Item response models for intratask change to examine the impacts of proactive interference on the aging of working memory span

Ryan P. Bowles
Charlottesville, VA

B.S., Wake Forest University, 1996
M.A., University of Chicago, 1999
M.A., University of Virginia, 2003

A Dissertation Presented to the Graduate Faculty of the University of Virginia in Candidacy for the Degree of Doctor of Philosophy

Department of Psychology
University of Virginia
June, 2006

John J. McArdle, Ph.D., *Chair*

John R. Nesselroade, Ph.D.

Timothy A. Salthouse, Ph.D.

Karen M. Schmidt, Ph.D.

David Shreve, Ph.D.

Abstract

Many psychological theories imply the existence of intratask change, that is, change that occurs as a task is being performed, but few statistical models incorporate this concept. Intratask change usually involves three complicating issues: items are not repeated, outcomes are categorical, often dichotomous, and subjects differ in the amount they change. In this research, I developed a family of item response models applicable to the study of intratask change. I used these models, which I call intratask change item response models (ICIRMs), to test a psychological theory involving intratask change: the hypothesis that the age-related decrease in working memory span is at least in part caused by an age-related increase in the effects of proactive interference. Proactive interference is an intratask change concept, as it accumulates throughout a working memory span task and leads to decreased performance. Previous research has ignored the dynamic aspects of PI, so an ICIRM is needed as a direct test of the theory. In order to be identified, ICIRMs require randomized item presentation order; therefore, I collected data on a working memory span task over the internet. A series of ICIRMs were fit to these data, and an exponential ICIRM was found to provide optimal fit. Contrary to expectations, however, average intratask change was positive, indicating increased performance over the task. Age-related differences in intratask change did not account for the age-related decline in working memory span. Simulations suggested that the conclusions on the shape and direction of intratask change were valid, but the conclusions on relations to age may not be reliable. I interpret the results as suggesting that strategy production may be a more important source of individual differences in working memory span than proactive interference.

Table of Contents

Abstract ... ii

Table of Contents ... iii

List of Tables ... vi

List of Figures ... vii

Acknowledgements ... viii

Dedication ... ix

I. Introduction ... 1

II. The role of proactive interference in the aging of working memory span ... 5

Working memory ... 6

Age and WM span ... 7

III. Models for intratask change ... 10

Learning models ... 12

Dynamic item response models ... 16

Across-occasion dynamic IRMs ... *17*

Intratask dynamic difficulty IRMs ... *21*

Intratask change as trait IRMs ... *27*

Multilevel longitudinal models for binary data ... 28

Expanded family of ICIRMs ... 32

Identification ... *36*

IV. Methods 39

WM span task 40

Data collection methodology 42

Internet data collection issues 44

Informed consent 44

Protection of children 45

Protecting anonymity 45

Maintaining experimental protocols 46

Sample characteristics of internet samples 46

Participants 47

Models 52

Hypotheses 54

V. Results 55

Descriptive statistics 56

No change 58

Comparison of change functions 59

Relations to age 66

Summary 69

VI. Simulations 70

Change model differentiation 71

Results 71

Parameter recovery 73

Results ... *74*

VII. Discussion ... 77

Summary of Working Memeory Span Results ... 79

Summary of Simulation Results ... 80

Discussion of Simulation Results ... 81

Discussion of Working Memory Span Results ... 83

Intratask change as strategy production ... 84

Strategy use and proactive interference ... 87

Age and intratask change ... 88

Comparison to previous studies ... 91

Definition of working memory span ... 94

Future Studies ... 95

Conclusions ... 96

References ... 99

Appendix A: Estimation of ICIRMs using WinBUGS ... 119

Appendix B: WinBUGS Script for Exponential ICIRM ... 121

Appendix C: SAS script for simulations with exponential ICIRM ... 122

List of Tables

Table 1. Psychologically relevant change models 33

Table 2. Proportion correct by item length 57

Table 3. Fit statistics for ICIRMs 61

Table 4. Item difficulty results 62

Table 5. Individual change parameter results 63

Table 6. Correlations among change parameters 65

Table 7. Effects of age on individual change parameters of exponential ICIRM 66

Table 8. DICs for simulated data 72

Table 9. Estimated parameter values for simulated data with exponential ICIRM 75

Table 10. Estimated parameter values for simulated data with sample size = 1000 77

List of Figures

Figure 1: Linear operator model for $\alpha = .6$ and $\lambda = .9$. 14

Figure 2: Luce's beta model for $\beta = 1.6$ 15

Figure 3: Path diagram for a latent growth curve for binary outcomes............................ 31

Figure 4. Age histograms for current data and B&S comparison data............................. 51

Figure 5. Education levels for current data and B&S comparison data. 52

Figure 6. Relation between time and proportion correct recall 57

Figure 7. Predicted probabilities from exponential ICIRM... 64

Figure 8. Predicted probabilities for two ages. 69

Acknowledgements

I wish to express my thanks to my advisors, Jack McArdle and Karen Schmidt, for their guidance and advice through the years. Thanks especially to Jack for his supervision on this dissertation. I also want to express special thanks to John Nesselroade for his help particularly in the last year and for our stimulating discussions on measurement and invariance. Special thanks also to Tim Salthouse for our collaborations including the study on which this dissertation is based. I also wish to thank my dissertation committee as a whole: Jack McArdle, John Nesselroade, Steve Nock, Tim Salthouse, Karen Schmidt, and David Shreve for their input on this study.

Dedication

To my wife, Lori, and son, Alexander. This dissertation would not be possible without their support and patience.

I. Introduction

Many research studies in psychology investigate processes leading to change within an individual. Change may occur as a result of an experimental manipulation, or as a result of some associated process such as development or aging. Psychological theories posit the processes by which the change occurs, and individual differences in the expected results. The dynamics aspects of psychological theories have become increasingly complex as new methodologies have been developed, although theorizing has often been constrained by an overemphasis on simplistic statistical techniques (Nesselroade & Schmidt McCollam, 2000). Some recent efforts in quantitative psychology have been directed toward the development of more sophisticated models of change (Böckenholt, 2005; Collins & Sayer, 2001; Moscowitz & Hershberger, 2002; Singer & Willett, 2003) and this has allowed psychological theory testing to move toward more explicit and specific tests of the change processes inherent in the theories (e.g., Ferrer & McArdle, 2004; Ram et al., 2005).

Most of these statistical models have been developed to examine across-task change or across-occasion change. In general, these models have taken as observations total score outcomes of psychological tasks, such as cognitive tasks or questionnaires, and model individual change at the level of the total scores. Some recent efforts have been aimed at more sophisticated merging of models that examine a nonlinear change model together with a psychometrically justifiable measurement model (McArdle, Grimm, Hamagami, Bowles, & Meredith, 2006; Ram et al., 2005).

However, some psychological theories imply *intratask change*, that is, trait change that occurs *as a task is being performed*. The key feature of these theories is that only one observation is available for any given level of the traits inherent in the theory. A new observation is made only after the traits have changed. Few current data analysis models address this type of dynamic system, because of several issues:

1. Tasks often consist of a number of non-repeated items.[1] Thus, the items are changing at the same time that the subject is changing. Separately identifying two dynamic trends can be problematic (Ferrer, Salthouse, Stewart, & Schwartz, 2004).
2. The observations taken during the course of a task are generally categorical responses (with the one common exception of reaction time). The data may be counts (Jansen, 1997), choices among unordered options (Böckenholt, 1993), or dichotomous (usually right/wrong). These types of data offer limited information from each observation. In this research, I focus in particular on tasks with dichotomous outcomes.
3. Subjects change different amounts over the course of the task. A common way to address this heterogeneity is to combine observations from multiple subjects into a single estimate of change. However, this form of aggregation may mask important individual differences (Molenaar, 2004; Nesselroade & Molenaar, 1999) and yield misleading results (Ram et al., 2005). Instead, subject-specific change can be considered a trait of the subject.

In this research, I examine models for the assessment of intratask change with dichotomous outcomes and the application of the models to the understanding of the role of proactive interference in the aging of working memory. In section II, I describe a

---

[1] For consistency, I refer to any stimulus which elicits a single observable response as an *item*. Several other terms are used in various disciplines, with the most common alternative being *trials*.

psychological theory that includes intratask change, the hypothesis that the age-related decline in working memory (WM) span is at least in part caused by an age-related increase in the effects of proactive interference (PI; Bowles & Salthouse, 2003b; Hasher & Zacks, 1988; May, Hasher, & Kane, 1999). PI is a reduction in the ability to correctly respond to a memory item due to interference from previously presented items. PI is clearly a concept of intratask change, with an expected decline in the trait level after each item presentation. This theory contains all three complicating aspects of intratask change:

1. Tests of WM span consist of a number of non-repeated sequences of material that a subject is presented and then asked to recall.
2. Each sequence recall can be scored dichotomously (right/wrong).
3. Individual differences, and in particular, age differences in the effects of PI are an essential part of the theory.

Previous research designed to address this hypothesis has ignored the inherent dynamic nature of PI or addressed it only partially, and therefore is incomplete and perhaps incorrect. A model for intratask change is necessary to directly test the dynamic predictions of the theory.

In section III, I describe existing models for intratask change and expand on the current models for more complicated change processes. I begin with learning models developed under behaviorist and stimulus-response psychology, traditions in which intratask change was central. I then present a number of dynamic item response models, which fall in the general family of item response theory (IRT) models. I next present closely related models for multilevel longitudinal binary data. Finally, I develop a family of item response models for the assessment of intratask change. I call these models intratask change item response models (ICIRMs). I particularly focus on both linear and nonlinear change processes of the type presented by Cudeck and Klebe (2002) and McArdle, Ferrer-Caja, Hamagami, and Woodcock (2001).

I applied ICIRMs to the analysis of a working memory span task. Section IV describes the methods used for collecting data appropriate for understanding PI as intratask change. Because of the difficulty of separately identifying subject-specific change and changes in item effects due to non-repetition of the items, an item presentation order that varies randomly or partially randomly across participants is most appropriate. Although some studies have employed a presentation order that varies randomly within participant (particularly those based on the version of operation span developed by Engle, Cantor, & Carullo, 1992), to my knowledge, no WM span tasks have been administered in an order that varies across participants. Therefore, new WM span data with randomized item presentation order were collected, using an internet data collection plan that capitalized on the large sample sizes available on the internet for tasks that take a small amount of time to complete (Birnbaum, 2000a; Musch & Reips, 2000). The analysis of these data is reported in Section V.

In section VI, I describe a series of simulations designed to assess the validity of the conclusions based on the WM span data. In particular, I addressed whether the correct shape, direction, and magnitude of intratask change can be accurately and precisely estimated. I first considered whether the true shape of intratask change can be recovered by analyzing simulated data with ICIRMs incorporating several different forms of intratask change. Second, I assessed whether the direction and magnitude of intratask change can be recovered by analyzing simulated data with the true ICIRM and examining whether the parameters of the models are recovered. I used as the basis for the simulations the results from the WM span data.

Finally, in section VII, I discuss the results of the simulations and WM span data. I interpret the results of the WM span data analyses in terms of existing theory, focusing particularly on PI as an explanation for the results. I propose a new hypothesis, the *strategy production hypothesis,* that may account for individual differences in WM span task performance. I conclude with a summary of the dissertation goals and findings.

II. The role of proactive interference in the aging of working memory span

Aging is associated with a decline in working memory (WM) span, the amount of information that can be simultaneously stored and manipulated. Some researchers have suggested that at least part of the age-related decline in working memory span is caused by an age-related increase in the effects of proactive interference (PI; Bowles & Salthouse, 2003b; Hasher & Zacks, 1988; May et al., 1999). PI is a reduction in the ability to use WM because of interference from previous presented information. As the effects of PI increase over the course of a WM span task, the increasing PI hypothesis is a theory incorporating intratask change.

In this section, I describe the increasing PI hypothesis. I begin by describing WM and how it is assessed with WM span tasks. Then, I discuss the relation between age and WM span and the inhibition deficit hypothesis, a theory that explains this relation.

Finally, I describe research on the increasing PI hypothesis, an aspect of the inhibition deficit hypothesis, which has almost entirely ignored the implied intratask change.

Working memory

Working memory is a cognitive system for the simultaneous storage and manipulation of information (Baddeley, 1986). A number of accounts of the structure and processes involved in working memory have been proposed (Miyake & Shah, 1999), but all highlight the role of control processes that coordinate the abundance of information that could potentially be stored (Baddeley & Hitch, 1974) or activated (Cowan, 1995; Schneider & Detweiller, 1987) in WM. These control processes differentiate WM from pure storage accounts of short-term memory (Atkinson & Shiffrin, 1968), but also make WM overlap substantially or perhaps be indistinguishable from other cognitive constructs involving control processes, such as attention (Conway & Engle, 1994; Engle, Conway, Tuholski, & Shisler, 1995; Rosen & Engle, 1997) and general fluid intelligence (Kyllonen & Christal, 1990).

Working memory is assessed with tasks demanding simultaneous storage and processing of information. WM assessments aim to measure the amount of information that can be simultaneously stored and manipulated, a capacity known as WM span. Typical WM span tasks involve performing a cognitive task (such as comprehensive reading) while simultaneously remembering one or more pieces of information (e.g., a series of words; Daneman & Carpenter, 1980; Turner & Engle, 1989). Despite using a capacity approach to assess a processing concept, WM span tasks have generated strong evidence of the validity of WM span as a cognitive construct. WM span tasks tend to be as least moderately reliable (Daneman & Merikle, 1996), and factor analyses indicate that

WM span tasks form a single factor closely related to but distinct from short-term memory and general fluid intelligence (Engle, Tuholski, Laughlin, & Conway, 1999; Kyllonen & Christal, 1990).

Age and WM span

WM span has a negative relation to age, with a correlation of approximately r = -.27 (Salthouse, 1994a; Verhaeghen & Salthouse, 1997). The average 70-year old scores at approximately the 21st percentile among all adults on tasks designed to measure WM span tasks (Verhaeghen, Marcoen, & Goossens, 1993). The age differences appear to be independent of the particular task used to measure working memory span (Salthouse, 1994a).

Research suggests that the age-related decline in many forms of cognition may be a direct result of the decline in working memory span. Working memory is an important predictor of many higher-level cognitive abilities, including spatial visualization (Embretson & Schmidt McCollam, 2000; McCollam (Schmidt), 1997; Salthouse, Babcock, Mitchell, Palmon, & Skovronek, 1990), reading comprehension (Daneman & Carpenter, 1980), following directions (Engle, Carullo, & Collins, 1991), and reasoning (Kyllonen & Christal, 1990; Salthouse, 1993a; Süβ, Oberauer, Wittman, Wilhelm, & Schulze, 2002). Most of the age-related decline in these higher-order cognitive abilities is shared (Salthouse, 1994b), and working memory accounts for much of this shared age-related variance: 40% of the age-related variance in reasoning, 30% in spatial visualization, and 46% in episodic memory (Verhaeghen & Salthouse, 1997). Research into the causes of age-related declines in working memory span may therefore inform understanding of age-related declines in cognition in general.

The inhibition deficit hypothesis (Hasher & Zacks, 1988) is one of a number of theories have been proposed to account for the age-related decline in WM span (Light, 1991; Salthouse, 1996; Zacks, Hasher, & Li, 2000). The inhibition deficit hypothesis posits that an important class of control processes in WM are inhibitory mechanisms that "when normally functioning, serve to limit entrance into working memory to information that is along the 'goal path' of comprehension" (Hasher & Zacks, 1988, p. 212). These inhibitory mechanisms serve three functions within WM (Hasher, Zacks, & May, 1999): (a) they restrict access to WM to currently relevant information, (b) they delete no longer relevant information, and (c) they restrict retrieval of information from WM to currently relevant information. Less efficient inhibitory mechanisms allow WM to become crowded with irrelevant information, yielding less storage capacity available for relevant information and therefore lower measured WM span.

Hasher and Zacks (1988) propose that aging is associated with a decline in the efficiency of the inhibitory mechanisms. The decreased efficiency has an effect on older adults' WM through each of the three functions of the inhibitory mechanisms (Stoltzfus, Hasher, & Zacks, 1996): (a) more information is active in WM, (b) information in WM is less relevant, and (c) irrelevant or previously relevant information is more likely to interfere with currently relevant information. The last property suggests that older adults should be relatively more susceptible to the effects of PI, and that greater susceptibility to PI should account for some of the age-related decline in WM span (Bowles & Salthouse, 2003b; May et al., 1999).

Although many aspects of the inhibitory deficit hypothesis have proven problematic when tested empirically (e.g., Burke, 1997; McDowd, 1997), research has generally supported the prediction that age differences in performance on WM span tasks are caused at least in part by differential susceptibility to PI (Lustig & Hasher, 2002; May et al., 1999; Whitney, Arnett, Driver, & Budd, 2001). Furthermore, within-subject manipulations designed to reduce the effects of PI lead to increased scores on measures of WM span (Lustig, May, & Hasher, 2001; May et al., 1999). May et al. (1999) found that this effect is stronger for older adults than for younger adults, and interpreted the results as suggesting that differential susceptibility accounts for some of the age-related decline in WM span (see also Lustig et al., 2001).

