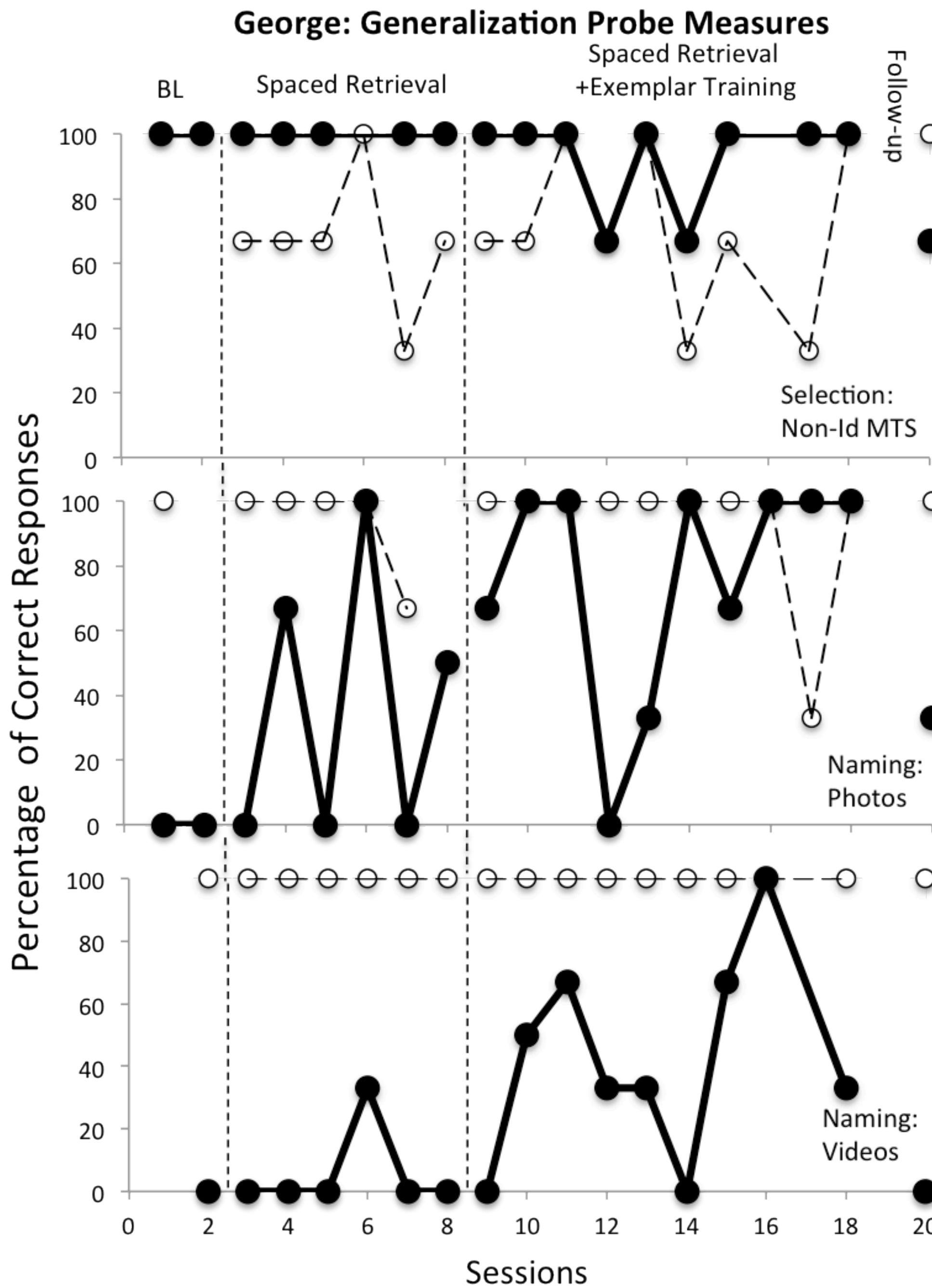

Figure 6. George's performance across selection-based matching-to-sample of non-identity photographs (top panel), topography-based naming of photographs (middle panel), and topography-based naming of videos (bottom panel). Across all probe measures, example probes are displayed on the solid line and closed circle data paths, and non-example probes are displayed on the dashed line with open circles data paths.

measure in that condition. Mean naming for examples and non-example photos was 36% (range 0-100%) and 93% (range = 67-100%), respectively. During exemplar training, correct naming during example photo probes continued to be variable, but less so (M = 76%, range = 0-100%), with naming stabilizing across the final three probes at 100%. George's correct naming in the presence of non-example photos during exemplar training was stable at 100%, except for session 17, in which he responded with the target staff member's name (i.e., George said "Kate" when presented with pictures of other staff members). Evidence of rote responding did not persist into the following session, however, as correct naming across non-example photos increased to 100%. At the one-week follow-up, naming across non-examples maintained at 100%, whereas naming across example photos dropped to 33%.

Finally, the bottom panel of Figure 6 shows George's percentage of correct naming responses during video probes. At baseline, George demonstrated stable responding across examples (0%) and non-example (100%) video probes. Across the spaced retrieval condition, George's naming during non-example video probes remained stable at 100%. George's naming in the presence of example video probes remained at zero percent, except for a transient increase in correct responding in the middle of the spaced retrieval condition (M = 6%, range 0-33%). When exemplar training was introduced, George's responding during non-example video probes continued at 100%. His naming during example video probes demonstrated a marked increase in variability, with only one probe at 100% correct responding (M = 43%, range = 0-100%). At the one-week-follow up, naming performance returned to baseline levels.

Susan

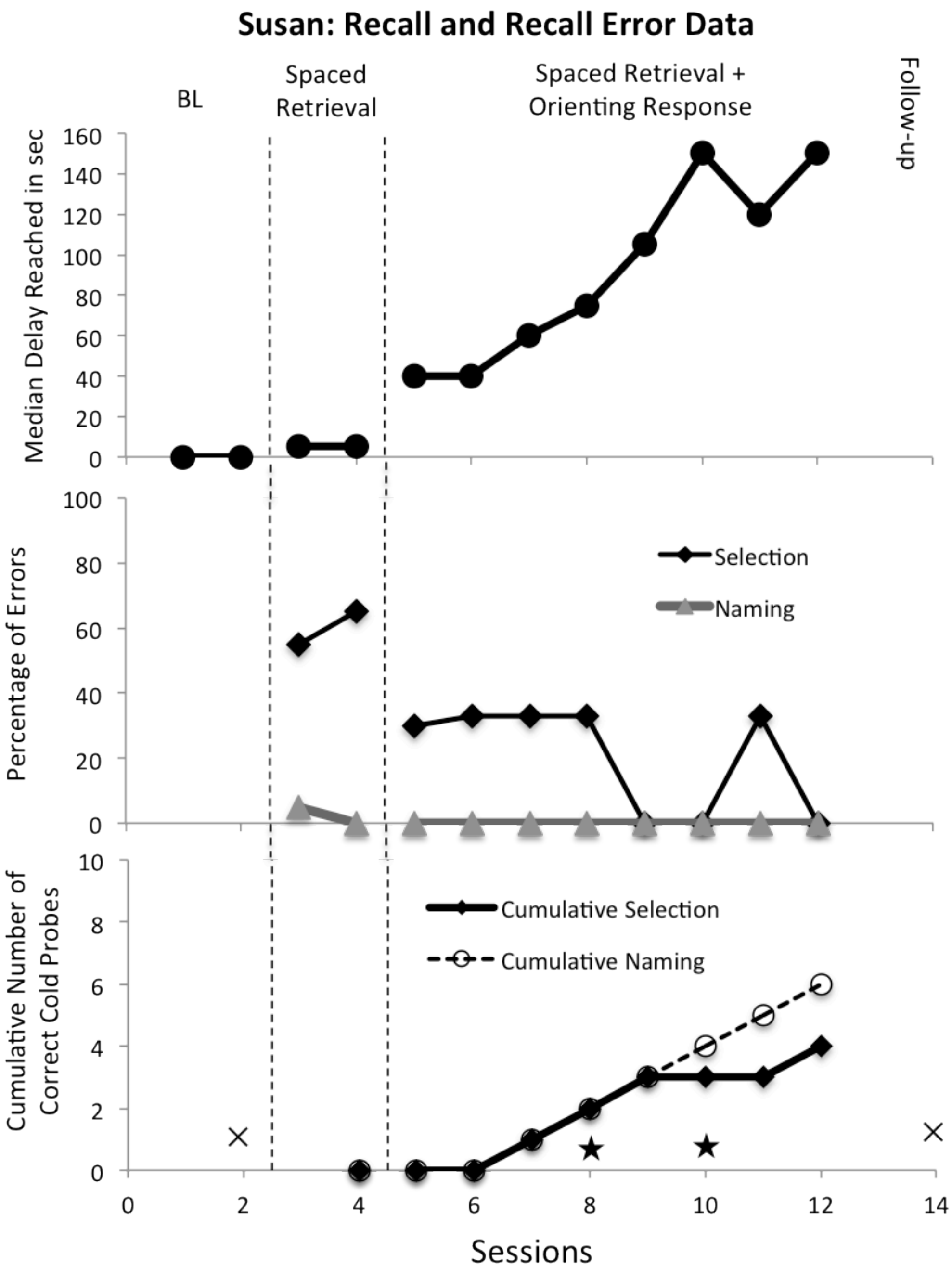

Figure 7. Susan's Recall Performance and Error Analysis. The top panel depicts median delay values reached with correct recall trials across sessions. The middle panel displays Susan's percentage of errors for selection and naming responses during spaced retrieval sessions. The bottom panel depicts Susan's cumulative number of correct selection and naming responses during cold probes administered at the start of each session. Also depicted in the bottom panel are failed and passed live-person probes, as indicated by X's and stars, respectively.

Susan's performance across conditions is depicted in Figure 7. The top panel represents the median delay value reached in seconds on correct trials. In baseline, Susan demonstrated stable low performance as indicated by the zero median delay values across two sessions. Upon the introduction of spaced retrieval, median delay values increased to 5-s but persisted at this level. When the orienting response was introduced during spaced retrieval sessions, a level effect was observed, in which the median delay value increases to 40-s for two sessions before demonstrating an increasing trend. After four sessions with this trend, the median value remained stable between 120-s and 150-s.

The middle panel depicts Susan's errors across spaced retrieval sessions. The only naming errors occurred during the first spaced retrieval session (across 5% of trials), and then remained at zero levels throughout the remainder of her participation. Selection errors were high during the spaced retrieval condition, with a mean of 60% and a range of 55-65%. Once the orienting response requirement was introduced, a level change occurred with selection errors, dropping from 65% to 30%. Selection errors were stable until dropping to zero levels after four sessions in to this condition. Mean selection errors were 20% (range = 0-33%).

The bottom panel of Figure 7 visually demonstrates the cumulative number of correct selection and naming responses during cold probes conducted at the beginning of every session after an initial spaced retrieval session. Susan did not engage in correct selection or naming behavior during such probes until the beginning of the third session of the spaced retrieval plus orienting response condition. Selection and naming rates remained steady for an additional two sessions, at which point she only engaged in

correct naming. Following two sessions of this pattern, Susan engaged in correct selection and naming responses.

The bottom panel of figure 7 also shows failed and passed live-person probes, as indicated by X's and stars, respectively. During the second session of baseline, Susan did not engage in the correct naming response in the presence of the target staff member. During the next two live-person probes, which occurred during the spaced retrieval plus orienting response condition, she passed both times on sessions eight and ten. During a one-week follow-up, the naming response failed to maintain.

Susan's performance on generalization probe measures is depicted in Figure 8. The top panel shows correct responding during selection-based nonidentity matching-to-sample probes. Susan's correct matching across example and non-example photos were low, with means of 25% and 0%, respectively, during baseline. With the introduction of spaced retrieval, matching improved with a mean of 50% across examples and 25% across non-examples. During the spaced retrieval plus orienting response condition, correct matching was at low and variable levels, with matching across examples and non-examples never correctly occurring above 50%. Mean matching with example photos was 19% (range = 0-50%), while mean matching with non-example photos was 11% (range = 0-33%) during this condition.

The middle panel depicts Susan's correct naming responses across example and non-example photo probes. Naming was stable in baseline, with 0% correct with example photos and 100% correct for non-example photos. During spaced retrieval, correct naming remained at zero levels for example photos. Correct naming with non-examples dropped from baseline to a mean of 42%. On such probes during spaced

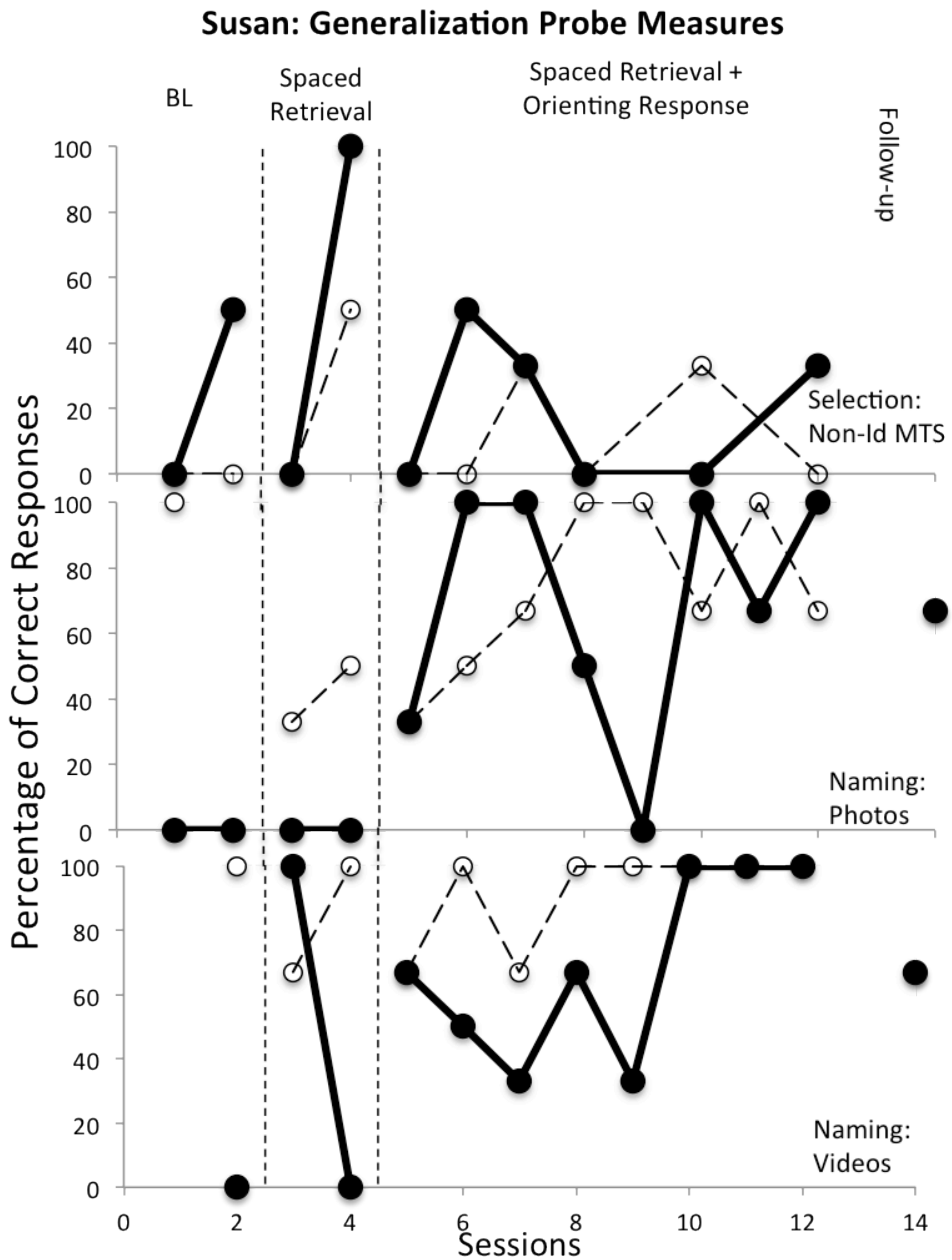

Figure 8. Susan's performance across selection-based matching-to-sample of non-identity photographs (top panel), topography-based naming of photographs (middle panel), and topography-based naming of videos (bottom panel). Across all probe measures, example probes are displayed on the solid line and closed circle data paths, and non-example probes are displayed on the dashed line with open circles data paths.

retrieval, Susan exhibited a tendency to name photos of non-targeted staff as Kate, the target staff member. With the introduction of the orienting response to spaced retrieval,

Susan's correct naming responses increased rapidly over time, going from zero levels to 100% in two sessions within this condition. Naming was variable during example photo probes throughout succeeding sessions (M = 69%, range = 0-100%). Correct naming during non-example photo probes took a similar pattern of variability, but with an overall slight increasing trend (M = 73%, range = 33-100%). During the one-week follow up, Susan's naming maintained at 67% across example and non-example photo probes. This level of responding was lower as compared to the previous session, but still higher than at baseline and spaced retrieval conditions.