Although consistent with the role of PI as a determinant of age differences in WM, these studies failed to account for the finding that the effect of proactive interference *changes over the course of a task*. Intratask change occurs as follows: (a) PI on the first item is usually negligible, as long as interference from previous tasks has been released (Wickens, Born, & Allen, 1963), (b) The second item suffers interference from the first trial, and (c) the third item suffers interference from the first and second trials, etc. (Keppel & Underwood, 1962). Therefore, the effects of PI increase with later presentation (Keppel, Postman, & Zavortnik, 1968), although the rate of increase decelerates (Underwood, 1957). The precise functional form of the change is not known, although some research suggests that an exponential function may describe the increase in the effects of proactive interference (Underwood, 1957; Wixted & Rohrer, 1993).

Research on individual differences in the effects of proactive interference on working memory has not accounted for the dynamic nature of PI. Instead, researchers have looked at how external measures of proactive interference relate to WM span (Kane & Engle, 2000; Rosen & Engle, 1998; Whitney et al., 2001) or at how manipulations designed to reduce PI affect WM span (May et al., 1999). These research methodologies offer indirect means for examining PI, but do not directly address PI as intratask change. A notable exception is Bowles and Salthouse (2003b), who looked at differences in the dynamics across age groups, but not across individuals, and examined only relative age group differences in intratask change, not absolute amount of change. They concluded that older adults found later presented WM span items relatively more difficult than earlier presented items, consistent with an age-related increase in the susceptibility to the effects of PI. These age group differences accounted for approximately half of the age-related decline in WM span. However, they looked only at group differences, not individual differences, in the effects of PI. Because of the finding that group differences need not apply to individuals (Allport, 1937; Estes, 1956; Nesselroade & Molenaar, 1999), the findings of Bowles and Salthouse (2003b) may be inaccurate. In addition, they looked only at relative differences in change, and were not able to identify the direction of change.

Research into age-related individual differences in PI is needed. Assessing individual differences in the effects of PI requires the assessment of individual differences in change over the course of a WM span task. Therefore, because WM span tasks generally have dichotomous outcomes, an item response model for intratask change is appropriate for examining the inhibition deficit hypothesis.

III. Models for intratask change

Intratask change is remarkably prevalent in psychological theories, yet statistical modeling has generally lagged far behind theorizing. One of the most common types of intratask change, learning, was for many years the essence of psychological research primarily in the behaviorist and stimulus-response paradigms. However, mathematical modeling of the dichotomous outcomes of learning, i.e., models addressing how subjects become progressively more likely to respond correctly, did not develop until much later. Many models of learning were proposed, but they generally suffered from intractable assumptions resulting from the lack of statistical methodology. These shortcomings included ignoring individual differences in change and the limitation of tasks to those with repeated items. The use of these models for the most part died with the conversion of psychology to the cognitive framework, when learning was no longer the primary focus of psychology.

Another family of models for intratask change developed out of the IRT tradition. Most dynamic IRMs were developed for one of three reasons, (a) to assess dynamics at the trait level (i.e., across-occasion change) with a more psychometrically valid measurement model, (b) to measure a single stable latent trait by controlling for dynamics inherent in the testing situation, or (c) to account for local dependence among items, where the response to one item affects the probability of a correct response to a later item. Although all of these models are dynamic in nature, very few addressed individual differences in intratask change.

A third family of models arose out of the general linear modeling framework, which includes such techniques as structural equation modeling and logistic regression. Multilevel longitudinal models for binary data, although not originally developed as models for intratask change, can be used as such. These models are closely related to dynamic IRMs, and in many cases, mathematically equivalent. Because they are part of the general linear modeling framework, however, they are limited to change processes that are linear in the subject parameters.

In this section, I present each of these three families of models for intratask change. I then expand on current models for intratask change by developing a general framework for linear and nonlinear change models. I call these models intratask change item response models (ICIRMs). Finally, I discuss identification of these models.

Learning models

The concept of learning has been an important aspect of psychology nearly from its beginning. Such researchers as Thorndike (1898; 1932), Watson (1914), and Skinner (1938) were primarily concerned with the learning process by which a stimulus and response become progressively more strongly related as they are presented together. Learning was generally assessed through changes in reaction time or response likelihood, where response likelihood is the probability that a particular behavior is given in response to a stimulus. Many models were developed to address response time outcomes, such as the classical exponential and power function learning curves (Newell & Rosenbloom, 1981). In addition some influential learning models predicted continuous unobservable outcomes, with no clear relation to observable reality, such as Hebbian learning theory (Hebb, 1949) or, much later, the Rescorla-Wagner model (Rescorla & Wagner, 1972).

Models for changes in the likelihood of a response did not gain prominence until the 1950s and 1960s when the field of stochastic learning theory developed (Bush & Mosteller, 1955; Sternberg, 1963). The basis of stochastic learning theory is that a subject is presented with a stimulus to which there are a finite number of response alternatives. Each alternative has a probability of occurrence. The models predicted how these probabilities changed in response to a feedback event. For example, for two alternatives, positive reinforcement to a response would make the probability of that response subsequently higher and the probability of the alternative response lower.

Most stochastic learning models for two alternatives, in which one alternative can be considered preferable (i.e., correct), were of the form

$$P(X_{i+1} = 1) = \alpha P(X_i = 1) + (1 - \alpha)\lambda \tag{1}$$

where $P(X_{i+1} = 1)$ is the probability of the correct response to item (a.k.a., trial) *i+1*, λ can be interpreted as the asymptotic probability of a correct response, and α is a learning rate reflecting how quickly the subject reaches λ (Bush & Mosteller, 1951; Bush, Mosteller, & Thompson, 1954; Estes, 1950). Various forms of these models generally differed in how λ and α were determined, but all were linear, and were called linear operator models (Sternberg, 1963). An example of the predicted probabilities in the linear operator model is given in Figure 1.

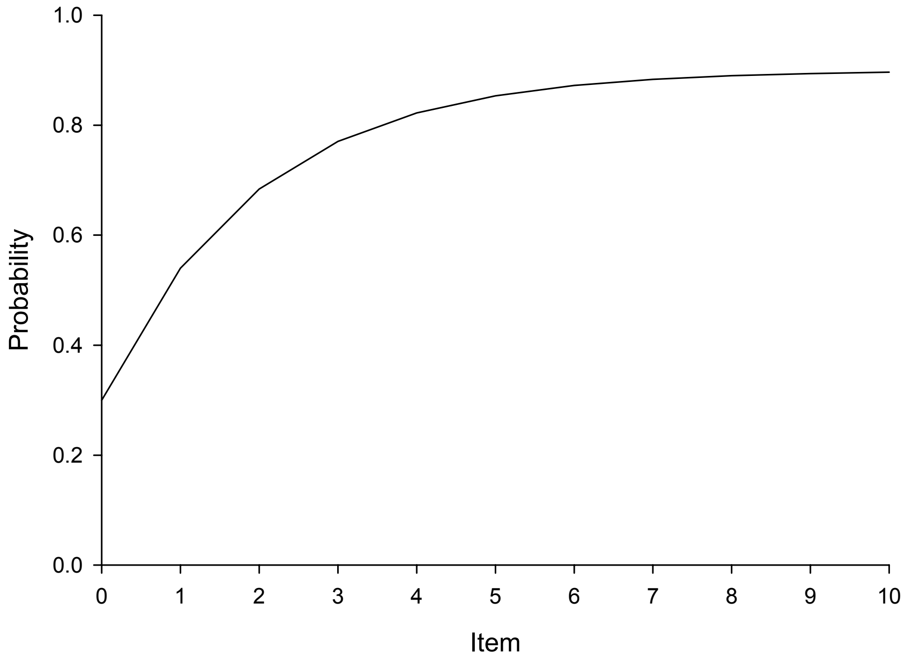

Figure 1: Linear operator model for α = .6 and λ = .9.

The most common alternative models were based on Luce's (1959) beta model, which is of the form

$$P(X_{i+1} = 1) = \frac{\beta P(X_i = 1)}{[1 - P(X_i = 1)] + \beta P(X_i = 1)_i} \tag{2}$$

where β can be considered a learning rate. These models are derived not from modeling probabilities, but from modeling *response strengths*, latent variables reflecting the tendency to select a response relative to all other alternatives, from which probabilities are derived. For two alternatives, the probability is determined by

$$P(X_i = 1) = \frac{v_{1i}}{v_{1i} + v_{0i}} \tag{3}$$

where v_{1i} is the response strength for item i. The beta model is derived from assumption that learning occurs through a proportional change in the response strength of the correct response, $v_{1(i+1)} = \beta v_{1i}$. Alternative forms of the beta model were based on different functional forms for the response strength change, and did not necessarily have meaningful derivable probabilistic expressions (Luce, 1959). An example of the predicted probabilities in the beta model is presented in Figure 2.

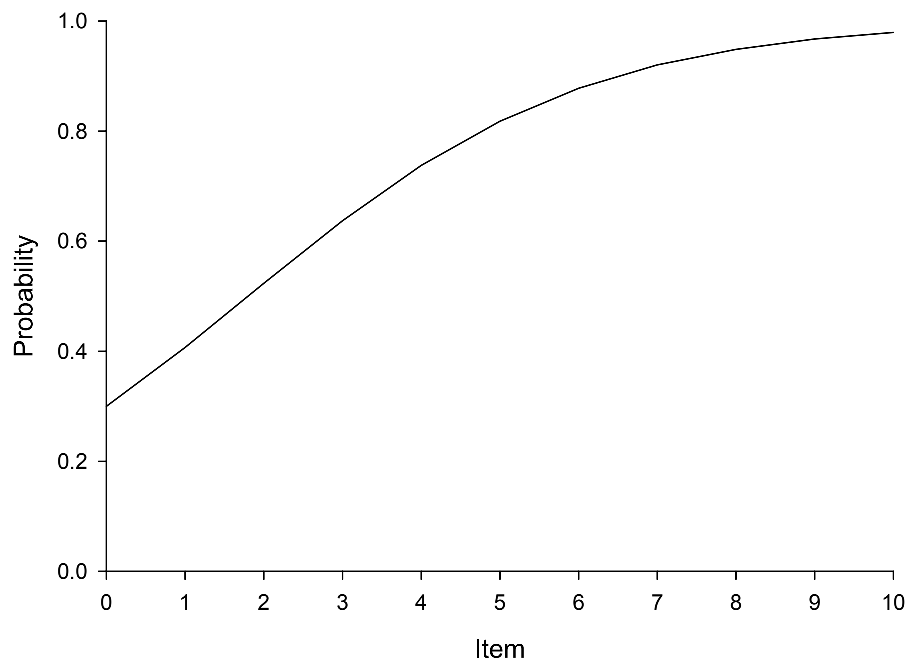

Figure 2: Luce's beta model for β = 1.6

Stochastic learning models included quite restrictive assumptions. It was generally assumed that subjects were replicates of each other (e.g., there was no variance in the initial probability of a correct response) and that the inherent difficulty of each trial was constant (Verhelst & Glas, 1993). Some researchers at the time recognized that these homogeneity assumptions may not be correct in reality (e.g., Sternberg, 1963). However, constraints on estimation tools precluded the development of more realistic models.

It appears that stochastic learning theory had little impact on further development of intratask change models. This may be because as stimulus-response and behaviorist psychology declined along with animal learning experimentation, models most relevant to behaviorist experiments declined as well. Although some later researchers discuss the relationship between more recent dynamic IRMs and stochastic learning theory (particularly Verhelst & Glas, 1993, 1995), further development of intratask change models appears to have arisen out of the item response theory and generalized linear modeling traditions with little appreciation for stochastic learning models.

Dynamic item response models

Item response models (IRMs), usually known as Item Response Theory (IRT) models, are a family of models for analyzing tests, questionnaires, and other instruments containing multiple items with ordered categorical data. The simplest IRM is the Rasch (1960/1980) model for dichotomous responses, under which subjects are represented by a single parameter reflecting their static latent trait, and items are represented by a single parameter reflecting their static difficulty. The model is

$$P(X_{ni}=1)=\frac{\exp(\theta_n-\beta_i)}{1+\exp(\theta_n-\beta_i)} \qquad (4)$$

where θ_n is the trait level of subject *n* and β_i is the difficulty of item *i*. Alternatively, the Rasch model can be expressed in the logit form

$$\ln\left[\frac{P(X_{ni}=1)}{P(X_{ni}=0)}\right]=\theta_n-\beta_i \qquad (5)$$

where $\theta_n - \beta_i$ is known as the logit.[2]

In some cases, the item difficulty parameters, β_i, are considered to be linear combinations of a number of features inherent in the items

$$\beta_i = \sum_{k=1}^{K} b_k z_{ki} \tag{6}$$

where z_{ki} is the value of feature *k* in item *i*, and b_k is the coefficient on the z_{ki} (similar to a regression coefficient). This yields the model

$$P(X_{ni} = 1) = \frac{\exp(\theta_n - \sum_{k=1}^{K} b_k z_{ki})}{1 + \exp(\theta_n - \sum_{k=1}^{K} b_k z_{ki})}. \tag{7}$$

This model is known as the linear logistic test model (LLTM; Fischer, 1973).

As with the Rasch model and the LLTM, most IRMs are designed to assess one or more traits that do not change either across occasion or within task. These models may not yield correct conclusions when the trait is changing. Instead, a dynamic IRM may be required. A few IRMs have been developed to generalize to dynamic processes. Dynamic IRMs generally fall into three categories (Verhelst & Glas, 1995): across-occasion dynamic IRMs, intratask dynamic difficulty IRMs, and intratask change as trait IRMs.

Across-occasion dynamic IRMs

Most effort at generalizing item response models to dynamic processes has been in developing across-occasion dynamic IRMs. Two families of across-occasion IRMs have appeared. The first family consists of models in which change occurs at the trait

[2] Although the logit form is a clearer expression of logistic probability expressions, in this research I use the probability form because a number of models are not expressible in a logit form.

level, with the latent trait measured with an IRM. The second is based on incorporating the change into item parameters, thus allowing estimation using an LLTM.

Across-occasion trait level change IRMs. Trait level change IRMs are based on the idea that change occurs at the trait level, with, at each occasion, the trait level indicated by multiple item responses. In these models, the change model and the IRM are essentially independent. This allows for a simple multistage estimation technique. Trait levels can be estimated separately at each occasion from the within-occasion item responses, and then the trait level estimates can be subjected to across-occasion dynamic modeling. The multistage approach has been used successfully in a number of studies (Bond & Fox, 2001; Dawson, 2000; Lee, 2003; McArdle et al., 2006; McArdle & Hamagami, 2004; Ram et al., 2005).

Alternatively, the change model can be incorporating directly into the IRM. A basic model of this type that makes few restrictions on the shape of the change is the Multidimensional Rasch Model for Learning and Change (MRMLC; Embretson, 1991). The MRMLC is

$$P(X_{ni}=1)=\frac{\exp[f(\Theta_n)-\beta_{it}]}{1+\exp[f(\Theta_n)-\beta_{it}]} \tag{8}$$

where Θ_n is a vector of trait levels, with

$$f(\Theta_n)=\theta_{1n}+I(t\geq 2)*\theta_{2n}+I(t\geq 3)*\theta_{3n}+...+I(t=T)*\theta_{Tn} \tag{9}$$

where $I(t\geq k)$ is an indicator function equal to 1 if the current measurement occasion is the *k*th or later, θ_{1n} is the initial ability and θ_{2n}, θ_{3n}, …, θ_{Tn} are trait level changes. The trait level changes, called ‘modifiabilities’ by Embretson, form a step function, and are essentially equivalent to difference scores in the latent trait between occasions of measurement. In the MRMLC, item difficulty, β_{it}, is indexed by both the item identifier, *i*, and the measurement occasion, *t*. The particular items administered at time *t* need not be identical to the items administered at time *t+1*. Likewise, the difficulty associated with item *i* need not be equal at all time points. In order to link the measurement scales across occasions, that is, to ensure that the latent trait is measured in the same metric at every occasion, at least one identification constraint must be imposed between each occasion, usually on the items (Embretson, 1991; Fischer, 1995). At a minimum, one item could be administered at each occasion with difficulty assumed equal. In practice, though, Embretson and other researchers usually either fix all item difficulties across measurement occasion, so that $\beta_{it} = \beta_i$ for all *t*, or use an LLTM with all item feature coefficients assumed to be time-invariant (Embretson, 1995).

More recent across-task dynamic IRMs have included more explicit nonlinear dynamic processes. For example, Ram et al. (2005) incorporated a cyclical model of emotions into an IRM using data from a daily emotion survey, assuming that item parameters were time-invariant. They found that individuals differed substantially in the parameters associated with a seven-day cyclical trend (i.e., amplitude, phase shift, unexplained variance), thus highlighting the need for allowing for individual differences in dynamic processes. They also found that the dynamic trend could be identified robustly when as few as two randomly selected items per occasion were administered, although this finding may be dependent on the well-defined and restricted nature of the change model they used.