The bottom panel of Figure 8 provides a visual representation of Susan's naming responses across example and non-example video probes. Baseline responding for example video probes was 0%, and 100% for non-examples. During spaced retrieval, naming became variable with both types of video probes. With the introduction of spaced retrieval, Susan correctly named example videos during 100% of probes and correctly naming 67% of non-example video probes. On the following session, naming during example video probes dropped to 0%, while responding to non-example probes improved to 100%. Requiring an orienting response during spaced retrieval trials was associated with moderately variable correct naming with example video probes until the sixth session where correct naming stabilized at 100% for the final three sessions. Mean correct naming with example video probes was 69% (range = 33-100%). Susan's correct naming in the presence of non-example video probes remained fairly stable, alternating between 67% and 100% before staying constant at 100% for the final five sessions (M = 92%). At Susan's one-week follow-up, correct naming occurred during 67% of both

example and non-example video probes, revealing a moderate decline in performance as compared to observed trends from the previous condition.

Harriet

Harriet's performance across baseline, spaced retrieval, and spaced retrieval plus exemplar training sessions is depicted in Figure 9. The top panel displays the median delay value reached in seconds on correct trials. Harriet's baseline delay value was at zero levels, indicating that responding did not occur at any attempted delay. Harriet's median delay value demonstrated a level change to 20-s when the spaced retrieval condition began, and remained steady between 20-s and 40-s median delay values. When exemplar training was introduced, no detectable change in trend, level, or variability occurred. After three sessions, Harriet met criterion for another photo exemplar to be targeted. This produced no detectable effects in terms of the median delay measure. The addition of a textual prompt produced a change in trend across seven sessions during this condition, from 60-s to 360-s. Harriet's delay values became variable for the final two sessions.

The middle panel of Figure 9 shows Harriet's errors across spaced retrieval sessions. Harriet engaged in selection errors at a very low amount, and only across the first two sessions of spaced retrieval. Naming errors were variable during this condition, ranging between 7 and 57%. Mean selection and naming errors were 0.6% and 23% respectively. During exemplar training, Harriet's selection errors continued to remain at low levels (M=2%), with the only selection errors occurring during the first session in this condition. Harriet's naming errors had a slight increasing trend throughout the spaced retrieval condition (M = 36%, range = 15-50%). Finally, the textual prompt

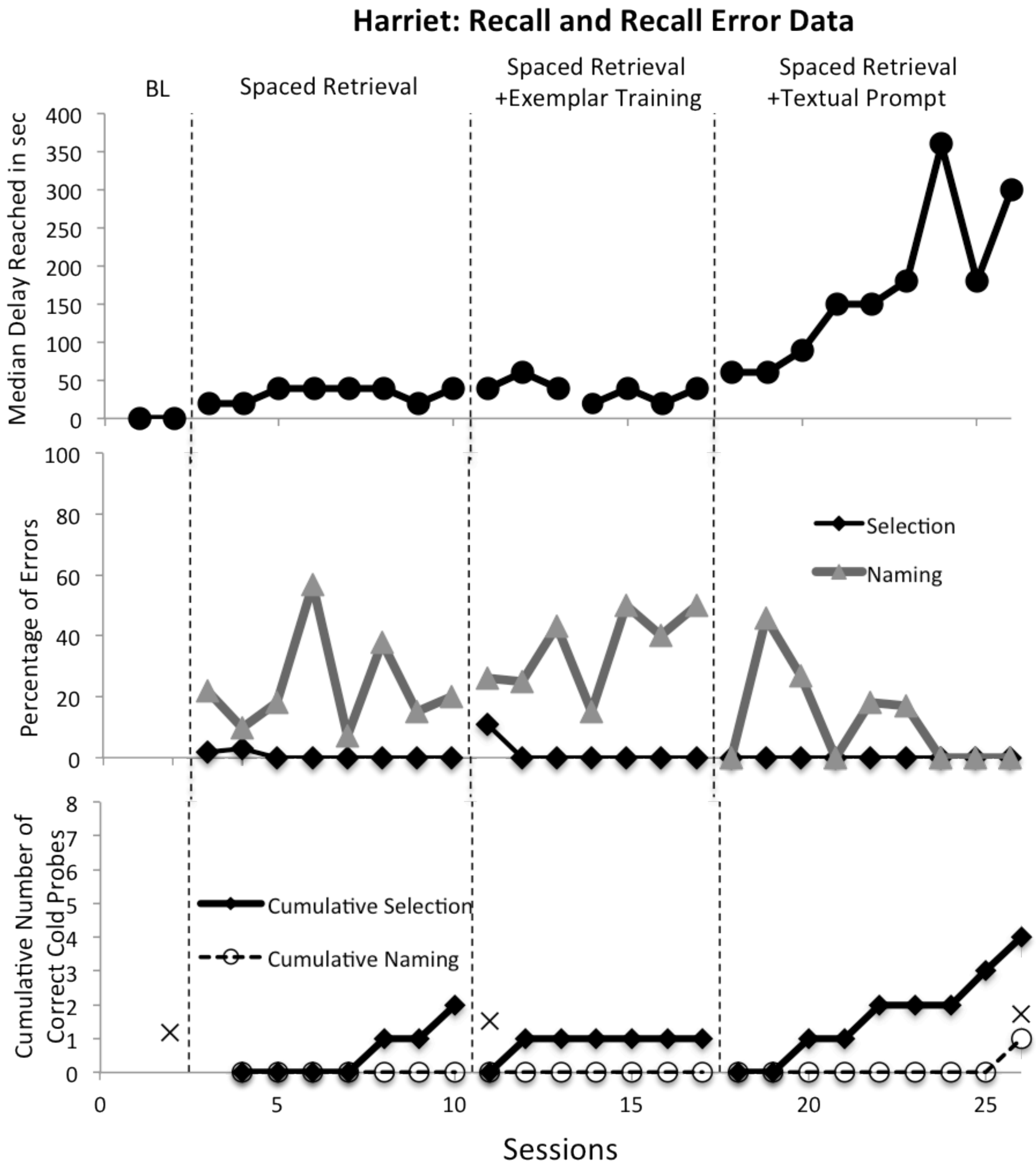

Figure 9. Harriet's Recall Performance and Error Analysis. The top panel depicts median delay values reached with correct recall trials across sessions. The middle panel displays Harriet's percentage of errors for selection and naming responses during spaced retrieval sessions. The bottom panel depicts Harriet's cumulative number of correct selection and naming responses during cold probes administered at the start of each session. Also depicted in the bottom panel are failed and passed live-person probes, as indicated by X's and stars, respectively.

addition to spaced retrieval resulted in no selection errors, and produced a decreasing trend over five sessions with naming errors until reaching zero levels for the final three sessions (M=12%, range = 0-46%).

The bottom panel of Figure 9 depicts the cumulative number of correct selection and naming responses during cold probes conducted at the beginning of every session after an initial spaced retrieval session. During the spaced retrieval condition, Harriet engaged in correct selection responses on sessions eight and ten. During the following condition with exemplar training implemented, only one correct selection response occurred. Finally, with the addition of a textual prompt, Harriet's correct selection response rate increased, with a total number of four correct selections during this condition. Harriet did not engage in correct naming during cold probes until the last session.

Also depicted in the bottom panel of Figure 9 are failed and passed live-person probes, as indicated by X's and stars, respectively. Harriet did not engage in correct naming during any three live-person probes taken across baseline, exemplar training, or spaced retrieval plus a textual prompt.

Figure 10 is a graphical representation of Harriet's performance across generalization probe measures. The top panel shows correct responding during selection-based nonidentity matching-to-sample probes. Harriet's baseline matching performance was stable in the presence of example photos, while slightly more variable during non-example selection probes. Means were 100% and 84% (range = 67-100%), respectively. During spaced retrieval, matching responses across example and non-example photo probes was steady with means of 84% and 92%. The introduction of exemplar training

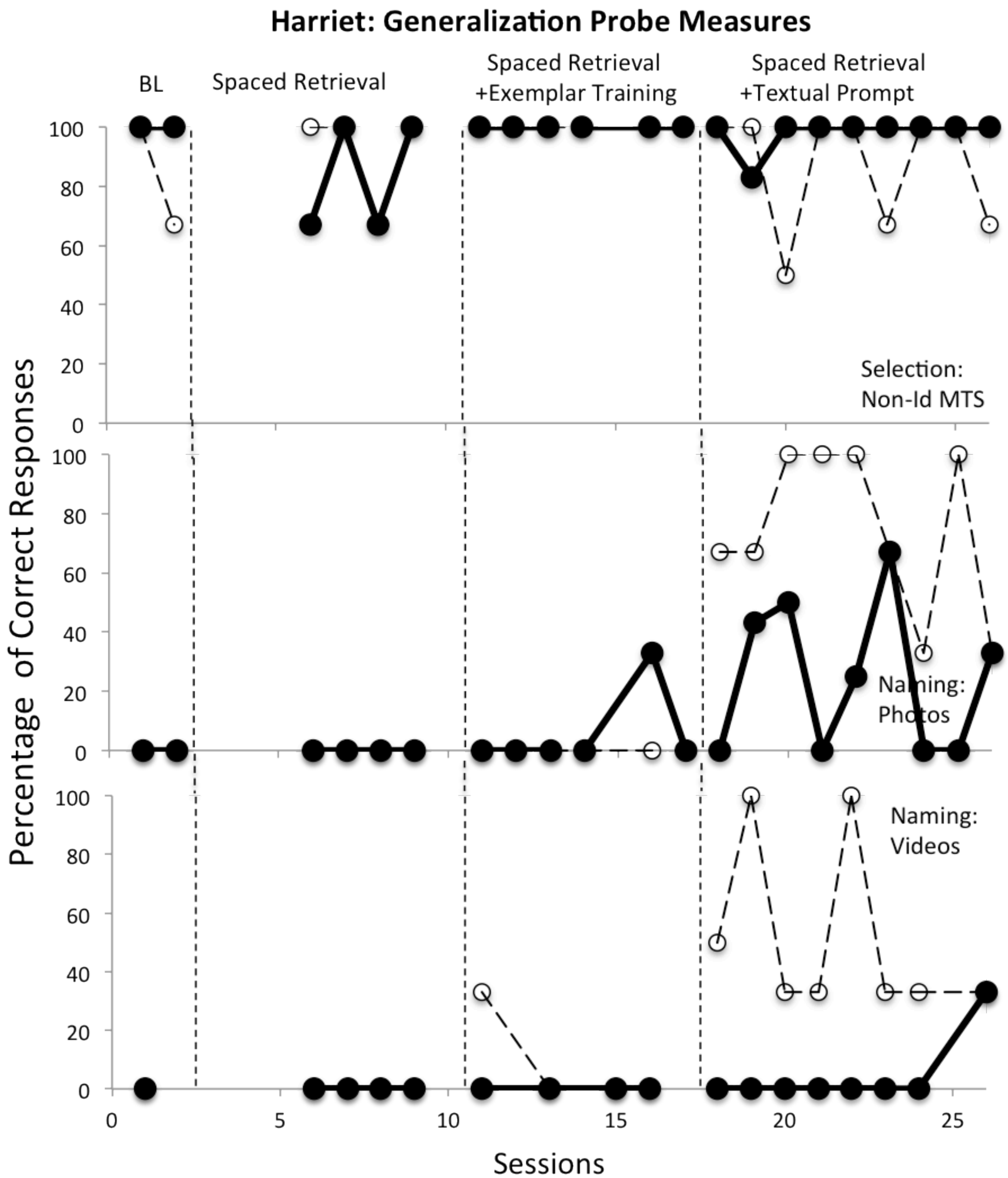

Figure 10. Harriet's performance across selection-based matching-to-sample of non-identity photographs (top panel), topography-based naming of photographs (middle panel), and topography-based naming of videos (bottom panel). Across all probe measures, example probes are displayed on the solid line and closed circle data paths, and non-example probes are displayed on the dashed line with open circles data paths.

was associated with 100% matching performance in the presence of example and non-examples. Finally, Harriet's matching with example photos remained stable, while

matching non-example photos became more variable (M = 87%, range = 50-100%) during the textual prompt condition.

The middle panel of Figure 10 depicts Harriet's correct naming responses across example and non-example photo probes. During baseline and spaced retrieval conditions, Harriet engaged in zero correct naming with both example and non-example photo probes (i.e., she provided incorrect names for every photo presented). This pattern continued for non-examples throughout the exemplar training condition. A spike in correct naming was observed for example photo probes on session 16, where Harriet correctly named one out of three example probes. Harriet's correct naming across examples and non-examples increased with the addition of a textual prompt to spaced retrieval trials, and were also variable. Mean correct naming for example photo probes was 24% (range = 0-67%), while mean correct naming for non-example photo probes was higher at 74% (range = 33-100%).

The bottom panel of Figure 10 demonstrates Harriet's naming responses across example and non-example video probes. Harriet's performance took a similar pattern as it did with naming photo probes under baseline and spaced retrieval conditions, with zero levels of correct naming across example and non-example video probes. During the first exemplar training condition session, a transient effect was observed on Harriet's naming of non-example videos, where she correctly named one out of three videos of non-targeted staff members. The spaced retrieval plus textual prompt condition produced a change in level and variability for correct naming with non-example video probes (M=52%, range = 33-100%). Harriet only engaged in one correct naming response in the presence of example video probes on the final session during this condition.

DISCUSSION

The present study sought to evaluate a spaced retrieval intervention on name-face relations with older adults who exhibited difficulty in naming and recalling the names of caregivers. Four participants were included that demonstrated cognitive impairment as evidenced by their scores on cognitive screeners, including the MMSE and CDT, along with behavioral measures taken during baseline on their naming and recalling repertoires. Multiple probe measures were used to examine generalization effects throughout the length of their participation. Moreover, the present study investigated the effects of modified procedures within the spaced retrieval intervention that were matched to stimulus control deficits exhibited by participants. Following a spaced retrieval condition, two participants were exposed to serial exemplar training. One participant was prompted to engage in an orienting response at the start of spaced retrieval trials, and one participant was exposed to a textual prompt embedded within spaced retrieval trials. One-week follow up probe measures were obtained across all four participants to determine the extent to which intervention gains had maintained.

Measures of Recall

Taken together, these data suggest that spaced retrieval alone did not produce transfer of the naming response to the actual sight of the staff member despite within-session gains of delayed recall. Further, although within-session gains were observed for three out of four participants, performance was not significantly altered between-sessions as measured by cold probes.

When tailored interventions were administered, improved outcomes occurred across recall measures and generalization probes to at least some extent for all

participants. These data suggest that tailoring the intervention based on participants' responding allows for greater outcomes to be reached. Overall, the implementation of conditions that addressed each participant's stimulus control deficits was associated with additional increases in median delay value and performance during cold probes, as well as correct responding during live-person probes.

Comparisons to Previous SR Studies Targeting Name-Face Relations

Hawley and Cherry's (2004) protocol involved six sessions of spaced retrieval training. By the end of training, the six participants reached a range of recall delay values between 60 and 360-s. Also by the end of training, three out of six participants passed the live-person probe after the target person had entered the room and the researcher directly asked if they knew the person's name. Hawley, Cherry, Boudreaux, and Jackson (2008) conducted nine sessions of spaced retrieval training, which resulted in four out of six participants exposed to the titrating spaced retrieval delay values passing the live-person probe by either spontaneously stating their name upon entering the room (n = 1), or after being directly asked (n = 3). Obtained maximum recall delays for participants where not reported in this study. Nine sessions of spaced retrieval training were also used in Cherry, Walvoord, and Hawley (2010) with four participants. None of the participants correctly named the target person during live-person probes either spontaneously or prompted in that study.

The current study also saw increases in delayed recall during spaced retrieval sessions. After seven sessions of spaces retrieval, Barry reached a maximum delay of 300-s. Following six sessions with George, his maximum recall delay value was 780-s. After Harriet was exposed to 10 sessions of spaced retrieval, she demonstrated successful

delayed recall at 40-s. Susan only received two sessions of spaced retrieval, with a maximum delay reached of 10-s, before the intervention was further tailored due to her stimulus control deficits. While the current study provides a partial replication of prior research on space retrieval, the effects appear to be limited to a relatively specific set of assessment conditions (i.e., delay recall values reached), a limitation that has implications for the clinical utility of the spaced retrieval procedure. For example, none of the participants in the current study passed a live-person probe during the spaced retrieval condition, which provides similar results to Cherry, Walvoord, and Hawley's (2010) report, but is in contrast to Hawley and Cherry (2004) and Hawley, Cherry, Boudreaux, and Jackson's (2008) reports. Based on previous results, it could be reasonably expected that participants in the current study exposed to at least six sessions had a higher probability of passing a live-person probe. Number of spaced retrieval conditions did vary in the current study, from two to eight, but none demonstrated successful transfer to the live-person probe during that time.