LLTM change models. Another class of across-occasion dynamic IRMs are based on the LLTM. All of these models collapse change into an item parameter. This requires that the dynamics cannot be subject-specific, although they can be group-specific. The simplest LLTM change model is based on the idea that, between each occasion, all subjects change the same amount

$$P(X_{it} = 1) = \frac{\exp(\theta_n + \delta_t - \beta_i)}{1 + \exp(\theta_n + \delta_t - \beta_i)} \tag{10}$$

where δ_t is the amount of change, assumed to be 0 for t = 1 (Fischer, 1976, 1989). So-called virtual items (V-items) can be created, such that the item difficulty of V-item *it*, β^*_{it}, is equal to $\beta_i - \delta_t$. The item features are therefore all dummy indicators of the occasion and the actual (non-virtual) item.

Many generalizations of this model exist (Fischer, 1977, 1989, 1995). The change parameter can be further decomposed into aspects such as group effects (e.g., treatment vs. control) as long as the groups can be linked to a common measurement scale (e.g., by assuming no group differences at pretest). This requirement can be relaxed if, for example, the items change across measurement occasions. In this case, provided the groups can be linked to a common measurement scale at any single occasion (i.e., that the item parameters are invariant across groups at some occasion), group differences in change can be assessed, but not magnitudes (Bowles & Salthouse, 2003b). Furthermore, it is not even necessary that the items measure the same latent trait, so long as the amount of change is the same across all dimensions and the items do not change across measurement occasions (Fischer, 1977).

Relation to intratask dynamic IRMs. In both types of across-task dynamic IRMs, the key issue is linking multiple assessments to the same measurement scale. The linking is accomplished in the same way, by repeating items or item features across measurement occasions, and assuming the associated parameters are invariant. Thus, these models may be applicable to intratask change only if a single item is repeated, or if the different items are assumed to have the same parameter values (i.e., are mathematically or statistically equivalent). Even with this constraint, further assumptions about the trait level parameters are needed because the trait level cannot be estimated with only one item. One option, similar to the LLTM models, is to assume that all subjects change the same amount between each item presentation. This model is a version of Fischer's (1977, 1989) linear logistic model with relaxed assumptions (LLRA). Alternatively, a highly constrained change model could be used, such as the Ram et al. (2005) model, which can capitalize on across-item information to estimate a small number of change parameters. However, an essentially unconstrained change model would not be identifiable, such as the Embretson (1991) model, for which there is one change parameter for each item presentation so that no across-item information is available. In any case, the across-occasion trait level change IRMs must be modified if the items are not repeated.

Intratask dynamic difficulty IRMs

Most of the intratask dynamic IRMs do not allow for subject-specific change processes. Instead, change is conceptualized as a change in the difficulty of the task. These models can be categorized by the way change occurs. Feedback models incorporate dynamics by allowing change in item difficulty to depend on the previous responses. Sequential local dependence models, although developed to address violations of local dependence, are similar to feedback models, but with the dynamics involving only the previous response. Finally, latent state models conceptualize change as transitions in the state a subject is in, with each state characterized by a different set of item difficulties.

Feedback models. In the late 1970s, Kempf (1977) developed what he called the dynamic test model, which can be considered the first dynamic item response model for intratask change developed in the IRT tradition. The dynamic test model is based on the decomposition of the probability of a vector of sequential responses into a multiplicative system of conditional probabilities

$$P(\mathbf{X}) = P(X_i \mid X_{i-1}, X_{i-2}, ..., X_1) * P(X_{i-1} \mid X_{i-2}, X_{i-3}, ..., X_1) * ... * P(X_1) \tag{11}$$

where **X** is an I x 1 vector of responses and X_i is the observed response to item i (i.e., the ith item presented). Let s_{ni} be the vector of responses by subject n before item i, such that $s_{ni} = [X_{n1}\ X_{n2}\ ...\ X_{n(i\text{-}1)}]$. Let $\psi(s_{ni})$ describe how a subject's responses are affected by the previous responses. Kempf proposed the following model

$$P(X_{ni} = 1) = \frac{\zeta_n + \psi(s_{ni})}{\zeta_n + \omega_i} \tag{12}$$

where ζ_n is the baseline ability for subject n and ω_i is the difficulty of item i. If $\psi(s_{ni})$ is 0, then the model is equivalent to the Rasch model (with $\zeta_n = \exp(\theta_n)$ and $\omega_i = \exp(\beta_i)$). However, if $\psi(s_{ni}) \neq 0$, then the model does not have a logit form and therefore must include the awkward constraint that $\psi(s_{ni}) < \omega_i$ in order for the probability to be properly bounded by 1. The function ψ is not subject-dependent, so individual differences in dynamics occur only as a result of different responses to earlier items. Thus, the dynamics in the dynamic test model are entirely attributable to feedback to previous responses, and not to any subject-specific change

The nature of the function $\psi(s_{ni})$ determines the nature of the dynamics in Kempf's model. Kempf suggested that s_{ni} may be summarized by the total score of all previous items,[3] $r_{ni} = X_{n1} + X_{n2} + ... + X_{n(i-1)}$. It can be shown that the initial ability, ζ_n, can be estimated independently of the change function $\psi(s_{ni})$. After deriving this result, Kempf essentially ignores the dynamics, instead focusing on measuring the static initial ability. Thus, the dynamic test model is used simply to "control" for the dynamics, not to examine the dynamics directly.

A closely related model is the dynamic Rasch model (Verhelst & Glas, 1993, 1995; Verguts & De Boeck, 2000). The dynamic Rasch model includes an additive effect of feedback to the baseline level within a Rasch model yielding the model

$$P(X_{ni} = 1) = \frac{\exp(\theta_n + f_i[s_{ni}] + g_i[z_{ni}] - \beta_i)}{1 + \exp(\theta_n + f_i[s_{ni}] + g_i[z_{ni}] - \beta_i)} \tag{13}$$

where $f_i[s_{ni}]$ is defined similarly to $\psi(s_{ni})$ in Kempf's model, although possibly dependent on the particular item *i*, and $g_i[z_{ni}]$ is an item-specific effect of direct feedback from the experimenter (essentially a time-varying covariate). This model overcomes the probability bounds problem of Kempf's model by maintaining a logistic form, because the effects of previous responses and direct feedback are additive with the initial ability and item difficulty. The effects of previous responses and feedback are item-specific, and

[3] This assumption is known as commutativity in the stochastic learning theory literature. The fact that Kempf did not mention this property provides further evidence that dynamic item response models developed independently of stochastic learning theory, particularly in light of the fact that the Bush and Mosteller (1951) model, perhaps the most commonly used stochastic learning model, can be shown to be a specialized case of the Kempf model (Verhelst & Glas, 1993).

there is no subject-specific dynamics. Therefore, $f_i[s_{ni}]+g_i[z_{ni}]-\beta_i$ can be considered as the effective difficulty of a "virtual" item described by the presented item plus the response and feedback history. In this way, the dynamic Rasch model can be formulated as an LLTM. As with Kempf's model, the dynamics are "controlled" for through the nuisance item parameters, and dynamics are not a focus of the model.

Sequential local dependence models. A second class of dynamic IRMs for intratask change conceptualize change as an effect of the presentation order of the items. These models are designed to address sequential local dependence (SLD; Bowles & Grimm, 2006), which is also known as surface local dependence, (Chen and Thissen, 1997), order dependence, (Hoskens & De Boeck, 1997), and recursive conditional dependence (Tuerlincx & De Boeck, 2004). SLD occurs when the response to an item affects the probabilities associated with a later item, usually the next item presented. The first SLD model was proposed by Ackerman and Spray (1987). A slightly modified version (Chen & Thissen, 1997) is based on the idea that, with probability, π, the response to the second item is identical to the response to the first item, while with probability $(1-\pi)$, the response to the second item follows a standard IRM (such as the Rasch model) where

$$\Pr(X_{ni}=1\,|\,X_{n(i-1)}=1)=\pi_i+(1-\pi_i)\frac{\exp(\theta_n-\beta_i)}{1+\exp(\theta_n-\beta_i)} \tag{14}$$

$$\begin{aligned}\Pr(X_{ni}=1\,|\,X_{n(i-1)}=0)&=1-\Pr(X_{ni}=0\,|\,X_{x(i-1)}=0)=1-\left[\pi_i+(1-\pi_i)\left(1-\frac{\exp(\theta_n-\beta_i)}{1+\exp(\theta_n-\beta_i)}\right)\right]\\&=(1-\pi_i)\frac{\exp(\theta_n-\beta_i)}{1+\exp(\theta_n-\beta_i)}\end{aligned} \tag{15}$$

Note that while π can be item-specific, it is assumed to be identical for all subjects. This is a result of the origin of local dependence models in the educational testing literature, which concentrates heavily on item and test characteristics while generally ignoring subject-specific effects. Therefore, the dynamics in the SLD models are attributed to item effects, not to change in the subject.

An alternative approach to SLD is the constant interaction approach of Hoskens and De Boeck (1997; see also Jannarone, 1986). Under the additive interaction approach, the previous response affects the difficulty of the next item instead of directly affecting the probability. The model for the second of two adjacent items is

$$P(X_{ni}=1)=\frac{\exp(\theta_n-\beta_i+\beta_{i(i-1)}X_{n(i-1)})}{1+\exp(\theta_n-\beta_i+\beta_{i(i-1)}X_{n(i-1)})} \tag{16}$$

where $\beta_{i(i-1)}$ is an interaction term reflecting the reduced difficulty on item *i* resulting from a correct response on item *i-1*. This model is equivalent to an intratask feedback model, although with direct feedback from the previous item only.

Although models for SLD are clearly dynamic in nature, addressing change was not a focus in the development of the models. Instead, the dynamics are an emergent property from the ordered nature of SLD. When considering the dynamics, however, it becomes clear that these models are versions of feedback models, where the feedback is a result of the response to the previous item only. Thus, these models have essentially the same properties as feedback models: change as a property to be controlled for, and no subject-specific change process.

Latent state models. In latent state models, subjects are assumed to be in one of a small number of states, with the item difficulty varying across states. The particular state a subject is in for any response is unobservable. The simplest latent state IRM is the Saltus model (Latin for leap; Wilson, 1989). Under the Saltus model, item difficulty depends on the state the subject is in. The model is

$$P(X_{ni}=1)=\frac{\exp(\theta_n-\beta_i-\sum_h \phi_{nh}\tau_{hk})}{1+\exp(\theta_n-\beta_i-\sum_h \phi_{nh}\tau_{hk})} \tag{17}$$

where ϕ_{nh} is a dummy variable indicating whether subject *n* is in state *h*, and τ_{hk} is the effect of being in state *h* on the item difficulty, assumed to be identical for all items of type *k*. The Saltus model was conceived as addressing 'leaps' or discontinuities in development, where a set of items (indexed by *k* in the model) becomes suddenly easier relative to the other items. Although the Saltus model is dynamic in conceptualization, in reality it is designed to model cross-sectional data with subjects falling statically into different latent states. No dynamic process that allows for transitions between states is hypothesized, so a subject is assumed to be in the same state for all item responses.

Adding a latent transition model to the Saltus model yields a version of the model of Rijmen, De Boeck, and van der Maas (2005; see Humphreys, 1998 for a related model). They incorporated a latent Markov transition model into an LLTM. Under the Rijmen et al. model, within each state, an LLTM is assumed to hold, with the item feature coefficients varying across states. Furthermore, there is a transition matrix that may be partially or entirely latent, that governs transitions between states. The transition matrix is invariant across subjects and across time, so that all subject-specific change is expressed in the particular sequence of states the subject goes through. This model can be seen as a generalization of a feedback model, with latent transitions between states instead of observed (i.e., total score on previous items).

Summary. All the dynamic intratask IRMs described in this section share a common property: change is conceptualized as an effect on the item difficulty. Subjects can differ in only two ways: a single (initial) trait level and the particular sequence of stages they go through.[4] These stages can be unobserved, as with the latent state models, or observed, as with the feedback and sequential local dependence models for which the response to the current item indicates the stage the subject will be in for the next item. The dynamic process is assumed to be the same for all subjects, although subjects may differ in the particular outcome of the process. Thus, change is not conceptualized as an attribute of the subject, but as a property of the task that, in many cases, is a nuisance to be controlled for in the measurement of the trait level.

Intratask change as trait IRMs

IRMs that treat change as a subject-specific trait are surprisingly uncommon. To my knowledge, only one has been proposed, Klauer and Sydow's (2001) learning model, which incorporates a linear latent growth curve model into a modified version of the Rasch model. The learning model can be expressed as

$$P(X_{ni}=1) = r + (1-r)\frac{\exp(\theta_n + \delta_n t - \beta_i)}{1+\exp(\theta_n + \delta_n t - \beta_i)} \tag{18}$$

where r is a guessing parameter equal across all items, θ_n is initial ability, and δ_n is learning ability. In this model, the change process (learning) is considered a subject-

---

[4] An exception is the fully unconstrained model of Rijmen et al. (2005), under which the trait level can differ across states, although there are still no interindividual differences in the transition matrix.

specific trait (learning ability, δ_n). Although Klauer and Sydow considered alternative change processes, all that they describe are some form of linear function of an initial ability and a learning ability, perhaps because the estimation technique they used requires the first derivative of the model equation, which can be very complex for nonlinear change models.

Multilevel longitudinal models for binary data

Intratask change models have also been developed in the generalized linear modeling framework, in which they are typically called multilevel longitudinal models for binary data. These models are quite closely related to dynamic IRMs. The Rasch model in its logit form is equivalent to a multilevel logistic regression with items nested within subjects, and can be expressed as

$$\ln\left[\frac{P(X_{ni}=1)}{P(X_{ni}=0)}\right]=\theta_n-\sum_{i=1}^{I}\beta_i D_i \qquad (19)$$

with the additional assumption of a model for the subject trait level parameter, $\theta_n = \mu + \varepsilon$, where D_i is a dummy variable equal to 1 if the item response is to the *i*th item, μ is the mean trait level, usually set to 0 as an identification constraint, and ε is a normally distributed variable with mean 0. The subject-specific trait level, θ_n, is therefore equivalent to a random intercept term (Kamata, 2001; Wilson & De Boeck, 2004). The LLTM can also be easily expressed as a logistic regression by replacing $\sum_{i=1}^{I}\beta_i D_i$ with the linear combination of item features, $\beta_i = \sum_{k=1}^{K} b_k z_{ki}$.

Similarly, the Rasch model can be expressed in latent response propensity notation (Christoffersson, 1975; Muthén, 1984). In this framework, there is a latent response propensity, x^*, with an associated threshold, τ, determining the observed binary response, such that the observed response X is 1 if $x^* > \tau$ and 0 otherwise. x^* is then expressed in a way similar to the logistic regression, with

$$x_{ni}^* = \theta_n - \sum_{i=1}^{I} \beta_i D_i + \varepsilon_{ni} \tag{20}$$

where ε_{ni} is an unobserved residual. The distribution of ε_{ni} determines the probabilities of observed responses, and is typically normally distributed or logistic distributed. In the latter case, the model is equivalent to a logistic regression. In this framework, item difficulty is generally expressed through the threshold parameters (Takane & de Leeuw, 1987). However, if item difficulty is part of the regression equation as above, all thresholds can be set to 0 as an identification constraint.

Incorporation of dynamics into these models has generally been done in the same ways that dynamics are incorporated into regression models for continuous variables. In particular, three dynamic generalizations for binary data have been generally implemented. Historically the first to receive attention was the lagged logistic regression (Bonney, 1987; Liang & Zeger, 1989). In this type of model, the outcome variable is regressed on additional variables that reflect the effect of previous responses or a known function of the previous responses, such that

$$x_{ni}^* = \theta_n - \sum_{i=1}^{I} \beta_i D_i + f(X_{i-1}, X_{i-2}, \ldots, X_1) + \varepsilon_{ni} \tag{21}$$

where f is the known function of previous responses. In practice, f is usually constrained to be a linear combination of a limited number of lagged responses. This lagged logistic regression model is equivalent in general to a feedback IRM, and more specifically to the general form of the dynamic Rasch model (Verhelst & Glas, 1993).

Another common way to incorporate dynamics into a logistic regression is to add a latent growth curve into the regression equation. This type of longitudinal analysis was developed primarily in the structural equation modeling (SEM) literature for continuous outcomes (Meredith & Tisak, 1990; McArdle, 1986; McArdle & Epstein, 1987; McArdle & Nesselroade, 2002), and then was applied to binary data as techniques for SEM with categorical data have become available (Muthén, 1984; Muthén & Muthén, 2004). Research has focused almost exclusively on purely linear models, in which subjects are defined by two traits, a level and a slope, with fixed known coefficients (i.e., factor loadings). This model is equivalent to Klauer and Sydow's (2001) learning model. A path diagram for the linear latent growth curve model with binary outcomes is shown in Figure 3. Additional growth factors that vary across subjects can also be incorporated, but because of the origin of these models in factor analysis, structural equation models, and, more generally, generalized linear models, they are limited to linear combinations of the subject-specific growth parameters.

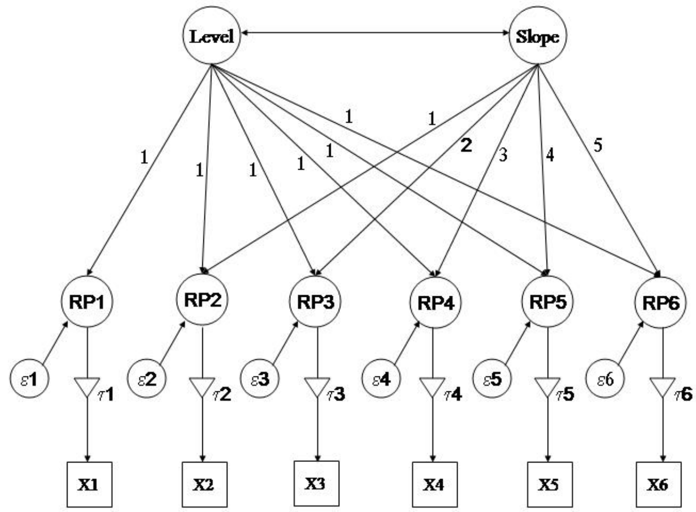

Figure 3: Path diagram for a latent growth curve for binary outcomes. The RPs are latent response propensities and the τs are thresholds.

The third type of dynamics that has been incorporated into a logistic regression is autoregressive error (Muthén, 1996). This type of dynamic model has been less well-used because the possibility of autoregressive error is most directly applicable in a latent response propensity framework. Under an autogressive framework, the residual is given a dynamic structure, affected by one or more previous residuals. For example, with a single autoregressive term (AR-1), the residual is expressed as

$$\varepsilon_{ni} = \rho * \varepsilon_{n(i-1)} + \nu_{ni} \tag{22}$$

where ν_{ni} is an error term. The autoregressive error model has no direct counterpart among dynamic IRMs, although models for SLD are similar in the dependence on the previous time point.

All three types of dynamic logistic regression models can be included in a single model with appropriate identification constraints (Muthén, 1996). These models are quite similar, and sometimes identical, to dynamic IRMs, although the modeling perspective is different: regression vs. item response modeling. I know of no logistic regression models that have incorporated more complex nonlinear change models, likely because this type of model is no longer in the generalized linear modeling framework.

Expanded family of ICIRMs

In this research I expand the family of ICIRMs to incorporate more types of change processes. ICIRMs developed to date have examined only change processes linear in the subject parameters, generally with a subject-specific initial trait level and a subject-specific slope (Klauer & Sydow, 2001). However, linear change models may not be appropriate in many psychological contexts (e.g., McArdle et al., 2002). The particular choice of the change model should depend on the theoretical expectations of the change process, as well as the statistical fit of the change model. I expand the family of ICIRMs by incorporating more complex nonlinear change models into the Rasch model. The general form of the model is

$$P(X_{ni}=1)=\frac{\exp(f(\Theta_n,t)-\beta_i)}{1+\exp(f(\Theta_n,t)-\beta_i)} \tag{23}$$

where Θ_n is a vector of latent trait levels for subject n and f is a (possibly) nonlinear function of Θ_n and time t, or equivalently, order of presentation. The function $f(\Theta_n,t)$ replaces θ_n in the Rasch model, and so I call f the *effective trait level* of person n at time t. f may take many forms. Some common psychologically relevant forms (Cudeck & Klebe, 2002; McArdle et al., 2002; Heathcote, Brown, & Mewhort, 2000) are presented in Table 1.

Table 1

Psychologically Relevant Change Models

Change process	Functional form	Parameter interpretation
	No change	
No change	$f(\Theta, t) = \theta_{1n}$	θ_{1n} = fixed trait level
	Linear in the person parameters	
Linear	$f(\Theta, t) = \theta_{1n} + \theta_{2n} t$	θ_{1n} = initial trait level θ_{2n} = trait level slope
Quadratic	$f(\Theta, t) = \theta_{1n} + \theta_{2n} t + \theta_{3n} t^2$	θ_{1n} = initial trait level θ_{2n} = trait level slope θ_{3n} = trait level acceleration
	Nonlinear in the person parameters	
Exponential	$f(\Theta, t) = \theta_{1n} - \theta_{2n} \exp(-\theta_{3n} t)$	θ_{1n} = asymptotic trait level θ_{2n} = total change from initial level so $\theta_{1n} - \theta_{2n}$ = initial trait level θ_{3n} = rate of change
Dual exponential	$f(\Theta, t) =$ $\theta_{1n} + \theta_{2n}[\exp(-\theta_{3n} t) - \exp(-\theta_{4n} t)]$	$\theta_{1n} + \theta_{2n}$ = initial trait level, but each parameter individually has no clear psychological interpretation θ_{3n} = decline rate θ_{4n} = growth rate

Table 1 cont.

Change process	Functional form	Parameter interpretation
	Nonlinear in the person parameters	
Power function	$f(\Theta,t)=\theta_{1n}-\theta_{2n}t^{-\theta_{3n}}$	θ_{1n} = asymptotic trait level θ_{2n} = total change from initial level so $\theta_{1n}-\theta_{2n}$ = initial trait level θ_{3n} = rate of change
Linear-linear spline	$f(\Theta,t)=\theta_{1n}+\theta_{2n}*t+\theta_{3n}*[\max(t-\theta_{4n},0)]$	θ_{1n} = trait level at change point θ_{2n} = rate of change before change point θ_{3n} = addition to rate of change after change point, so $\theta_{2n}+\theta_{3n}$ = rate of change after change point θ_{4n} = change point, where rate of change changes

Time-invariant covariates can be incorporated into ICIRMs as predictors of the latent trait parameters. For example, if age differences in exponential learning are being examined, the exponential change rate parameter can be determined by the function

$$\theta_{3n}=\gamma * Age_n+\varepsilon_{3n} \tag{24}$$

where γ is a subject-invariant regression coefficient predicting the effect of age on the rate of learning, and ε_{3n} is a subject-specific unobserved residual. Furthermore, time-varying covariates can also be included if the latent trait parameters are allowed to vary over time. However, this is only possible if the latent trait parameter can be decomposed into a deterministic effect of observed time-varying covariates and an unobserved time-invariant residual. For example, the dynamic Rasch model (Verhelst & Glas, 1993), originally expressed as

$$P(X_{ni}=1)=\frac{\exp(\theta_n+f_i[s_{ni}]+g_i[z_{ni}]-\beta_i)}{1+\exp(\theta_n+f_i[s_{ni}]+g_i[z_{ni}]-\beta_i)} \tag{25}$$

can be expressed as a no change ICIRM with a time-varying trait level. The dynamic Rasch model is then expressed as

$$P(X_{ni}=1)=\frac{\exp(f(\Theta_n,t)-\beta_i)}{1+\exp(f(\Theta_n,t)-\beta_i)}=\frac{\exp(\theta_{0nt}-\beta_i)}{1+\exp(\theta_{0nt}-\beta_i)} \tag{26}$$

where

$$\theta_{0nt}=\varepsilon_n+f_i[s_{ni}]+g_i[z_{ni}] \tag{27}$$

with ε_n being a time-invariant residual term (equivalent to θ_n in the original formulation of the dynamic Rasch model).

This family of ICIRMs is based directly on the Rasch model, but with the static trait level reformulated as a dynamic effect. These models can be readily generalized to other IRMs, particularly those in the Rasch family (Rost, 2001), in which subject parameters and item parameters are additively related to each other. For example, the Partial Credit Model (PCM; Masters, 1982) is a generalization of the Rasch model for ordered categorical data. The PCM is expressed as

$$P(X_{ni}=x)=\frac{\exp\left\{\sum_{k=0}^{x}(\theta_n-\tau_{ik})\right\}}{\sum_{z=0}^{m}\exp\left\{\sum_{k=0}^{z}(\theta_n-\tau_{ik})\right\}} \tag{28}$$

where τ_{ik} is an item-specific category threshold parameter, equivalent to β_i for dichotomous data. The PCM can be used as an ICIRM by replacing θ_n with the effective trait level, $f(\Theta_n, t)$. IRMs not in the Rasch family may be more difficult to convert to ICIRMS because most include an interaction between the trait level and item difficulty (called a discrimination parameter in IRT). It is not obvious how the interaction should be included when replacing the trait level, θ_n, with $f(\Theta_n, t)$. To my knowledge, no IRMs for intratask change have been based on IRT models not part of the Rasch family.

Identification

In almost all IRMs, trait level and item difficulty are additively related, so that an identification constraint is needed in order to establish the zero-point of the latent trait scale. This is the case for all the ICIRMs presented here. As the zero-point of the latent trait scale has no substantive meaning, this identification constraint is psychologically unimportant. A common identification constraint, to which I adhere in this research, is to assume that the difficulty of one of the items is equal to 0.

A more important issue for identification of ICIRMs is identification of the change process separately from the progression of item difficulty in the task. That is, item difficulty may be changing at the same time that the trait level is changing. Separate identification of simultaneous change processes can be problematic (Ferrer et al., 2004; Salthouse, Schroeder, & Ferrer, 2004). Change due to one process may be explained equally well by change in the other process. For the ICIRMs, this identification problem is manifest in the additive relation between trait level and item difficulty. At each item presentation, the mean trait level can be absorbed into the item difficulty. This can have important effects on the psychological interpretation, and could possibly yield inaccurate conclusions.

Consider first the linear change model. At time *t*, the logit (log odds) is

$\theta_{1n} + \theta_{2n}t - \beta_{i(t)}$, where the notation $\beta_{i(t)}$ is to emphasize that item *i* is administered only at time *t*. At any time *t*, the mean of $\theta_{2n}t$, $\sum_n \theta_{2n}t = \bar{\theta}_2 t$ can be absorbed into the item difficulty. That is, the logit

$$\theta_{1n} + \theta_{2n}t - \beta_{i(t)} = \theta_{1n} + (\bar{\theta}_2 + \theta^*_{2n})t - \beta_{i(t)} \tag{29}$$

is mathematically indistinguishable from

$$\theta_{1n} + \theta^*_{2n}t - (\beta_{i(t)} - \bar{\theta}_2) = \theta_{1n} + \theta^*_{2n}t - \beta^*_{i(t)}. \tag{30}$$

Thus, the mean of the change parameter cannot be identified apart from change in the item difficulty, so it cannot be determined whether change is generally positive or negative, which may have an important impact on the interpretation of estimation results. That is, the choice of identification constraint is mathematically arbitrary, but not psychologically arbitrary. Klauer & Sydow (2001) addressed this issue by setting the mean of the change parameter to 0, and looked only at whether there were variance and group differences in the change parameter. It is important to note that no matter the choice of identification constraint, the shape of the change function remains linear.

The identification problem in similar, albeit more complicated, when considering the quadratic change model. At time t, the logit is

$$\theta_{1n} + \theta_{2n}t + \theta_{3n}t^2 - \beta_{i(t)}. \tag{31}$$

At time *t*, the mean of $\theta_{2n}t + \theta_{3n}t^2$, $\bar{\theta}_2 t + \bar{\theta}_3 t^2$ can be absorbed into the item difficulty, so two psychologically relevant parameters are not identifiable: mean change and mean acceleration. Nonetheless, the shape of the change function remains quadratic regardless of the choice of identification constraint.

The identification problem for nonlinear change models is substantially more complex. For the exponential change function at time t, the logit is

$$\theta_{1n} - \theta_{2n}\exp(-\theta_{3n}t) - \beta_{i(t)}. \qquad (32)$$

The mean of $\theta_{2n}\exp(-\theta_{3n}t)$, which can be absorbed into the item difficulty, has no closed form expression is terms of the means of θ_{2n} and θ_{3n}. In fact, it is well-known that the curve of the means need not take the same form as the individual curves (Allport, 1937; Estes, 1956; Nesselroade & Molenaar, 1999), and it can be shown that the shape of the curve $f(t) = \dfrac{\sum_n \theta_{2n}\exp(-\theta_{3n}t)}{N}$ more closely resembles a power function change curve than an exponential change curve (Anderson & Tweney, 1997). Therefore, absorbing the mean, *f(t)*, into item difficulty is similar to subtracting a power function from an exponential function. This function need not be exponential in form. Therefore, the choice of identification constraint may affect the shape of the individual change curves. This may lead to an incorrect conclusion about the true change function, which could lead to dramatic mistakes in the psychological interpretation of the ICIRM.

There are two ways to remove the identification problem. First is to assume a change function for item difficulty. With no change function expressed, the implicit assumption is that item difficulty follows a step function with unknown steps. Step functions cannot be estimated concurrently with trait level change (McArdle & Anderson, 1990; McArdle & Woodcock, 1997). However, the identification problem can be eliminated if a more highly constrained function is selected,[5] so that the trait level change cannot be absorbed into the item difficulty. However, this solution may depend critically

[5] Provided the two change functions are separately identifiable. For example, if both change functions are linear, then the slope parameters are not separately identifiable.

on the validity of the item difficulty change model. Little is known about appropriate change models for item difficulty, other than that most tests are given in order of increasing difficulty. Furthermore, tests comparing alternative change models for item difficulty may have minimal power or may be not be available because of the identification problem.

Alternatively, the covariation of item change and trait level change may be eliminated if the item presentation order and item identification are not identical. This can be accomplished if the items are presented in a different order for each subject, so that, overall there is no pattern to item difficulty. An obvious way to achieve this is to present the items in a random or partially random order. In this research, I use data with fully randomized item presentation order to estimate ICIRMs.

IV. Methods

I now describe the methods used to address intratask change on a WM span task. Because of the need for randomized item presentation order in order to separate change in WM span from change in the items, existing data sets involving WM span tasks are not appropriate. Instead, new data on adult WM span were collected over the internet. The foremost advantage of internet data collection is the availability of large sample sizes, particularly if the task is short (Birnbaum, 2000a; Musch & Reips, 2000). For example, over 600,000 participants completed an experiment on implicit attitudes over 18 months (Nosek, Banaji, & Greenwald, 2002b). As WM span tasks typically take less than 10 minutes to complete, internet data collection is an effective methodology for this research. I therefore developed a dedicated web site, www.internetcognition.com, to collect WM span data appropriate to this research. In this section, I first describe the WM span task used in the data collection, followed by the data collection methodology as implemented on the web site. I then describe issues with internet data collection, and the means employed to address these issues. Finally, I describe the characteristics of the sample collected and the models used to analyze the data.

WM span task

One WM span task, a version of operation span (Mogle, 2006; Turner & Engle, 1989; Unsworth, Heitz, Schrock, & Engle, 2005), was administered.[6] The version of operation span employed involved a series of arithmetic problem- letter combinations. Participants solved the arithmetic problem (processing), then were presented a letter which they attempted to remember (storage). Each arithmetic problem consisted of two mathematical operations, the first division or multiplication (set off by parentheses), and the second addition or subtraction. A potential response was presented as part of the problem, and the participant indicated whether the response was correct (right mouse click) or incorrect (left mouse click). An example arithmetic problem is

$(6 \div 2) + 3 = 5$?

Immediately following the response, a letter was presented for 2000 milliseconds. Consistent with previous implementations of computerized versions of operation span (Unsworth et al., 2005; Mogle, 2006), the possible letters were limited to a set of 12: F, H. J, K, L, N, P, Q, R, S, T, Y. After a specified number of arithmetic problem- letter

[6] This version was modified from the original version (Turner & Engle, 1989) by the use of letters instead of words, consistent with all previous implementations of online versions of operation span.

sequences, participants were asked to enter the letters in the order seen using an electronic keypad by clicking on the letters. After entering all the letters, the participant clicked on ‘Enter’ and the next series of arithmetic problems- letters began. Participants were presented a total of 15 series of between 2 and 8 arithmetic problems- letters, a length typical of most WM span tasks, which allowed for maximizing the amount of data per participant while keeping total task time to at most 15 minutes including instruction. Series length varied randomly between 2 and 8 inclusive, and participants did not know the length of the series beforehand. Roughly half way through data collection, a programming error was discovered that had limited series length to 7. Thus, only half of the participants were presented with a maximum series length of 8. Across all participants and items, the average item was of length 4.72.

Consistent with Engle and colleagues’ typical practice for operation span (e.g., Turner & Engle, 1989; Unsworth et al., 2005), a series on which all letters were recalled in the correct order was considered correct regardless of the correctness of the arithmetic problems. Also in line with typical practice, participants who did not achieve at least an 85% success rate on the arithmetic problems were removed from the data set. Thus, for this task, an item response is the recall of a series of letters, with a correct response indicating complete recall.