These discrepancies call attention to the importance of making data-based decisions on when to implement condition changes during an intervention. The studies just mentioned all report using a predetermined spaced retrieval session sequence that terminates after a set endpoint (i.e., after six or nine sessions) regardless of how participants respond during those sessions. The current study, on the other hand, changed conditions based upon each participant meeting a performance criterion. It could be the case that the participants in the previous studies would have reached greater recall delay values if the protocol did not adhere to a predetermined condition sequence. This could be more likely if recall delay values were in actuality showing an increasing trend, which

highlights the added importance of displaying results visually to allow for detection of these changes in data over time.

Error Measures

Previous researchers have described spaced retrieval as an errorless procedure (e.g., Camp, Foss, O'Hanlon, & Stevens, 1996). The middle panels of figures 3, 5, 7, and 9, however, indicate that various errors were made during spaced retrieval, and some to a fairly high extent (e.g., Harriet and Barry). For the procedure to be errorless, errors should be nearly impossible to commit by supporting correct performance through strategies such as a zero second prompt delay, or increasing the salience of the target photograph from the distractors by adding additional visual cues that are systematically faded over a period of time based on a performance criterion.

Only after the introduction of spaced retrieval modifications were concomitant decreases in errors observed. Nevertheless, errors continued to occur during all conditions, as the modified procedures still did not include true "errorless" strategies. This is important to consider, in that intervention descriptions should accurately reflect the procedures they entail. Ensuring accurate descriptions allows for improved replication of findings, reduced conceptual confusion, and allows for better interpretations of the mechanism of action responsible for behavior change. For example, a differential reinforcement of alternative behavior (DRA) intervention specifies both the procedure (i.e., a reinforcer that was previously maintaining challenging behavior is now presented contingent upon an adaptive, alternative response, and is withheld for the challenging behavior) and the mechanism of action (i.e., the increase in this alternative behavior is a result of reinforcement, and the challenging behavior decreases as a result

of being placed on extinction). Such precision in describing interventions allows the procedure to be technological and conceptually rigorous (see Baer, Wolf, & Risley, 1968, for the importance of these and other scientific dimensions). Spaced retrieval should therefore be procedurally coherent to aid in the scientific endeavors of further evaluation and dissemination.

Generalization Probes

At baseline, participants demonstrated the ability to match pictures of various caregivers to pictures of those same caregivers, but at different camera angles and wearing different attire. This suggests that errors made during spaced retrieval and across generalization probes were not due to poor visual acuity. The exception to this was observed with Susan during such non-identity matching-to-sample tasks. Her non-identity matching was variable throughout the course of her participation. .

To better rule out visual impairments, a session of identity matching-to-sample was conducted during baseline. This entailed having Susan match a sample photograph of a staff member to the exact same photograph of that staff member in the comparison array. Under these conditions, Susan's matching performance improved to 80%. Thus, she exhibited difficulty in matching the "same" face at different variations, but performed well when asked to identify another exact configuration of stimulus properties that made up the staff member's face. This discrepancy in responding across identity and non-identity matching-to-sample, along with her high rates of selection errors relative to her naming errors, suggests the behavioral symptoms of prosopagnosia.

Prosopagnosia is a type of agnosia that describes poor facial recognition not due to sensory deficits. Susan's behavior during identity matching tasks may have been

guided by a specific constellation of stimuli that she would have been unable otherwise to respond to as a whole (i.e., two identical pictures of a person's face). In other words, she could indicate that they were the same faces based upon all the stimulus properties being identical. With non-identity matching, many, but not all, of the stimuli that comprise the two faces of the same person were the same (i.e., the faces are at different angles, producing shadows that highlight different features of the face), resulting in poor performance. Additional evidence to suggest prosopagnosia is found in Susan's variable pattern of correct responses in the presence of both S+ and S- probes, where she would at times state the target person's name during non-examples, and state that she did not know who the person was during example probes.

During the spaced retrieval condition, photograph-naming probes revealed no transfer with Harriet, Barry, and Susan, and highly variable levels of transfer with George to untrained photographs of the target staff member. Video-naming probes during this same condition also revealed no transfer of responding across participants. In general, this suggests that spaced retrieval alone did not produce generalization.

The one transient exception to this pattern of poor generalization during spaced retrieval can be found during Susan's first session of spaced retrieval, where she engaged in 100% correct responding during S+ video probes. This performance dropped back down to 0% during the next session. As noted above, Susan displayed difficulty in identifying different pictures of the same person. During the first video presentation of the S+, the target talked about upcoming activities. Given that she was the activities coordinator, this may have provided for a context that evoked Susan's naming response

due to a specific behavioral history with the target staff member, which then remained in strength across the other S+ video probes and S- video probe.

With the introduction of exemplar training, transfer of the naming response was seen across both photograph and video probes for Barry, and video probes for George. Barry's pattern of responding across generalization probes is of particular interest, as naming transferred to 100% of photographs during the first session of exemplar training, while evidence of transfer across video probes did not occur until the third session in. These outcomes may have occurred because although untrained photographs were used, they still shared many stimulus properties with the training stimulus used during spaced retrieval. The delay in transfer across video probes may have been due to the greater discrepancy in shared stimulus properties between the training photograph and untrained video probes (i.e., it was of the same person, but presented in a different modality). Both Barry and George passed live-person probed during this condition as well. This provides some evidence that exemplar training in spaced retrieval can produce generalization to untrained members of a stimulus class.

Susan's behavior demonstrated evidence of transfer when a prompt to engage in an orienting response was introduced to her spaced retrieval training. Anecdotally, Susan was observed to repeat an orienting response prompt following the sound of the cue to initiate a spaced retrieval trial. That is, she stated the attire of the target person while she scanned the array of photographs. This behavior also occurred during the live-person probe. Specifically, she stated "Woman with the white blouse and black sweater. Must be Kate!" when the target entered the room. The target was not wearing the same attire as in the training photograph, suggesting that it was not the attire alone that caused

correct naming to transfer. This instead suggests that verbal mediation played a large role in the transfer across generalization probes, where an intraverbal relation was established between the target person's name and the target's attire. In other words, Susan stated the target's attire, and hearing herself say this potentially functioned as a self-prompt that strengthened the target's name.

Harriet did not demonstrate transfer across the exemplar training condition, but did show modest increases in her naming response transferring across naming-photograph probes when a textual prompt was added. Evidence of transfer across video probes continued to be poor, however. Interestingly, her naming performance during non-example probes demonstrated the greatest improvement. This suggests that the textual prompt restricted the range of her responses during non-example probes. Harriet exhibited a tendency to state a variety of names in the presence of both example and non-examples throughout spaced retrieval condition. Moreover, she would often demand during error corrections that the name she stated was indeed the correct name. The textual prompt had the immediate effect of increasing the probability that she correctly stated "I don't know" in the presence of non-examples.

Of note, the target staff member for Barry, Susan, and George provided anecdotal reports of George and Susan independently approaching her during scheduled group activities in the assisted living facility and stating her name. These instances were reported to the researcher on the day of George's 16th session, on which evidence for transfer to the actual person was also formally obtained during a live-person probe. For example, the target staff member reported that George began to frequently approach her and initiate conversation by saying "How's Kate doing?". George would also add after

greeting her that his “Memory is pretty good”. Susan also would initiate conversation by greeting the target by name during bingo activities. Interestingly, Susan continued to engage in verbal mediation prior to saying the target’s name (e.g., “black sweater and white blouse”). Although anecdotal in nature, these reports provide some amount of social validity evidence for the procedures in the current study. The anecdotal reports suggest that socially important gains were seen by improving the quality of interactions between the participants and the target staff member outside of sessions.

Lastly, one week-follow up probes revealed maintenance of non-identity matching-to-sample performance that was equivalent to prior levels across all participants. Barry, Susan, and George’s naming during non-example photograph and video probes also remained at steady levels during follow-up. For both Barry and George, naming during example photograph probes remained higher compared to baseline performance, but was significantly lower as compared to the previous condition. Barry and George’s performance during example video probes revealed a return to zero levels of correct responding. For Susan, naming during both example and non-example photograph probes at one-week follow-up remained at levels commensurate with the previous condition. Similar to Barry and George, Susan’s correct naming during video probes decreased from prior levels, although not to the same degree.