This version of operation span has been shown to have good measurement properties when intratask change is ignored. Internal consistency (alpha = .78) and test-retest reliability ($r = .83$) were high in a sample of young adults (age ≤ 35; Unsworth et al., 2005). It also loads on a single working memory factor along with paper and pencil and other computerized versions of working memory span tasks (Unsworth et al., 2005), regardless of whether the task is administered in a supervised laboratory setting or over the Internet (Mogle, 2006). Furthermore, reported correlations between age and operation span, although not common in the literature, are in the expected negative direction (Hambrick & Engle, 2002).

Data collection methodology

In this section I describe the procedure employed in collecting data at www.internetcognition.com. All participants who visited the web site went through the same order of presentation:

1. Introduction. The participant was initially presented with a brief description of the experiment and then asked to press a button to begin. This was followed by a description of how to exit the experiment at any time.
2. Request for informed consent. The participant was presented with an electronic version of a standard informed consent agreement. If the participant selected "I agree" then the experiment continued with a screen asking participants not to write down any of the following information.
3. Pretest questionnaire. Seven forced choice questions were asked, with responses selected from drop-down menus. The questions were [response options in brackets]:
    a. What is your gender? [Male, Female]
    b. What is your age? [Under 18, 18, …, 97]
    c. Starting with kindergarten, how many years of formal schooling have you completed? [0 years, 1 year, …, 20 years, more than 20 years]
    d. How healthy are you? [Excellent, Very Good, Good, Fair, Poor]

e. Have you participated in this experiment before? [Yes, No]

f. How many people are in the room with you right now? [just me, 1, 2, 3, 4, more]

g. Can you give 15 uninterrupted minutes to complete this experiment? [Yes, No]

4. Instructions and practice for WM span task. Participants were first presented a very brief general description of the operation span task, followed immediately by more detailed instructions. They were then presented three arithmetic problems in order to practice responding with the left click-incorrect, right click-correct combination. This was followed by practice on the full task, using two series of length 2. The experiment continued regardless of the level of success on the practice items, although the responses to the practice items were recorded.
5. Brief vocabulary test. A short three question multiple-choice synonyms test was presented after the practice. The vocabulary test served two purposes. First, the items are a subset of Salthouse's (1993b) Synonyms Vocabulary Test, which has been presented as part of at least 18 studies (Bowles & Salthouse, 2003a). Therefore, the responses to these three items can be compared to responses from a large laboratory sample. Second, the vocabulary test served to release PI from the practice items.
6. Operation span task.
7. Posttest questionnaire. Three forced choice questions were asked, with responses selected from drop-down menus. The questions were [options presented in brackets]:

   a. Were you able to concentrate the entire time you were doing this experiment? [Yes, No]
   b. How many people are in the room with you now? [just me, 1, 2, 3, 4 or more]
   c. Can we use your results on this experiment for our research? [Yes, No]
8. Completion screen. The participant was thanked for participating, and provided with an estimate of their memory span equal to the length of the longest series to which he recalled all letters correctly. This estimate was not used for any research purposes, but merely as an interesting reward for the participant.

Internet data collection issues

Internet-based research offers several challenges not encountered in the laboratory (Nosek, Banaji, & Greenwald, 2002a; Reips, 2000). Nosek et al. (2002a) provide a comprehensive review of the challenges involved in the Internet data collection. Some of the most important are described below, as well as the manner in which the experiment addressed those challenges.

Informed consent

One advantage of Internet-based research is that the absence of face-to-face interaction with the experimenter removes the most obvious source of coercion (Nosek et al., 2002a). Informed consent, however, remains a vital part of the research process (Frankel & Siang, 1999). This research involved two layers of consent. The first occurred near the beginning, where participants received information about the general purpose of the study, the type of data that is recorded, and the quality of data privacy as part of an electronic version of the standard informed consent agreement. Consent was assumed to be given if the participant clicked on the 'I agree' button. The second layer of consent occurred after the task was completed, when participants were asked whether their responses could be used for the research. If they choose "No", then the data was not recorded.

Protection of children

Protection of children can be a difficult issue for Internet data collection because, while in a laboratory setting, the participation of a child is not likely to pass unnoticed, recognizing these participants over the Internet can be difficult (Nosek et al., 2002a). In this research, participants were required to enter their age before participating. Visitors to the website who selected 'Under 18' for their age were allowed to complete the experiment in the same manner as adult participants, but their data was not recorded. No direct notice was given that children were not allowed to participate in order to minimize any tendency for children to lie about their age in order to try the tasks.

Protecting anonymity

Directly identifying information was not collected. The only demographic data collected was age, education, and self-reported health. Because recruitment was by word of mouth (see the section on the participants below), there was a slight chance that I could identify participants based on their date of participation. For example, close acquaintances would most likely have participated relatively early, so there was a chance that I could identify them from their age and education. In order to minimize this, the date of participation, although collected as part of the data, was deleted from the analyzed data before examination of the data. These efforts at maintaining anonymity and data security functionally render data collected in the research more anonymous than data collected in a laboratory setting, where the participant interacts directly with an experimenter (Nosek et al., 2002a).

Maintaining experimental protocols

The absence of the laboratory setting creates difficulties in standardizing the participant's setting. Distractions such as a ringing telephone are impossible to prevent. Efforts to account for variations in experimental setting occurred in three areas. First, before beginning the task, participants were presented with a list of requirements for completing the research, including instructions to minimize potential distractions, to work alone, and to reserve 15 minutes to complete all aspects of the task. Second, pretest and posttest questionnaires were administered that were designed to assess how well the participant could concentrate on the task. The pretest questionnaire included: How many people are in the room with you right now? and Can you give 15 uninterrupted minutes to complete this experiment?. The posttest questionnaire included: How many people are in the room with you now? and Were you able to concentrate the entire time you were doing this experiment? Finally, the data were examined for obvious problems, especially associated with overly long testing time.

Sample characteristics of internet samples

Internet-based samples, although large, tend to be weighted toward younger participants. Of the 600,000 participants in the Nosek et al. (2002b) experiment, about 50% were under the age of 30, and almost 90% were under the age of 50, compared to

U.S. population figures of about 42% and 73%, respectively (U.S. Census Bureau, 2000). Despite the strong skewing, it is important to note that even in experiments developed early in the history of Internet-based research, samples tended to have at least 5 to 10% of participants over the age of 50 (Bailey, Foote, & Throckmorton, 2000; Gosling, Vazire, Srivastava, & John, 2004). With the increased use of the Internet in daily life, Internet usage has become less heavily weighted toward the young, although still quite skewed (Lenhart et al., 2003). Thus, despite samples skewed toward the young, a substantial number of older adults can be expected to participate in Internet-based psychological studies. In fact, the sample collected in this research was approximately 20% over the age of 50.

Participants

The sample consisted of 403 people at least 18 years old who provided at least the first layer of informed consent, that is, the standard informed consent at the start of the experiment.[7] The participants were recruited through an email recruitment initially sent to my friends, family, and acquaintances. The recruitment email contained a request that the email be sent on to the participants' friends, family, and acquaintances.

I identified seven potential exclusion criteria that could be used to select only those data that were likely consistent with experimental protocols. Participants could be excluded if they:

1. Refused consent for their data to be used. Participants who responded "No" to the posttest question "Can we use your results on this experiment for our research?"

[7] Four records were deleted because the same id was recorded twice for two different ids.

were excluded from the data. No participants responded "No" and so none could be excluded on this basis.

2. Were unable to concentrate throughout the entire task. 86 participants responded "No" to the posttest question "Were you able to concentrate the entire time you were doing this experiment?"
3. Had participated before. 5 participants responded "Yes" to the pretest question "Have you participated in this experiment before?"
4. Exited before completing the WM span task. A small number of participants (*n* = 11) did not complete all 15 items of the WM span task.
5. Spent too long on the arithmetic problems. There are no standards for determining how long a participant should take to complete the operation span task or the component arithmetic problems, nor was there a clear break in the distribution of response times that would suggest differences between responses for which experimental protocols were maintained and those in which protocols were not maintained. Therefore, I chose an arbitrary cutoff of 10000 milliseconds for the average response time to the arithmetic problems that eliminated obvious departures from protocol without removing too many participants. 14 participants who averaged more than 10000 milliseconds to complete each arithmetic problem.
6. Spent too long on the letter recall. As with the arithmetic problems, there are no standards for recall time. I chose an arbitrary cutoff of 5000 ms per letter to be recalled. 5 participants had response times above this threshold.

7. Had poor accuracy on the arithmetic problems. Consistent with the practice established by Engle and colleagues (e.g., Engle et al., 1999; Unsworth et al., 2005), the cutoff was established at 85% success on the arithmetic problems. 59 participants had a success rate below 85%.

Results were similar regardless of which exclusion criteria were employed. Therefore, in this dissertation, I report results from two samples: the full sample consisting of 403 participants, and a conservative sample of 260 participants remaining after all seven exclusion criteria were used.

In the full sample, participants ranged in age from 18 to 86, with a mean of 35.9 and median of 31. 63% were female. The participants were highly educated; one participant reported 6 years of formal education, but otherwise the minimum was 12 years. More than 22% reported at least 20 years. The median level of education was 17 years. If a response of "more than 20 years" is assumed to indicate 22 years of education, then the mean was 17.3 years. Some of the younger participants were likely at the maximum education for their age: the maximum years of education for participants under the age of 20 was 15. Thus, for some of the youngest participants, reported education does not perfectly reflect their final lifetime level of formal education. The correlation between age and education was -.02 and was not statistically significantly different from 0 (p = .72). Participants were on the whole very healthy: 72% said they were in "Very Good" or "Excellent" health, while only 4% said they were in "Poor" or "Fair" health. If the health responses are treated as a five-point interval rating scale, the average health was 3.9 or "Very Good". The correlation between age and health was -.003 (p =.95). Participants performed very well on the three-item vocabulary test. Approximately 75%

correctly answered at least 2 correctly, and 42% correctly answered all 3, with an overall average of 2.1. Older adults tended to have higher vocabulary scores: the correlation between age and vocabulary score was .25 ($p < .01$).

The conservative sample was quite similar. The maximum age was smaller at 76, but the mean and median ages were approximately equal. The conservative sample was somewhat less skewed female, but the sample was still 56% female. The conservative sample was slightly better educated on average, with a mean of 17.5 years. Reported health was approximately the same as the full sample. Performance on the vocabulary test was slightly better: 80% correctly answered at least 2 correctly, and 46% correctly answered all 3, with an overall average of 2.2.

In order to compare these samples to a typical laboratory sample, I compared these data to those from Bowles & Salthouse (2003a; henceforth B&S), who aggregated the data from 18 studies by Salthouse and colleagues. Figure 4 provides a histogram of the age distribution. Approximately 45% of the sample was younger than 30, and 22% were age 50 or older, compared to 15% and 50% for the comparison data.

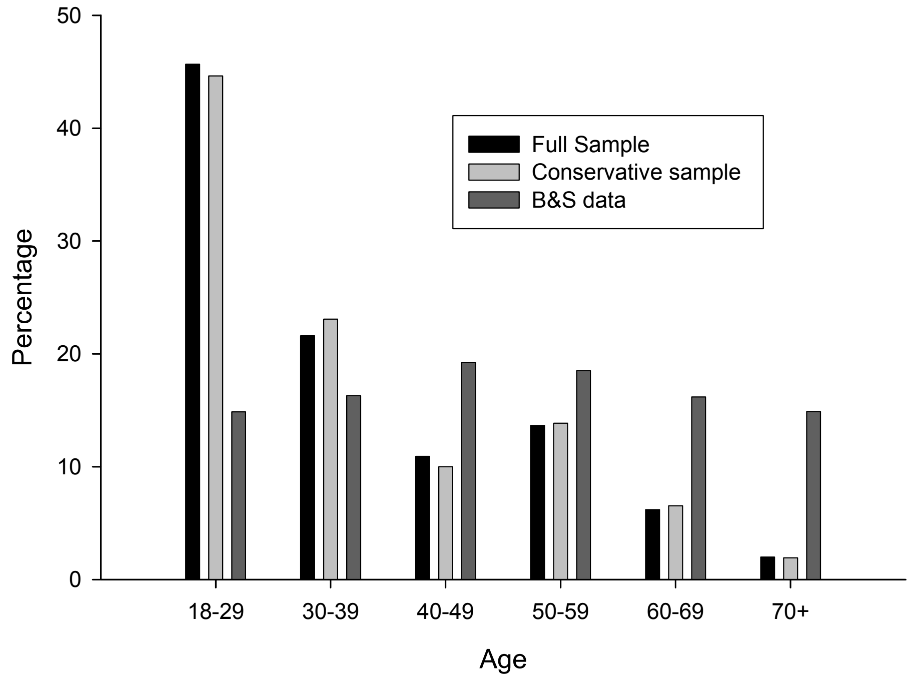

Figure 4. Age histograms for current data and B&S comparison data.

Educational levels for these participants were substantially higher, as the average education for the B&S data was 15.1 years, more than 2 years less. This difference was somewhat larger for younger adults, as shown in Figure 5, although the correlations were essentially equal (full sample: -.02; conservative sample: -.05; B&S data: -.03). Health levels were approximately equal, with the B&S data having average health of 3.9 (compared to 3.9 and 4.0), although the correlation with age was somewhat smaller (full sample: -.003; conservative sample: .03; B&S data: -.14). Finally, average vocabulary scores for the B&S data were equal to the conservative sample (2.2 correct; 2.1 for full sample), and the age correlation was approximately equal (full sample: .25; conservative sample: .21; B&S data: .23). In total, based on the admittedly limited demographic data collected, I conclude that the samples in this study are younger and better educated, but are not substantively different in the relations between age and other variables.

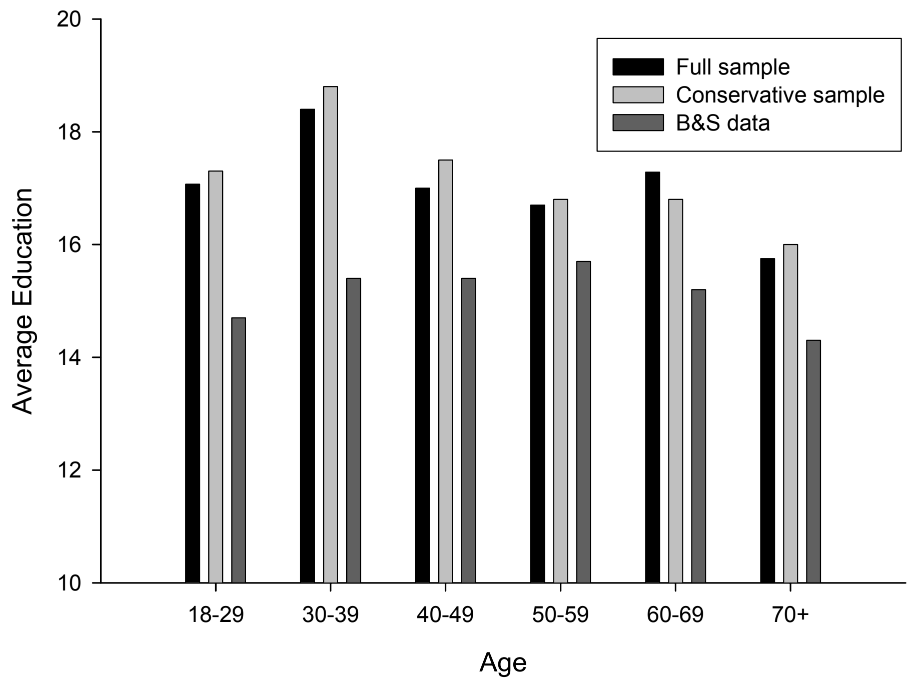

Figure 5. Education levels for current data and B&S comparison data.

Models

The data from the WM span task were analyzed with ICIRMs incorporating a number of change functions. The change functions, $f(\Theta, t)$, considered were:

1. No change: $f(\Theta, t) = \theta_{1n}$
2. Linear change: $f(\Theta, t) = \theta_{1n} + \theta_{2n}(t-1)$
3. Quadratic change: $f(\Theta, t) = \theta_{1n} + \theta_{2n}(t-1) + \theta_{3n}(t-1)^2$
4. Exponential change with common rate parameter *r*:

   $f(\Theta, t) = \theta_{1n} - \theta_{2n} \exp[-r(t-1)]$

5. Exponential change with individual differences in the rate parameter:

$$f(\Theta,t) = \theta_{1n} - \theta_{2n} \exp[-\theta_{3n}(t-1)]$$

6. Dual exponential change with common rate parameters r_d and r_g:[8]

$$f(\Theta,t) = \quad \theta_{1n} + \theta_{2n}\{\exp[-r_d(t-1)] - \exp[-r_g(t-1)]\}$$

7. Power function: $f(\Theta,t) = \theta_{1n} - \theta_{2n}(t-1)^{-\theta_{3n}}$

8. Linear-linear spline with knot point fixed at item 10:

$$f(\Theta,t) = \theta_{1n} + \theta_{2n} * (t-1) + \theta_{3n} * [\max(t-10,0)]$$

9. Linear-linear spline with common knot point K:

$$f(\Theta,t) = \theta_{1n} + \theta_{2n} * (t-1) + \theta_{3n} * [\max(t-K,0)]$$

10. Linear-linear spline with individual differences in the knot point:

$$f(\Theta,t) = \theta_{1n} + \theta_{2n} * (t-1) + \theta_{3n} * [\max(t-\theta_{4n},0)]$$

Note that the term (*t-1*) is included in these equations rather than *t* in order to allow for interpretation of the initial level or intercept at $t = 1$ instead of $t = 0$. Items of a common length (i.e., the same number of arithmetic problems- letters) were assumed to have equal difficulty. Thus, there were 7 item difficulties, one each for items of length 2 to 8. As an identification constraint, the difficulty of item 6 was fixed at 0.

Estimation of the ICIRMs were done in WinBUGS (Spiegelhalter et al., 2003) running within SAS (Zhang, McArdle, Wang, & Hamagami, 2006). A technical description of WinBUGS is provided in Appendix A. WinBUGS yielded an error when attempting to estimate the logistic form of the ICIRMs. Therefore, I used the virtually identical normal ogive version

[8] The dual exponential model with individual differences in the rate parameters did not converge.

$$P(X_{ni}=1)=\Phi\left[\frac{1}{1.7}(f(\Theta_n,t)-\beta_i)\right] \tag{33}$$

where Φ is the cumulative distribution function for the standard normal distribution, and the scaling factor, 1/1.7, is necessary to approximately equate the normal ogive and logistic versions of the model.

Comparisons between models were based on the Deviance Information Criterion (DIC; Spiegelhalter, Best, Carlin, & van der Linde, 2002). Like the AIC (Akaike, 1973) and BIC (Schwarz, 1978), the DIC is a parsimony adjusted fit statistic that balances changes in misfit against changes in model complexity. Lower values of the DIC indicate better parsimony-adjusted fit and a model that provides a more optimal balance between fit and complexity.

Hypotheses

Based on previous research (Underwood, 1957; Wixted & Rohrer, 1993), I hypothesized that the exponential change ICIRM would provide the best fit among the ten models described above, and that the intratask change would be negative, reflecting decline in ability to respond correctly due to increased PI. The three parameters of the exponential change function, asymptotic trait level, total change from initial level, and rate of change, have direct psychological interpretation in terms of WM span and PI. The asymptotic trait level is the WM span fully impacted by PI. The total change from initial level is the total effect of PI on WM span. Therefore, the difference between the asymptotic level and the total change, i.e., the initial level, is WM span free of the effects of PI. Finally, rate of change is the rate of growth in the effects of PI. If the inhibition deficit hypothesis (Hasher & Zacks, 1988) accounts for age-related differences in WM span, then it is predicted that the rate of growth will be positively related to age because older adults are impacted by PI at a faster rate than younger adults. Furthermore, the total change from initial level will be negatively related to age because older adults will have a greater negative total effect of PI than younger adults. Finally, because previous research indicates that differential susceptibility to PI accounts for about half of the age-related decline in WM span (Bowles & Salthouse, 2003b), I hypothesize that the initial trait level will be negatively related to age in order to account for the remaining age relation. To summarize, my hypotheses are:

1. An exponential ICIRM will provide the optimal fit to the WM span data.
2. The total change from initial level will be negative, reflecting decreasing ability as PI is built up.
3. The rate of change will be positively related to age, reflecting faster buildup of PI for older adults.
4. The total change from initial level will be negatively related to age, reflecting greater total impact of PI for older adults.
5. The initial trait level will be negatively related to age, reflecting age-related declines independent of age difference in intratask change.

V. Results

In this section I describe the results for the analysis of the WM span data using ICIRMs. I first present descriptive statistics for the working memory span data from both samples, including observed probabilities. Next, I describe the results from typical practice, that is, ignoring intratask change (i.e., with a no change ICIRM) and the relation of working memory span to age in order to compare these data to extant literature. I then compare the fit of numerous ICIRMs incorporating different change functions to identify the change function that best describes the intratask change. Next, I compare the results from the optimally fitting ICIRM for each sample to the no change ICIRM in order to assess the substantive impact of ignoring intratask change. Finally, I describe results on the relations between the individual change parameters to age. Throughout the results, the alpha level was .05.

Descriptive statistics

The proportion of correct responses overall was .74 for the full sample and .77 for the conservative sample. As shown in Table 2, proportions varied negatively with item length. The frequency with which each item was presented was approximately equal, except for item length 8, which was potentially administered only to roughly half of the participants after a programming mistake was corrected. Figure 6 displays the proportion correct recall across time. There appears to be a tendency for the proportion correct to increase early in the task and stabilize later in the task.

Table 2

Proportion Correct by Item Length

	Full Sample		Conservative Sample	
Item length	Proportion Correct	Frequency of Presentation	Proportion Correct	Frequency of Presentation
2	.96	.15	.97	.15
3	.92	.16	.93	.17
4	.89	.17	.90	.17
5	.79	.15	.81	.15
6	.64	.16	.66	.15
7	.46	.15	.48	.15
8	.25	.07	.25	.07

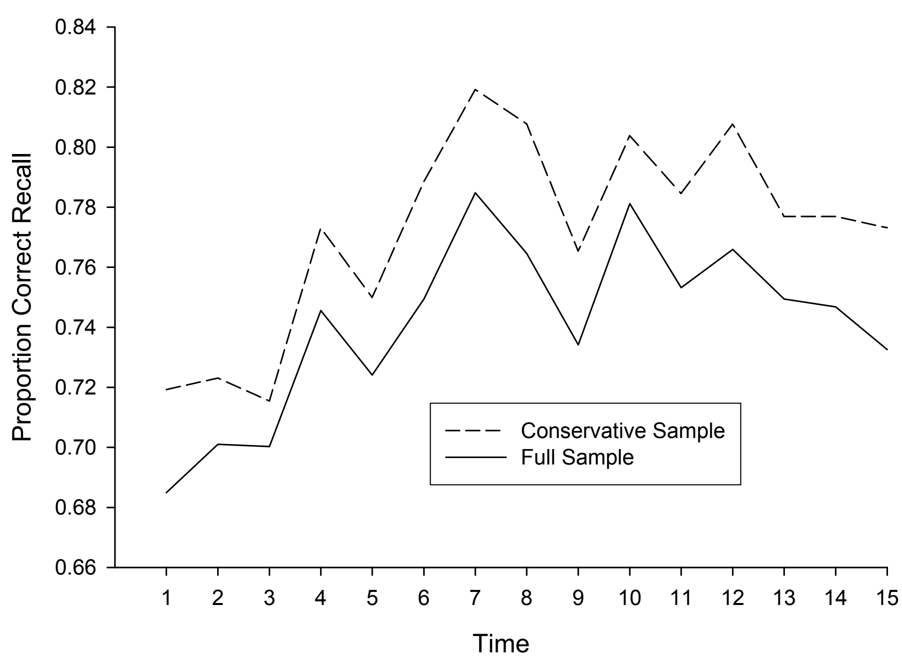

Figure 6. Relation between time and proportion correct recall

No change

The no change ICIRM ignores the possibility of intratask change, and is therefore in line with typical practice with working memory span tasks. The scoring of individual subjects on working memory span tasks varies considerably in the research literature (Conway et al., 2005). Some researchers have used absolute scoring, in which the span score reflects the item length at which accuracy of recall falls below a threshold (e.g., two out of three trials at a given length; Daneman & Carpenter, 1980; Waters & Caplan, 1996). Others have used partial credit load scoring, in which the span reflects the number of elements (e.g., letters) recalled regardless of whether the entire sequence is recalled correctly (e.g., Conway et al., 2005). Still others have used a scoring based on the Rasch model, which is equivalent to the no change ICIRM (Bowles & Salthouse, 2003b). In all scoring options, the scores are based on the same information, and all scoring methods are therefore highly correlated (Conway et al., 2005). Thus, results from the no change ICIRM can be validly compared to existing research, including research on the relation between working memory and age.

For the full sample with the no change ICIRM, as expected, item difficulty increased as the length of the series of arithmetic problems/ letters increased, from -2.96 for item length 2 to 2.21 for item length 8. Participants were relatively able, with an average working memory span trait level of 0.79 (95% CI: [0.57, 0.99]), indicating that the average person responded correctly to the average item with probability .77. This matches closely with the observed proportion correct of .74. There was substantial variance in the fixed trait level (SD = 1.86), indicating substantial individual differences in working memory span. As expected, the working memory span trait level was negatively related to age, with a standardized regression coefficient of -.20 (t_{258} = -3.35, p < .01). Squaring this value, $-.20^2 = .042$, yields the percentage of variance in WM span explained by age. The correlation is smaller than the meta-analytic result of -.27 (shared variance = .073) found by Verhaeghen and Salthouse (1997). I conclude that these data are consistent with previous studies of working memory and age, although with a smaller relation between WM span and age.

Results for the conservative sample were similar. Item difficulty increased as the length of the series of arithmetic problems/ letters increased, from -3.23 for item length 2 to 2.21 for item length 8. Average working memory span trait level was slightly higher, at 0.98 (95% CI: [0.70, 1.26]), indicating that the average person responded correctly to the average item with probability .79, compared to the observed proportion correct of .77. The variance in the working memory span was slightly smaller (SD = 1.72). The correlation with age was -.15 (t_{258} = -2.40, p = .02), yielding a percentage of variance explained of $-.15^2 = .022$.

Comparison of change functions

A key issue in this research is identifying the correct functional form of the intratask change. I hypothesized that exponential change with individual differences in all three parameters would provide the optimal fit for the data. In order to test this hypothesis, I analyzed both samples of the working memory data with ten ICIRMs incorporating different change functions.

Fit statistics for each of the models are presented in Table 3. All models had DIC values clustered within a small range (4592 to 4625 for the full sample; 2759 to 2783 for the conservative sample) except three: exponential change with individual differences in all three parameters (DIC = 4576 and 2664); the closely related power change (DIC = 4474 and 2714); and the linear-linear spline (DIC = 4544 and 2678), which can be considered a linear approximation to the exponential and power models. Thus, I conclude that, as hypothesized, the shape of the ICIRM is approximately exponential with individual differences in all three change parameters. However, the precise shape is not completely clear, as the power ICIRM provided optimal fit for the full sample while the optimal fitting model for the conservative sample was the exponential ICIRM. For the remainder of this dissertation, I will report results for the exponential ICIRM for both samples, noting that the power ICIRM had nearly identical results as the exponential ICIRM for both samples. The WinBUGS script for the exponential change ICIRM is in Appendix B.

Table 3

Fit Statistics for ICIRMs

	DIC	
Change process	Full Sample	Conservative Sample
No change	4625	2783
Linear change	4596	2761
Quadratic change	4615	2772
Exponential change with common rate parameter	4607	2769
Exponential change	4576	2664
Dual exponential change with common rate parameters	4602	2766
Power function	4474	2714
Linear- linear spline with knot point fixed at item 10	4592	2759
Linear- linear spline with common knot point	4595	2762
Linear- linear spline	4544	2678

The exponential change ICIRM has three individual change parameters, the asymptotic trait level, θ_{1n}, the total change from initial level, θ_{2n}, and the rate of change, θ_{3n}. The initial level can also be derived by subtracting the total change from initial level from the asymptotic trait level. There are also 6 item difficulties for items of length 2 – 5 and 7 – 8 (the difficulty of item length 6 was fixed to 0 as an identification constraint).

Item difficulties are reported in Table 4. As expected, difficulty increases with increased length.

Table 4

Item Difficulty Results

	Full Sample		Conservative Sample	
Item	Estimated Difficulty	95% Confidence Interval	Estimated Difficulty	95% Confidence Interval
Length 2	-3.03	[-3.39, -2.68]	-3.33	[-3.88, -2.82]
Length 3	-2.30	[-2.61, -2.01]	-2.29	[-2.69, -1.90]
Length 4	-1.87	[-2.15, -1.60]	-1.95	[-2.32, -1.58]
Length 5	-0.99	[-1.23, -0.74]	-1.10	[-1.43, -0.78]
Length 6	=0		=0	
Length 7	1.04	[0.81, 1.28]	1.09	[0.79, 1.39]
Length 8	2.25	[1.93, 2.58]	2.24	[1.83, 2.65]

Results for the individual change parameters are presented in Table 5. For the full sample, the average asymptotic level was 0.89, which indicates that the average person in this sample will respond correctly to the average item with asymptotic probability .79. The average total change from initial level was 0.59, which yields an average initial trait level of 0.29, yielding a predicted initial probability of .71. Note that, contrary to my hypothesis based on a proactive interference account, the average total change from initial level was positive, indicating that for most participants, the trait level increases throughout the task. The SD of the total change, however, was large relative to mean total change, indicating that for many participants, intratask change was negative. Finally, the average rate of change was 0.73. Results for the conservative sample were very similar. The average asymptotic level was slightly higher (1.15) as was the average total change from initial level (0.71). The average rate of change was also slightly higher (.91). These numbers yields the predicted probability curves displayed in Figure 7.

Table 5

Individual Change Parameter Results

Change factor	Parameter	Full Sample Parameter Estimate	Conservative Sample Parameter Estimate
Asymptotic level, θ_{1n}	Mean	0.89 [0.67, 1.10]	1.15 [0.85, 1.48]
	SD	1.93 [1.52, 2.42]	1.43 [0.98, 1.97]
Total change from initial level, θ_{2n}	Mean	0.59 [0.23, 0.89]	0.71 [0.17, 1.19]
	SD	1.17 [0.01, 3.11]	1.45 [0.11, 3.99]
Rate of change, θ_{3n}	Mean	0.73 [0.46, 1.89]	0.91 [0.34, 1.97]
	SD	0.08 [0.003, 0.76]	0.31 [0.04, 1.08]
Initial level	Mean	0.29 [-0.04, 0.64]	0.44 [0.02, 0.89]
	SD	1.72 [1.34, 2.22]	1.66 [1.21, 2.28]

Note: Numbers in brackets are 95% confidence intervals.

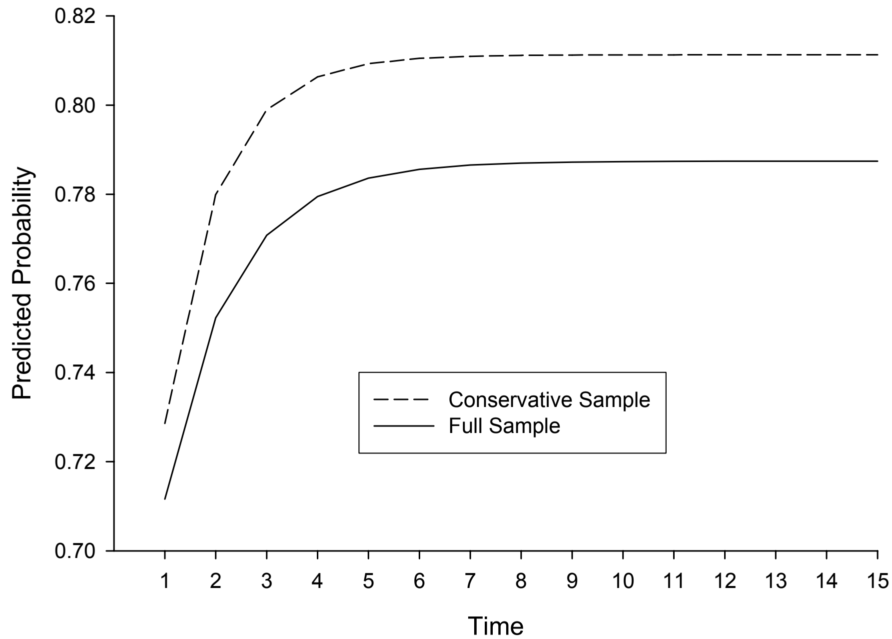

Figure 7. Predicted probabilities from exponential ICIRM

Correlations among the change parameters are listed in Table 6. Almost all were at least moderately high. The correlation between the asymptotic level and the initial level was .96 in the full sample and .93 in the conservative sample, indicating that the two were virtually identical. The correlation between the asymptotic level and the total change from initial level was -.12 and -.36, indicating that participants who had high asymptotic WM span increased least, or, put another way, began the WM span task closer to their asymptotic level. The correlation between the asymptotic level and the rate of change was .25 and .20, indicating that participants who had high asymptotic WM span reached their asymptotic level at a faster rate. Finally, the correlation between the total change from initial level and the rate of change was -.04 and -.21, indicating that participants who increased more from their initial level did so at a slower rate.