The similar pattern observed across three participants, in which naming performance in the presence of a video probe diminished more greatly over time is interesting, and relates to Skinner’s (1957) description of tact extensions. He argued that any training stimulus has multiple stimulus features, any of which can take control over responding outside of the training context. The number of similar stimulus features that a

novel stimulus shares with the original training stimulus may at least in part determine the extent to which the tact extends, or generalizes, to other stimuli. For example, a child may be taught to tact (i.e., name) a golden retriever as a dog in the presence of that dog. The child later sees a German shepherd, and without further training, tacts “dog”. According to Skinner (1957), this occurs because the stimulus features upon which reinforcement for the tact “dog” is based gains some amount of control over the response. When the child is exposed to a novel dog, most, but not all, of the stimulus features are present (e.g., both animals have long snouts and four feet, but are different in hair color and ears).

A similar interpretation may be made for why naming transferred first to novel photographs, and then shortly after to video presentations of the target staff member. Moreover, follow-up probes revealing greater decreases in naming during video probes as compared to photograph probes suggests that without additional training on a particular photograph of the staff member, stimulus presentations that shared fewer stimulus properties with the original training stimulus would be the first to lose control over the naming response.

Limitations

The present study had several limitations. Although a concurrent multiple baseline design was used to demonstrate replication of the intervention across participants, the modified spaced retrieval conditions were individualized, and therefore the effects were not replicated across participants because the same independent variable was not being manipulated. For replication to occur, the participants would have to have served as their own controls by first withdrawing treatment and reintroducing it.

Reversal to baseline performance would not be expected to occur, however, because the memory training effects would expect to persist across the withdrawal phase. With Harriet, however, a withdrawal of the textual prompt would be defensible. If withdrawing the textual prompt would result in decreased delayed recall, followed by a reintroduction with accompanying increases in delayed recall, then this would lend greater evidence that the textual prompt was the critical variable influencing improvement, and not as a result of repeated exposure to the stimuli.

Another limitation to the current study was that the live-person probe administrations were not equally spaced across conditions. Barry was administered a live-person probe during baseline and after two sessions of spaced retrieval, but was administered such probes after five sessions in to the exemplar training conditions. Thus, without probes being conducted at points in time closer to the end of one phase and the beginning of the next, the detection of any effects are not observed to occur contiguously with the change in the independent variable. Susan was not exposed to a live-person probe during the spaced retrieval condition, making conclusions that can be drawn on the differential effectiveness of spaced retrieval versus modified spaced retrieval impossible. Additionally, live-person probes were only conducted at one-week follow-up for Susan. Data that indicates maintenance of the naming response in the presence of the actual person would be beneficial to determine the socially valid impact the intervention has.

The current study also had a small sample size, with three out of four participants responding to intervention. Future research should replicate the effects of spaced retrieval interventions that target stimulus control deficits across larger numbers of older

adults with NCD to better identify the success rate with this population, and to further explore relevant variables that could influence performance during spaced retrieval.

Another limitation of the current study is that it did not conduct best practice preference assessment or reinforcer assessments for the participants. Information derived from such assessments would aid in maximizing the therapeutic effectiveness of the behavioral intervention. It could be argued that Harriet's poor performance was primarily the result of motivational deficits. In other words, an effective reinforcer was not used as a consequence for correct responding during spaced retrieval trials.

Finally, the follow-up probes were conducted only after one-week of training, resulting in limited statements that can be made regarding the long-term impact of the interventions on naming staff. Maintenance of skills is of particular concern when working with older adults, as the progression of NCD is irreversible and will result in inevitable decline. Maintenance of skills is therefore the goal, and additional follow-ups at longer intervals of time would help to ascertain the extent to which trained skills maintain.

Future Directions

The current study extends the findings of Hawley and Cherry (2004), Hawley, Cherry, Boudreaux, and Jackson's (2008), and Cherry, Walvoord, and Hawley's (2010) by incorporating multiple probe measures to assess for stimulus control deficits, such as poor discrimination between the S+ and S- and limited generalization. The addition of such probe measures also points out limitations of prior research, as rote responding was not assessed. Moreover, supplemental training above and beyond spaced retrieval produced improvement across the full range of probe measures. Future research should

therefore include, at the very least, tests to determine poor discrimination to evaluate the extent to which spaced retrieval results in rote responding, which may limit generalization effects.

Additional research is warranted. Future research could include live video probes. The present study attempted to address the logistical issue of target staff member unavailability for multiple live-person probes in across consecutive sessions by including pre-recorded video probes. With the increased ability of many electronic devices to connect to the Internet wirelessly, many devices have the capability to use live video applications. Live video feeds could provide for a reasonable alternative to the live-person probe, especially if the target person is a family member that lives a great distance away.

The current study also used a baseline recall measure at a 60-s value with a non-targeted staff member. Future research could obtain a more sensitive and valid baseline measure of delayed recall by sampling behavior at titrating recall delays based on performance without feedback. This would result in a baseline condition that is more similar to the spaced retrieval training condition, and would produce a gradient of recall delay values at which responding is always correct, partially correct, and at zero levels.

Future research could also use a multiple baseline design across target stimulus sets to better establish experimental control. By staggering the intervention across target photographs, improvements in performance would be more confidently ascribed to the introduction of the intervention, and not some extraneous variable.

The current study adopted a procedure by Cherry and Simmons-D’Gerolamo (1999), in which recall value days were titrated trial by trial. This may be reason for why

high rates of errors occurred for participants, as performance was unable to stabilize across trials before moving to a new delay value. Thus, future research could evaluate the relative impact on a changing criterion design, in which performance must stabilize (e.g., three correct trials at a given delay) before increasing the delay value.

Finally, future research should evaluate the effects of errorless learning procedures on naming acquisition rates and delayed recall values obtained. Such errorless procedures could entail the provision of zero second prompts, or salient S+ cues that are systematically faded as acquisition increases.

Conclusions

The current study sought to evaluate the effects of a spaced retrieval intervention with multiple probe techniques. Spaced retrieval alone produced increases in median delay values reached, but provided poor generalization across probe measures. The current study also evaluated the impact of tailored conditions that addressed stimulus control deficits identified during probe measures. During these conditions, improvements in both recall and generalization probe measures were observed. The current study adds to the literature by including various probe tactics to assess for treatment effects and to assess for targets to then alleviate with additional training. The current study also is the first to report on the use of probing across non-examples. This was found to be useful in identifying respondents that had a tendency to engage in rote responding. Without the inclusion of non-example probes, it would be difficult to parse out the participant simply "guessing" because the same stimulus is presented in every trial, or if there is evidence of discrimination between example and non-example staff members. The present study also is the first to formally report an obtained baseline of each participant's naming and recall

repertoires. A baseline provides greater evidence that the naming response does not already exist at some response strength prior to intervention, and allows for shifts in responding during intervention conditions to be compared to this baseline state. Finally, the current study utilized intervention targets that were socially valid. By using training stimuli that the older adult is likely to encounter throughout their day, the social importance of the results increases. Spaced retrieval and other interventions to target memory deficits should continue to be researched, given that memory problems are a commonly reported concern in older adults with NCD, particularly the Alzheimer's subtype.

REFERENCES

Alzheimer's Association. (2014). Alzheimer's disease facts and figures. Retrieved on August 31, 2014, from http://www.alz.org/alzheimers_disease_alzheimer_ statistics.asp.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1,* 91-97.

Baker, J. C., Hanley, G. P., & Mathews, R. M. (2006). Staff administered functional analysis and treatment of aggression by an elder with dementia. *Journal of Applied Behavior Analysis, 39*, 469-474.

Baker, J. C., LeBlanc, L. A., & Raetz, P. B. (2008). A behavioral conceptualization of aphasia. *The Analysis of Verbal Behavior, 24,* 147-158.

Baker, J. C., LeBlanc, L. A., Raetz, P. B., & Hilton, L. C. (2011). Assessment and treatment of hoarding in an individual with dementia. *Behavior Therapy, 42*(1), 135-142. doi:10.1016/j.beth.2010.02.006

Bier, N., Provencher, V., Gagnon, L., Van der Linden, M., Adam, S., & Desrosiers, J. (2008). New learning in dementia: transfer and spontaneous use of learning in everyday life functioning. Two case studies. *Neuropsychological Rehabilitation, 18* (2), 204-235.