Table 6

Correlations among Full Sample Change Parameters

Parameter	Asymptotic level, θ_{1n}	Total change from initial level, θ_{2n}	Rate of change, θ_{3n}	Initial level
		Full Sample		
Total change from initial level, θ_{2n}	-.12			
Rate of change, θ_{3n}	.25	-.04		
Initial level	.96	-.40	.24	
No change trait level	.99	-.17	.26	.97
		Conservative Sample		
Total change from initial level, θ_{2n}	-.36			
Rate of change, θ_{3n}	.20	-.21		
Initial level	.93	-.67	.24	
No change trait level	.96	-.44	.43	.94

Comparing the exponential ICIRM to the no change ICIRM allows for examination of the effect of ignoring intratask change on the understanding of WM span. I compared the individual change parameters from the exponential model to the no change model. Correlations between the no change trait level and the rate of change and total change from initial level were moderately high: .26 and -.16 respectively for the full sample, and .43 and -.44 for the conservative sample. These were in line with correlations between the asymptotic or initial level and the total change and rate of change parameters of the exponential ICIRM. The correlation between the no change trait level and the initial level and asymptotic level from the exponential ICIRM, on the other hand, were .97 and .99, respectively for the full sample, and .94 and .96 for the conservative sample, indicating that these two parameters share almost all of their variance and therefore contain the same information about performance on the WM span task. Thus, I conclude that, although ignoring intratask change does not yield invalid conclusions about one particular aspect of performance (initial or asymptotic performance), it provides only an incomplete understanding, neglecting to consider reliable individual differences in changes in performance across the task.

Relations to age

The final set of analyses examined the relations between age and the individual change parameters of the exponential ICIRM. Based on the inhibition deficit theory, I hypothesized that the rate of change would be positively related to age and the total change from initial level would be negatively related to age. Furthermore, because the inhibition deficit theory does not explain the entire age-related decline in WM span (Bowles & Salthouse, 2003b), I hypothesized that the initial level would be negatively related to change. To examine these hypotheses, I regressed each of the change parameters on age. Results are presented in Table 7.

Table 7

Effects of Age on Individual Change Parameters of Exponential ICIRM

Change Parameter (Outcome)	Intercept	Age Effect	Standardized Age Effect	*t*-value	*p*
		Full Sample			
Asymptotic level	1.44	-.016	-.19	7.52	<.01
Total change from initial level	0.39	.0055	.21	6.44	<.01
Rate of change	0.73	-.000059	-.04	0.69	.49
Initial level	1.05	-.021	-.24	3.98	<.01
		Conservative Sample			
Asymptotic level	1.45	-.0084	-.13	2.10	.04
Total change from initial level	0.50	.0057	.18	3.00	<.01
Rate of change	0.94	-.00094	-.10	1.67	.10
Initial level	0.95	-.014	-.17	2.83	<.01

Contrary to my hypothesis, the rate of change was not significantly related to age. Also contrary to my hypothesis, the total change from initial level was positively related to age. For each year older, the total change increased by .0055 in the full sample and .0057 in the conservative sample, yielding standardized coefficients of .21 and .18, respectively. This finding indicates that older adults increased in ability over the course of the task more than younger adults. Finally, consistent with my hypothesis, the initial and asymptotic levels were negatively related to age. For each year older, in the full sample, the initial level was .021 and the asymptotic level .018 lower; the comparable numbers were .014 and .0084 in the conservative sample. The standardized coefficients were -.24 and -.19 in the full sample and -.17 and -.13 in the conservative sample. Thus, older adults had lower initial and asymptotic trait levels than younger adults.

Interestingly, these effects combined yield virtually parallel predicted probability curves, as displayed in Figure 8. The higher total change for older adults, combined with the slower (although nonsignificant) rate of change, makes older adults look approximately the same as younger adults over the course of the 15 items of this task, except for a level effect associated with the age differences in the initial or asymptotic levels. Thus, over the course of the task, the difference between older and younger adults remains roughly constant. The age-related variance in the asymptotic level is only slightly lower than the age-related variance in the no change trait level: .017 compared to .022, a reduction of 23.1%. Alternatively, the age-related variance in the initial level was .030, an increase of 37.2%. These findings are not consistent with my earlier (Bowles & Salthouse, 2003b) finding that, after accounting for age-group differences in intratask change, the age-related variance in the initial level is about half of the age-related variance in the no change trait level.

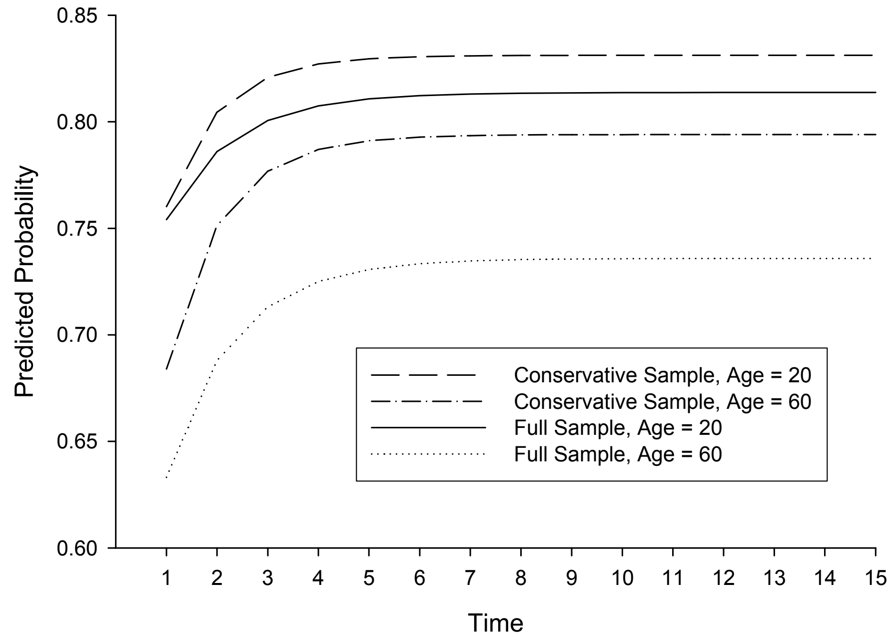

Figure 8. Predicted probabilities for two ages.

Summary

Results did not support PI as the most important cause of age-related changes in WM span. Although the functional form of intratask change was approximately exponential, consistent with previous research on PI (Underwood, 1957; Wixted & Rohrer, 1993), the direction of change was opposite of expectations: positive rather than negative. Instead of a reduction in the ability to respond correctly, as would result from the buildup of PI, intratask change was positive for most participants. On average, the adults in this sample started the task correctly recalling the average WM span item with predicted probability around .72, while on the final item, the probability was approximately .79, substantially higher.

The relations between age and the individual change parameters of the exponential ICIRM were generally consistent with expectations based on general cognitive age-related decline. Older adults performed more poorly initially on the WM span task than younger adults, and developed increased ability at a slower rate, although this difference was not statistically significant. However, older adults displayed greater potential increases, such that over the course of this WM span task, intratask change was approximately equal regardless of age. Thus, for this data, differential intratask change accounted for little of the age-related decline in WM span.

VI. Simulations

ICIRMs are a novel family of statistical models, and it is not clear that the results obtained accurately reflect the true underlying intratask change. In order to confirm the empirical results from the working memory span data, I ran a series of simulations to address three primary concerns about the nature of intratask change. First, interpretation of the results depends critically on identification of the correct shape of the intratask change, that is, the correct change function. To address this issue, I ran a series of simulations to see whether the exponential ICIRM is identified as the correct model when it is the true model. Second, a key finding is that average intratask change is positive. Therefore, it is important to confirm that the total change from initial level is estimated well enough to be confident on the direction of intratask change. Finally, in order to make conclusions on individual differences in intratask change, the magnitude of change needs to be estimated accurately and precisely. To address the last two issues, I ran a series of simulations to assess how effectively the parameters of an ICIRM can be recovered, focusing on exponential change. For all simulations, I used the parameter estimates from the conservative sample of the working memory span data as the true parameter values.

Change model differentiation

An important issue for assessing change is the ability to differentiate among the change models. Therefore, I ran additional simulations to address how well ICIRMs can be differentiated. As with the simulations above, I simulated data from the exponential ICIRM using the estimated parameters from the conservative sample as the true values. I then analyzed the data with ICIRMs incorporating the ten change functions used in the earlier empirical analyses. Comparisons between models were based on the DIC. I simulated 10 data sets.

Results

DICs for each of the ICIRMs, as well as the average across all ten simulated data sets, are presented in Table 8. For 9 of the 10 data sets, as well as the average, the power ICIRM was the optimally fitting model. The only exception was one data set for which the linear- linear spline ICIRM was optimal. The exponential ICIRM, which was the true model, was the second most optimal model for eight of the ten datasets, as well as the average. The only exceptions were one data set for which the linear- linear spline ICIRM was second most optimal, and one data set for which the exponential ICIRM with a common rate parameter was second most optimal.

Table 8

DICs for Simulated Data

	Data Set				
Change process	1	2	3	4	5
No change	2960	2901	2938	2905	2889
Linear change	2950	2884	2941	2893	2881
Quadratic change	2983	DNC	2983	2908	2921
Exponential change with common rate parameter	2953	2890	2929	2873	2845
Exponential change	*2929*	*2856*	*2921*	*2870*	*2849*
Dual exponential change with common rate parameters	DNC	DNC	DNC	DNC	DNC
Power function	**2820**	**2707**	**2902**	**2724**	**2745**
Linear- linear spline with knot point fixed at item 10	2955	2885	2941	2893	2880
Linear- linear spline with common knot point	2950	2885	2941	2893	2881
Linear- linear spline	2949	2879	2938	2888	2877

Note: DNC indicates that the model did not converge. DIC values in bold are the lowest within the column. Values in italics are second lowest.

Table 8 cont.

	Data Set					
Change process	6	7	8	9	10	Average
No change	2785	2908	2923	2914	3035	2915.8
Linear change	2766	2899	2905	2898	3015	2903.2
Quadratic change	2789	2942	2926	2904	3027	2931.4
Exponential change with common rate parameter	2776	2893	2909	2903	3005	2897.6
Exponential change	*2746*	*2875*	2895	*2890*	2998	*2882.9*
Dual exponential change with common rate parameters	DNC	DNC	DNC	DNC	DNC	DNC
Power function	**2684**	**2866**	**2841**	**2837**	*2978*	**2810.4**
Linear- linear spline with knot point fixed at item 10	2768	2898	2903	2896	3006	2902.5
Linear- linear spline with common knot point	2767	2896	2904	2898	3015	2903.0
Linear- linear spline	2765	2898	*2884*	2898	**2955**	2893.1

Note: DNC indicates that the model did not converge. DIC values in bold are the lowest within the column. Values in italics are second lowest.

Parameter recovery

In order to assess the ability to recover the true values of the parameters for the WM span task, I ran a series of simulations to examine how closely the parameter estimates matched the simulated true values. As the primary goal was to confirm the validity of the conclusions from the WM span task, I used as the basis of the simulations the structure of the conservative sample of the WM span task: 15 items, 260 subjects. I simulated multivariate normal individual change parameters using the mean and covariance structure estimated from the empirical data. Item responses were then simulated based on the exponential ICIRM with time-specific trait level determined by the exponential change function and the simulated individual change parameters, with item difficulties set to the empirical estimates. All simulations were performed in SAS using Moser's (2006) script for simulating multivariate normal data. Estimation using the simulated data was done with WinBUGS. The simulations were repeated 25 times. The script is in Appendix C. I employ two terms to describe the results of simulations: *accuracy*, defined as how close the average estimated parameter values were to the true values, and *precision*, defined as how much the estimated parameter values varied around their mean (low precision reflects high variability).

Results

Estimates from the simulations using WinBUGS are presented in Table 9. The estimated item difficulty parameters closely matched the true values, with a slight inward bias associated with estimation using normal priors (e.g., Warm, 1989). The estimates were quite precise, as the standard deviation of the estimates was small, particularly in relation to the variability in item difficulty and the variance of the asymptotic level. The mean asymptotic level was also estimated both accurately and precisely, although the standard deviation of the asymptotic level was overestimated by a factor of about 1.25. The mean total change from initial level was slightly underestimated and with somewhat less precision than the asymptotic level. The SD of the total change was also overestimated, by a factor of 1.35, and with low precision. The rate of change, on the other hand, was estimated very poorly: it was overestimated by a substantial amount, and with very little precision. However, the SD of the rate of change was accurately estimated with relatively high precision.

Table 9

Estimated Parameter Values for Simulated Data with Exponential ICIRM

Parameter		True Value	Mean of Estimated Values	SD of Estimated Values
Item Difficulty	Length 2	-3.20	-2.98	0.18
	Length 3	-2.30	-2.15	0.18
	Length 4	-1.90	-1.80	0.19
	Length 5	-1.10	-1.02	0.15
	Length 7	1.00	1.03	0.13
	Length 8	2.30	2.26	0.13
Asymptotic Level, θ_{1n}	Mean	1.15	1.13	0.20
	SD	1.43	1.80	0.28
Total Change from Initial Level, θ_{2n}	Mean	0.71	0.57	0.37
	SD	1.45	1.96	1.63
Rate of Change, θ_{3n}	Mean	0.91	2.18	1.39
	SD	0.31	0.33	0.38

Note: SD is the standard deviation. Mean and SD of estimated values are based on 25 simulations.

Because the rate of change was poorly estimated, I ran a second series of simulations to test whether a larger sample would have allowed for better estimation. All true parameter values were set to the same values as before; only the sample size was increased, from 260 to 1000. The simulation was repeated 10 times. Results are presented in Table 10. The rate of change was indeed estimated more accurately and precisely, although it was still overestimated. Furthermore, the SD of the rate of change was underestimated. Other results were similar to those with the smaller sample size, including that the total change from initial level was underestimated, and its SD overestimated.

Table 10

Estimated Parameter Values for Simulated Data with Sample Size = 1000

Parameter		True Value	Mean of Estimated Values	SD of Estimated Values
Item Difficulty	Length 2	-3.20	-2.94	0.14
	Length 3	-2.30	-2.15	0.08
	Length 4	-1.90	-1.80	0.10
	Length 5	-1.10	-1.02	0.11
	Length 7	1.00	0.97	0.08
	Length 8	2.30	2.27	0.07
Asymptotic Level, θ_{1n}	Mean	1.15	1.13	0.05
	SD	1.43	1.70	0.16
Total Change from Initial Level, θ_{2n}	Mean	0.71	0.49	0.11
	SD	1.45	2.14	1.23
Rate of Change, θ_{3n}	Mean	0.91	1.32	0.58
	SD	0.31	0.11	0.07

VII. Discussion

A primary goal of this dissertation was to develop a new class of models to address intratask change, that is, change in a person's trait level that occurs during the course of a task. Intratask change is an integral aspect of many psychological theories. However, available models and methods for examining intratask change have been limited by restrictive and often untenable assumptions or requirements, such as a requirement that the task consist of a single repeated item, or an assumption that all individuals are identical. Models for intratask change that allow for nonlinear change on tasks consisting of non-repeated items with dichotomous or categorical outcomes have not previously been generally available. This dissertation provides a new means for analyzing this type of data: the family of intratask change item response models (ICIRMs). ICIRMs are a generalization of standard item response models, but with the standard static trait level parameter replaced by a function reflecting change occurring during a task.

The family of ICIRMs was applied to a psychological theory that implies intratask change yet has never been examined in that framework. The increasing PI hypothesis, an aspect of inhibition deficit theory (Hasher & Zacks, 1988), states that older adults perform more poorly on WM span tasks in part because they are less able to suppress information from previous items that is no longer relevant. PI is an intratask change concept, as PI builds over the course of the WM task at a decelerating rate. Thus, the increasing PI hypothesis would predict that intratask change is negative and decelerating, with older adults experiencing greater decline over the course of the task.

In order to confirm the usefulness of ICIRMs in yielding valid conclusions about the shape, direction, and individual differences in the magnitude of intratask change, I examined the results of several simulations. The simulations addressed these issues in two ways. First, to examine the validity of conclusions about the shape of intratask change, I examined whether ICIRMs with different change function can be differentiated from a true exponential ICIRM. Second, to examine the validity of conclusions about the direction and magnitude, I assessed with the accuracy and precision with which the parameters associated with intratask change can be recovered. For the simulations, I used as the true values the results from the WM span data analyses. This allowed for interpretation of the results from the WM span analyses in terms of the simulated results.

In this section I describe and interpret the results of the WM span analyses and simulations. I first summarize the results from the analyses and simulations. I then discuss the results of the simulations, followed by a discussion of the results for the WM span data. I first focus on the finding that average intratask change is positive and the relation of this finding to PI accounts of WM. This is followed by a discussion of the results in terms of the age-related decline in WM span. Finally, I conclude with a brief summary of the contributions of this dissertation to psychological theory and methods.

Summary of Working Memeory Span Results

Results from analyses of WM span data using ICIRMs did not match expectations based on proactive interference and inhibition deficit theory. Based on this theory and research on the nature of proactive interference, I had five hypotheses about the shape, direction, and individual differences in the magnitude of intratask change.

First, as predicted, an exponential ICIRM provided the optimal fit to the WM span data. The exponential ICIRM had much lower DIC than almost all the other models, with the exceptions being other models that have shapes very similar to the exponential change function.

Second, contrary to expectations, intratask change was positive for the majority of participants. That is, participants' ability to respond correctly generally increased over the course of the task. Proactive interference yields negative intratask change. Therefore, proactive interference cannot be considered the dominant cause of intratask change.