Blazer, D. (2013). Neurocognitive disorders in DSM-5. *Am J Psychiatry, 170,* 585-587. doi:10.1 176/appi.ajp.2013.13020179

Bourgeois, M. S. (1991). Communication treatment for adults with dementia. *Journal of Speech and Hearing Research, 34,* 831-844.

Bourgeois, M. S. (1993). Effects of memory aids on the dyadic conversations of individuals with dementia. *Journal of Applied Behavior Analysis, 26,* 77-87.

Bourgeois, M. S., & Hickey, El. M. (2009). *Dementia: From diagnosis to management – a functional approach*. New York, NY: Psychology Press

Brown, F. J., Peace, N., & Parsons, R. (2009). Teaching children generalized imitation skills: A case report. *Journal of Intellectual Disabilities, 13,* 9–17. doi:10.1177/1744629509103514.

Brooks, J. O., & Hoblyn, J. C. (2007). Neurocognitive costs and benefits of psychotropic medications in older adults. *J Geriatr Psychiatry Neurol, 20,* 199-214.

Brush, J. A., & Camp, C. J. (1998a). Using spaced retrieval as an intervention during speech-language therapy. *Clinical Gerontologist, 19*, 51-64.

Brush, J. A., & Camp, C. J. (1998b) Spaced retrieval during dysphagia therapy: A case study. *Clinical Gerontologist*, *19,* 96-99.

Buchanan, J. A., & Fisher, J. E. (2002). Functional assessment and noncontingent reinforcement in the treatment of disruptive vocalizations in elderly dementia patients. *Journal of Applied Behavior Analysis, 35*, 99-103.

Buchanan, J. A. (2006). A review of behavioral treatments for persons with dementia. *The Behavior Analyst Today, 7*(4), 521-537.

Buchanan, J. A., Christenson, A., Houlihan, D., & Ostrom, C. (2011). The role of behavior analysis in the rehabilitation of persons with dementia. *Behavior Therapy, 42*(1), 9-21. doi:10.1016/j.beth.2010.01.003

Burgio, L. D., & Burgio, K. L. (1986). Behavioral gerontology: Application of behavioral methods to the problems of older adults. *Journal of Applied Behavior Analysis, 19*(4), 321-328. doi:10.1901/jaba.1986.19-321

Camp, C. J., & Schaller, J. R. (1989). Epilogue: Spaced-retrieval memory training in an adult day care center. *Educational Gerontology, 15*, 641-648.

Camp, C. J., Foss, J. W., O'Hanlon, A. M., & Stevens, A. B. (1996). Memory interventions for persons with dementia. *Applied Cognitive Psychology, 10*, 193-210.

Center for Disease Control (2010). *10 Leading causes of death by age group, Unites States*. Table retrieved on September 7th, 2014 from http://www.cdc.gov/injury/ wisqars/pdf/10l cid_all _ deaths_by_age_group_2010-a.pdf

Cherry, K. E., & Simmons-D'Gerolamo, S. S. (1999). Effects of a target object orientation task on recall in older adults with probable alzheimer's disease. *Clinical Gerontologist, 20*(4), 39-63.

Cherry, K. E., Hawley, K. S., Jackson, E. M., & Boudreaux, E. O. (2009). Booster sessions enhance the long-term effectiveness of spaced retrieval in older adults with probable alzheimer's disease. *Behavior Modification, 33*(3), 295-213.

Cherry, K. E., Walvoord, A. A. G., & Hawley, K. S. (2010). Spaced retrieval enhances memory for a name-face-occupation association in older adults with probable alzheimer's disease. *The Journal of Genetic Psychology, 171*(2), 168-181.

Creighton, A. S., Van der Ploeg, E. S., & O'Connor, D. W. (2013). A literature review of spaced-retrieval interventions: a direct memory intervention for people with dementia. *International Psychogeriatrics, 25*(11), 1743-1763.

Creighton, A. S., Van der Ploeg, E. S.,Camp, C. J., & O'Connor, D. W. (2015). Using spaced retrieval training to teach people with dementia to independently use their walking aids: Two case studies. *Clinical Gerontologist, 38,* 170-178.

Dixon, M., Baker, J. C., & Sadowski, K. A. (2011). Applying skinner's analysis of verbal behavior to persons with dementia. *Behavior Therapy, 42*(1), 120-126. doi:10.1016/j.beth.2010.05.002

Dwyer-Moore, K. J., & Dixon, M. R. (2007). Functional analysis and treatment of problem behavior of elderly adults in long-term care. *Journal of Applied Behavior Analysis, 40*, 679–683.

Engelman, K. K., Altus, D. E., & Mathews, R. M. (1999). Increasing engagement in daily activities by older adults with dementia. *Journal of Applied Behavior Analysis, 32*, 107-110.

Engelman, K. K., Altus, D. E., Mosier, M. C., & Mathews, R. M. (2003). Brief training to promote the use of less intrusive prompts by nursing assistants in a dementia care unit. *Journal of Applied Behavior Analysis, 36*, 129-132.

Fisher, J. E., & Carstensen, L. L. (1990). Behavior management of the dementias. *Clinical Psychology Review, 10*, 611-629.

Fisher, J. E., & Swingen, D. N. (1997). Contextual factors in the assessment and management of aggression in dementia patients. *Cognitive and Behavioral Practice, 4,* 171-190.

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

Fridriksson, J., Holland, A. L., Beeson, P., & Morrow, L. (2005). Spaced retrieval treatment of anomia. *Aphasiology, 19*(2), 99-109.

Friedman, B., Heisel, M. J., & Delevan, R. L. (2005). Psychometric properties of the 15-item geriatric depression scale in functionally impaired, cognitively intact, community-dwelling elderly primary care patients. *Journal of the American Geriatrics Society, 53* (9), 1570-1576.

Hawley, K. S., & Cherry, K. E. (2004). Spaced-retrieval effects on name-face recognition in older adults with probable alzheimer's disease. *Behavior Modification, 28*(2), 276-296.

Hawley, K. S., Cherry, K. E., Boudreaux, E. O., Jackson, E. M. (2008). A comparison of adjusted spaced retrieval versus a uniform expanded retrieval schedule for learning a name-face association in older adults with probable alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology, 30,* 639-649.

Heard, K., & Watson, T. S. (1999). Reducing wandering by persons with dementia using differential reinforcement. *Journal of Applied Behavior Analysis, 32*, 381-384.

Hunter, C. E. A., Ward, L., & Camp, C. J. (2012). Transitioning spaced retrieval training to care staff in an australian residential aged care setting for older adults with dementia: A case study approach. *Clinical Gerontologist, 35*, 1-14.

Hurd, M. D., Martorell, P., Delavande, A., Mullen, K. J., & Langa, K. M. (2013). Monetary costs of dementia in the united states. *New England Journal of Medicine, 368*, 1326-1334.

Kotler-Cope, S., & Camp, C. J. (1990). Memory interventions in aging populations. In E. A. Lovelace (Ed.), *Advances in Psychology* (pp. 231-261). North Holland: Elsevier Science Publishers.

LeBlanc, L. A., Cherup, S. M., Feliciano, L. & Sidener, T. M. (2006). Using choice making opportunities to increase activity engagement in individuals with dementia. *American Journal of Alzheimer's Disease and Other Dementias, 21*, 318-325.

Lin, L., Ya-Ju, H., Su, S., Watson, R., Tsai, B. W., & Wu, S. (2010). Using spaced-retrieval and montessori-based activities in improving eating ability for residents with dementia. *International Journal of Geriatric Psychiatry, 25*, 953-959.

Lopez, M.N., Charter, R.A., Mostafavi, B., Nibut, L.P., & Smith, W.E. (2005). Psychometric properties of the folstein mini-mental state examination. *Assessment, 12,* 137-144.

Lourenco, R. A., & Veras, R. P. (2006). Mini-mental state examination: Psychometric characteristics in elderly outpatients. *Journal of Public Health, 40* (4), 712-719.

Lundervold, D. A., & Lewin, L. M. (1992). *Behavior analysis and therapy in nursing homes*. Springfield, IL: CC Thomas.

Marzullo-Kerth, D., Reeve, S. A., Reeve, K. F., & Townsend, D. B. (2011). Using multiple-exemplar training to teach a generalized repertoire of sharing to children with autism. *Journal of Applied Behavior Analysis, 44,* 279–294. doi: 10.1901/jaba.2011.44-279

McCurry, S. M., & Drossel, C. (2011). *Treating dementia in context: A step-by-step guide to working with individuals and families*. Washington, DC: APA Books.

Murphy, S. L., Xu, J., Kochanek, K. D. (2013). Deaths: final data for 2010. *National Vital Statistics Reports, 61*(4), Hyattsville, MD: National Center for Health Statistics.

National Institute of Aging (2014) *Alzheimer's Disease Fact Sheet*. Retrieved on September 9th, 2014, from http://www.nia.nih.gov/alzheimers/publication/ alzheimers-disease-fact-sheet.

Oleson, C. R., & Baker, J. C. (2014). Teaching mands to older adults with dementia. *Analysis of Verbal Behavior, 30,* 113-127.

Omnibus Budget Reconciliation Act of 1987, Pub. L. No. 100–203, Title IV, Subtitle C, Nursing Home Reform (1987).

Oren, S., Willerton, C., & Small, J. (2014). Effects of spaced retrieval training on semantic memory in alzheimer's disease: A systematic review. *Journal of Speech, Language, and Hearing Research, 57*, 247-270.

Ortman, J. A., Velkoff, V. A., & Hogan, H. (2014, May). *An Aging Nation: The Older Population in the United States*. Population Estimates and Projections retrieved on September 7th, 2014 from http://www.census.gov/prod/2014pubs/p25-1140.pdf.

Schramm, U., Berger, G., Müller, R., Kratzsch, T., Peters, J., & Frölich, L. (2002). Psychometric properties of clock drawing test and MMSE or short performance test (SKT) in dementia screening in a memory clinic population. *International Journal of Geriatric Psychiatry, 17,* 254-260.

Sidman, M. (1971). The behavioral analysis of aphasia. *Journal of Psychiatric Research, 8,*413-422. doi: 10.1016/0022-3956(71)90034-3

Skinner, B. F. (1957). *Verbal behavior*. Engelwood Cliffs, NJ: Prentice-Hall.

Stokes, T., & Baer, D. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10,* 349–367.

Turner, R. S. (2003). Neurologic Aspects of Alzheimer's Disease. In Lichtenberg, P. A., Murman, D. L., & Mellow, A. M. (Eds.), *Handbook of Dementia: Psychological, Neurological, and Psychiatric Perspectives* (pp. 1-24). Hoboken, New Jersey: Joe Wiley & Sons, Inc.

Trahan, M. A., Kahng, S. W., Fisher, A. B., & Hausman, N. L. (2011). Behavior-analytic research on dementia in older adults. *Journal of Applied Behavior Analysis, 44*(3), 687-691.

Trahan, M. A., Donaldson, J. M., McNabney, M. K., & Kahng, S. W. (2014). Training and maintenance of a picture-based communication response in older adults with dementia. *Journal of Applied Behavior Analysis, 47*, 404-409.

Tsolaki, M., Pantazi, T., & Kazis, A. (2001). Efficacy of acetylcholinesterase inhibitors versus nootropics in alzheimer's disease: A retrospective, longitudinal study. *The Journal of International Medical Research, 29,* 28-36.

Wang, P. S., Brookhart, M. A., Setoguchi, So., Patrick, A. R., & Schneeweiss, S. (2006). Psychotropic medication use for behaviorial symptoms of dementia. *Current Neurology and Neuroscience Reports, 6,* 490–495 .

Werner, C. A. (2011, November). The Older Population: 2010. Retrieved on September 2nd, 2014, from the U.S. Census Bureau website: http://www.census.gov/prod/cen2010/briefs/c2010br-09.pdf.

Williamson, P. N., & Ascione, F. R. (1983). Behavioral treatment of the elderly: Implications for theory and practice. *Behavior Modification, 7*, 583-610.

World Health Organization and Alzheimer's Disease International (2012). *Dementia: A Public Health Priority*. Retrieved on September 7th, 2014, from http://www.who.int/ mental_ health/publications/dementia_report_2012/en/.

Wunderlich, K. L., Vollmer, T. R., Donaldson, J. M., Phillips, C. L. (2014). Effects of serial and concurrent training on acquisition and generalization. *Journal of Applied Behavior Analysis, 47*, 1-15.

Appendix A

Example Set of Target Staff Member

Appendix B

Example Spaced Retrieval Array









Appendix C

Spaced Retrieval Data Sheet

Cold Probe	Selection	+ -	1	2	3	4	5	6
	Verbal	+ -						
Exemplar #	Selection	+ -	+ -	+ -	+ -	+ -	+ -	+ -
	Verbal	+ -	+ -	+ -	+ -	+ -	+ -	+ -
Time (sec)		_____ sec	______sec	_____ sec	_____ sec	_____ sec	_____ sec	_____ sec

Instruction to P: "When the buzzer sounds, I would like you to hand me the picture of [Target name] and say that their name is [Target name].
Correct responses: "That's right!" *Increase delay. Rearrange photos.*

Time intervals: 5 10 20 40 60 90 120 150 180 300 420 540 660 780 900

Incorrect response: "Remember, when the buzzer sounds, I would like you to hand me the picture of [Target name] and say that their name is [Target name]. What is this person's name? Which picture would you hand me if you hear the buzzer?" *Decrease delay interval to previous interval at which responding was correct. Rearrange photos.*
Scoring: + indicates Correct Response. – indicates incorrect or response after 10 seconds of the buzzer sounding.

Appendix D

Probe Data Sheet

------------------ Examples -------- ------ Nonexamples -----

MTS	Selection	+ -	+ -	+ -	+ -	+ -	+ -
Naming	Naming	+ -	+ -	+ -	+ -	+ -	+ -
Video	Naming	+ -	+ -	+ -	+ -	+ -	+ -
Live Person	Negative	+ P Pv -	+ P Pv -				
	Positive	+ P Pv -	+ P Pv -				

Hidden Probe Delay: ________	Selection	+ -	+ -
	Verbal	+ -	+ -

S + Change probe Delay: _________	Selection	+ -	+ -
	Verbal	+ -	+ -

Appendix E

Treatment Integrity Checklist

Instructions: Check or tally whether each step was or was not completed by the experimenter. If there was no opportunity, mark not applicable (N/A).

Spaced Retrieval

	YES	NO
1. Experimenter (Exp) begins subsequent sessions with a recall probe.	❒	❒
2. Exp presents photo and name one at a time on board.	❒	❒
3. Exp issues instructions to participant.	❒	❒
4. Exp begins timer.	❒	❒
5. If participant responds correctly, Exp delivers praise.		
5a. Delay is **increased** to next scheduled interval.		
6. If participant responds incorrectly, Exp redelivers instructions and prompts correct response.		
6a. Delay is **decreased** to previous interval.		
7. After each trial, photos are rearranged on the matrix board.		

Exemplar Training

	YES	NO
1. Delay is initially held constant from SR training.	❒	❒
2. Given a correct response from the participant, delay remains the same.		
3. Given an incorrect response from participant, delay decreases.		
4. After each trial, photos are rearranged on the matrix board.		

Probes____ YES NO

	YES	NO
1. For MTS probe, Exp provides instruction.		
2. No programmed consequences given during any probe.		
3. New photos are introduced in array.	❐	❐

Integrity % = Total number of 'Yes' divided by total number of 'Yes' and 'No', multiplied by 100.

Total 'Yes' ______________

Total 'No' _______________

____
(yes)
÷ = ____________ x 100 = __________________%
(_______ + ________)
(yes) (no)

Appendix F

Human Subjects Review Board Approval

WESTERN MICHIGAN UNIVERSITY

Human Subjects Institutional Review Board

Date: December 22, 2014

To: Wayne Fuqua, Principal Investigator
Christopher Walmsley, Student Investigator for dissertation

From: Daryle Gardner-Bonneau, Ph.D., Vice Chair
Daryle Gardner-Bonneau

Re: HSIRB Project Number 14-12-05

This letter will serve as confirmation that your research project titled "Training Sufficient Exemplars in Spaced Retrieval to Promote Transfer of Control with Older Adults with Dementia" has been **approved** under the **expedited** category of review by the Human Subjects Institutional Review Board. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the application.

Please note: This research may **only** be conducted exactly in the form it was approved. You must seek specific board approval for any changes in this project (e.g., ***you must request a post approval change to enroll subjects beyond the number stated in your application under "Number of subjects you want to complete the study***)." Failure to obtain approval for changes will result in a protocol deviation. In addition, if there are any unanticipated adverse reactions or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the HSIRB for consultation.

Reapproval of the project is required if it extends beyond the termination date stated below.

The Board wishes you success in the pursuit of your research goals.

Approval Termination: **December 21, 2015**

251 W. Walwood Hall, Kalamazoo, MI 49008-5456
PHONE: (269) 387-8293 FAX: (269) 387-8276