Third, contrary to expectations, the rate of change was not related to age, although the estimate was negative. In light of the simulation results, the rate of change appears not to be estimated well. Therefore, the true rate of change may be related to age, although the poor estimation may mask this relation.

Fourth, contrary to expectations, the total change from initial level was positively related to age. Older adults grew more in the ability to correctly respond to WM span items over the course of the task than younger adults. However, when combined with the nonsignificant negative relation between age and the rate of change, the intratask change was essentially parallel for younger and older adults, suggesting that, overall, there was no relation between age and intratask change.

Fifth, as hypothesized, the initial trait level was negatively related to age. Older adults started with lower ability to respond correctly to WM span items than younger adults. Older adults also had lower asymptotic level than younger adults.

Summary of Simulation Results

The simulations were aimed at assessing the validity of the findings on WM span. In particular, they addressed the validity of conclusions that the shape of intratask change was exponential, that the typical direction of intratask change was positive, and that there were substantial individual differences in the magnitude of intratask change.

First, I examined whether the shape of intratask change can be recovered. When the exponential ICIRM was the true model underlying individual performance, the power ICIRM was identified consistently and strongly as the optimally fitting model. In almost all cases, the exponential ICIRM was the second most optimally fitting model. The power ICIRM and exponential ICIRM are closely related and very similar in shape (Anderson &

Tweney, 1997; Heathcote et al., 2000). Therefore, this finding is not particularly surprising, and indicates that the shape of the intratask change is essentially identifiable.

Second, I considered whether the direction of intratask change, as indicated by the total change from initial level, is accurately identifiable. The total change from initial level was slightly underestimated and the standard deviation of the individual differences in total change was overestimated, but the direction was accurately recovered.

Third, I examined whether the magnitude of change, as well as individual differences in the magnitude, are accurately recovered. The rate of change was strongly overestimated with low precision. Therefore, the magnitude of intratask change, which is determined by the rate of change together with the total change from initial level, is not recovered well. As a consequence, individual differences in the intratask change may not be estimated reliably. Because of the poor estimation of the rate of change parameter, I considered whether a larger sample size would lead to better estimation. I examined the impact of increasing sample size from 260 to 1000. This did indeed result in more accurate and precise estimation of the rate of change parameter, although it was still overestimated with underestimated variability.

Discussion of Simulation Results

The simulation results indicate that ICIRMs have the potential to be a powerful tool for evaluating and developing psychological theory. Even with only a moderately large sample size and a short task, the shape and direction of intratask change were recovered well, which is especially encouraging in light of the complexity of the family of models. However, reliably relating individual differences in intratask change to external variables such as age may require larger sample sizes.

One key finding from the simulations is that the shape of intratask was recovered fairly well. When the true model was the exponential ICIRM, the simulations indicated that the power ICIRM was the optimal model with the exponential ICIRM generally second best. These two models are very closely related and often indistinguishable, so this finding does not create a major concern for the use of ICIRMs. Nonetheless, it is interesting that the incorrect model provides the optimal fit. One possible explanation is that the true model is an exponential ICIRM for each individual, but in essence the model being fit is an average curve with individuals varying around the curve. When exponential curves are averaged, the resulting curve is often more similar to a power curve (Anderson & Tweney, 1997; Heathcote et al., 2000). This possibility creates a minor issue for the empirical data. If individual exponential ICIRMs average to yield a power curve having the optimal fit, then what does it mean for individuals when the average curve yields an exponential ICIRM as the optimal fit? Individuals may adhere to a different non-exponential curve that averages to an exponential curve, although undoubtedly the individual curve is quite similarly shaped to an exponential curve. The precise nature of this individual curve is not currently known and remains a question for future research.

The simulations also indicated that the direction of the intratask change is recovered well, although the total change from initial level was underestimated and the standard deviation overestimated. This suggests that the number of individuals with intratask change in the opposite direction of the typical (i.e., average) individual is overestimated. The distribution of total change based on the WM span results consists of a large proportion with positive intratask change, but also some individual with negative intratask change. Assuming the total change is indeed normally distributed and the estimated mean and SD are correct, 31% of the participants would be expected to have negative intratask change. The simulations indicated that the mean total change was underestimated by a factor of .57 / .71 = .80 and that the standard deviation of the total change was overestimated by 1.96 / 1.45 = 1.35. If the true mean of the total change from initial level is also underestimated by the same factor and the true standard deviation overestimated by the same factor, then the actual proportion of participants with negative intratask change is 20%, substantially lower although still substantial.

Finally, the simulations indicated that the rate of change parameter was not recovered well. This suggests that findings on individual differences in the rate of change parameter, and by extension, intratask change in general, may not be reliable. One potential solution to this problem is to increase sample size. This manipulation led to better estimation of the rate of change, indicating that ICIRMs can yield information on the relation between individual differences in change parameters and external variables.

Discussion of Working Memory Span Results

The goal of the empirical study on working memory span and age was to assess the theory that an age-related increase in PI accounts for the age-related decline in WM span. Previous research has suggested that differential susceptibility to PI may explain half of the age-related decline in WM span (Bowles & Salthouse, 2003b). However, these previous studies did not directly assess the role of PI in WM. Instead, they considered manipulations designed to affect the amount of PI (Bunting, 2006; Emery, 2006; Lustig et al., 2001; May et al., 1999), how external measures of PI relate to WM span (Kane & Engle, 2000; Rosen & Engle, 1998; Whitney et al., 2001), or relative group differences in intratask change (Bowles & Salthouse, 2003b). No previous study has examined WM in a dynamic framework in which individual differences in change across the WM span task could be examined, nor the direction of intratask change. By using an ICIRM that directly examines how the ability to respond correctly to a WM item changes over the course of the task, this study provides the most direct examination of the role of PI in the aging of WM.

Perhaps the most compelling result of the study is that average intratask change was positive. Thus, the results do not support PI as the most important cause of age-related changes in WM span. Although the functional form of intratask change was exponential, consistent with previous research on PI (Underwood, 1957; Wixted & Rohrer, 1993), the direction of change was opposite of expectations. Instead of a reduction in the ability to respond correctly, as would result from the buildup of PI, ability increased over the course of the task. On average, the adults in this sample started the task correctly recalling the average WM span item with probability of about .72, while on the final item, the probability was .79.

Intratask change as strategy production

The likely explanation for positive intratask change involves learning how to most effectively respond to WM span items through the development of more effective strategies for simultaneously processing and storing information. I use the term strategies generically, referring to any aspect of behavior that leads to more successful performance. Several terms for concepts related to individuals' spontaneous (i.e., without training) changes in the use of strategies during memory tasks have been employed; consistent with Dunlosky and Hertzog (1998), I call it the *strategy production hypothesis*.

Under this hypothesis, individual differences in WM span performance are caused at least in part by individual differences in the ability to implement strategies that maximize working memory capabilities.

A related idea is the hypothesis that individual differences in static strategy use may be a source of individual differences in WM span task performance, termed the *strategic allocation hypothesis* by Engle, Cantor, and Carullo (1992). Under this hypothesis, strategy use is considered a stable characteristic of the individual rather than a dynamic aspect of performance. Individual differences in WM span have been shown to be related to (a) individual differences in self-pacing, interpreted as evidence of differences in strategy use (Engle et al., 1992; Friedman & Miyake, 2004); (b) memory strategy training (McNamara & Scott, 2001; Turley-Ames & Whitfield, 2003); and (c) the quality of self-reported strategy use, both retrospectively posttest and after each WM span item (Dunlosky & Kane, 2006). For example, Dunlosky and Kane (2006) found that participants responded correctly to 75% of WM span items when using normatively effective strategies (i.e., strategies that have been shown to be generally effective in memory research, including imagery, sentence generation, and grouping) compared to 60% when using less effective strategies (reading, repetition).

Strategy use as a dynamic aspect of performance has received little attention in research on memory in general. Most research in this area has focused on metacognitive monitoring of strategy effectiveness, rather than changes in the use of strategies themselves. For example, Bieman-Copland and Charness (1994) found that strategy-specific judgments-of-learning more accurately reflected observed learning on the second trial of a memory task than on the first trial, indicating increased awareness of the effectiveness of the strategies. Dunlosky and Hertzog (2000) suggested that there are four aspects of metacognitive strategy monitoring that may influence individual differences in strategy production:

1. Variance in effectiveness of strategies. Strategies must vary in effectiveness in order for strategy production to be effective and useful. Individuals may differ in how effective certain strategies are relative to other strategies. Therefore, strategy production may be particular useful for individuals who could produce a strategy much more effective relative to the strategy currently in use.
2. Monitoring of differential effectiveness. Individuals must monitor strategy usage and effectiveness in order to recognize that strategies differ in effectiveness. Individuals may differ in their ability to do so, and therefore some may not recognize the efficacy of strategy production.
3. Updating of strategy knowledge. Individuals may differ in their ability to dynamically incorporate new information on strategy effectiveness, and may therefore allocate cognitive resources toward less than optimal strategy production.
4. Utilization. Individuals may differ in their utilization of the most effective strategies even if they have equal levels of understanding of the effectiveness. For example, individuals may differ in the efficiency with which they employ the optimal strategy.

Each of these may contribute to the individual differences in WM span, but only knowledge updating and utilization are likely to be expressed dynamically as intratask change. Which of these two is the dominant source of intratask change remains a topic for future research, although some research suggests that utilization may be the more important source of age-related differences (see below).

Strategy use and proactive interference

It is important to note that these results do not preclude PI as an effect. PI and strategy production may both be aspects of intratask change in WM span tasks, yielding change in opposite directions. In fact, given the high variability in the total change from initial level, a number of participants were predicted to have negative intratask change. Based on the simulation results, the proportion of participants with negative intratask change is approximately 20%. Thus, a number of participants are predicted to have negative intratask change, although a strong majority has positive intratask change.

This is not to claim that there are two qualitatively different classes of participants, one group with positive intratask change, one with negative intratask change. Instead, participants likely vary both in their resistance to PI and their ability to develop effective strategies, which implies that some participants who are poor in both may have negative intratask change overall. In this data, the average participant had high enough ability in both domains to have positive intratask change. However, in samples with lower average cognitive abilities, the balance may tilt further toward resistance to PI being the dominant source of intratask change, although even in these samples, many participants would be expected to have positive intratask change.

The two sources of intratask change are likely closely related. One aspect of strategy production may be within-task development of methods for minimizing the effect of PI. That is, a strategy may be aimed at reducing interference rather than increasing memory performance directly, such as by increasing the distinctiveness of information in memory (Bunting, 2006; Nairne, 2002). This hypothesis is supported by findings that interference on cognitive tasks can be minimized through strategy selection (e.g., Long & Prat, 2002). Alternatively, interference may reduce the ability to produce strategies that enable successful memory performance (e.g., Finlay, Hitch, & Meudell, 2000). A third possibility is that strategy production and resistance to PI share a common cause, although the nature of that common cause is not known but may be some sort of fluid ability or executive function. Thus, it is likely that PI and strategy production simultaneously effect WM span performance, although in this data, strategy production appears to be the dominant effect.

Age and intratask change

The empirical results also indicated that there were age-related differences in intratask change. The total change from initial level was positive related to age, indicating that older adults increase in ability more than younger adults. Combined with the nonsignificent negative estimated relation between age and the rate of change, the shape of the intratask change was virtually parallel regardless of age. Thus, individual differences in intratask change do not appear to account for an age-related decline in WM span observed in the data when analyzed in a more standard manner with a no change ICIRM. Instead, the no-change age-related decline was either slightly greater or slightly less than the initial or asymptotic trait level in the exponential ICIRM. This finding is not consistent with previous studies, particularly Bowles and Salthouse (2003b), who found strong age-group differences in intratask change accounting for a substantial portion of the age-related decline in WM span ignoring intratask change.

The finding that the total change from initial level is positively related to age is surprising given the general age-related decline found for almost all cognitive abilities. One explanation is that older adults have greater cognitive plasticity, and therefore can benefit from strategy production more. This hypothesis, however, is contrary to research that indicates that older adults, although having substantial cognitive plasticity, do not have as much plasticity as younger adults (Verhaeghen, Marcoen, & Goossens, 1992), as demonstrated with extensive training programs, which tend to benefit younger adults to a greater extent than older adults (e.g., Kliegl, Smith, & Baltes, 1990; Singer, Lindenberger, & Baltes, 2003). Alternatively, this result can be interpreted as indicating that older adults begin farther from their asymptotic level; that is, their performance is initially further from an optimal level. While perhaps a plausible alternative, there is no theoretical or empirical evidence that this is the case.

The unexpected relations between age and intratask change may be because of characteristics of typical Internet samples. Internet samples tend to consist of relatively able participants, with a stronger selection bias for older adults than for younger adults (Lenhart et al., 2003). Therefore, the older adults in this sample may be relatively more able in comparison to the younger adults, attenuating negative correlations between age and the rate of change, and perhaps leading to a positive relation between age and total change from initial level despite a true 0 or negative relation. This possibility may be reflected in the lower than expected correlation between age and static WM span as indicated by the results from the no change ICIRM.

The characteristics of this sample, however, do not clearly match expectations for Internet samples. Although participants, regardless of age, were more educated than agepeer laboratory participants, the relation between education and age was not different, contrary to the expectation that the relation should be greater, i.e., less negative or more positive. Furthermore, scores on a short vocabulary test were no different from the lab sample, and had the same relation to age. Therefore, it is not clear how these participants compare to typical lab or Internet samples in terms of cognitive ability.

Based on these unclear sampling characteristics, as well as the simulation results, I feel that the results on the relations between age and the individual intratask change parameters should be considered tentative pending replication in additional studies with different sampling characteristics. Combining the finding that average intratask change is positive with the Bowles and Salthouse (2003b) finding that age differences in intratask change are strong and account for about half of the age-related decline in WM span suggests that further studies would likely find that older adults experience smaller intratask change, perhaps due primarily to an age-related slowing in the rate of change. That is, older adults may produce strategies at a slower rate than younger adults, and this difference may account for a substantial portion of the age-related decline in WM span task performance.

A small literature has addressed the role of aging in strategy production in memory performance, although none has examined WM per se. Consensus has not yet been reached, as some researchers have found little effect of age on strategy production (e.g., Blatt-Eisengart & Lachmann, 2004; Dunlosky & Hertzog, 1998), while others have noted the importance of strategy use as an aspect of age-related differences in memory performance (Lachmann & Andreoletti, 2006). These contrary results may stem from differences in aspects of strategy production. In particular, Dunlosky and Hertzog (2000) found no relation between age and metacognitive updating, whereas Rogers, Hertzog and Fisk (2000) and Verhaeghen and Marcoen (1996) found that older adults are less likely to select the optimal strategy and use it efficiently. Thus, it appears that age differences in strategy production may arise from age differences in strategy utilization, perhaps because older adults are less able to gain from strategy usage as optimal strategies are identified.

Comparison to previous studies

Given that strategy production provides a more powerful explanation for intratask change on WM tasks than proactive interference, previous findings that emphasized the role of PI must be reconsidered. Researchers have concluded that individual differences in susceptibility to PI accounts for individual differences in WM span based primarily on three types of evidence: (a) allowing for relative group differences in intratask change due to differential susceptibility to PI reduces individual differences in WM span (Bowles & Salthouse, 2003b); (b) external measures of PI relate to WM span (Kane & Engle, 2000; Rosen & Engle, 1998; Whitney et al., 2001); and (c) manipulations designed to reduce the amount of PI reduce individual differences in WM span (Bunting, 2006; Emery, 2006; Lustig et al., 2001; May et al., 1999). I address each of these in turn.

Bowles and Salthouse (2003b) found that allowing for relative age-group differences in intratask change accounted for about half of the age-related decline in WM span. They interpreted the age-group differences as evidence that older adults are more susceptible to PI than younger adults. Their results were consistent with their conclusion, but they are also consistent with age-group differences in strategy production. That is, as described above, if older adults produce strategies at a slower rate, then relative group differences in intratask change would be observed.

A number of studies have found that external measures of PI are related to performance on WM span tasks (Kane & Engle, 2000; Rosen & Engle, 1998; Whitney et al., 2001). In each study, the relation between susceptibility to PI and WM span was negative; that is, individuals with high WM span also tended to be less susceptible to PI. However, the hypothesized direction of the cause cannot be known (due to the classic correlation vs. causation argument), and the studies differed in the interpretation of the direction. Some researchers interpreted the correlational evidence as indicative that PI directly affects WM span (Whitney et al., 2001), while others suggest a common cause or lean toward higher WM span causing resistance to PI (Kane & Engle, 2000; Rosen & Engle, 1998). This study indicated that, even if PI is a direct cause of individual differences in WM span, the correlation likely overestimates its importance, as susceptibility to PI may be related to strategy production, which is in turn related to WM span. Thus, the correlation between susceptibility to PI and WM span may reflect both a direct effect and an indirect effect through strategy production. The relative size of both effects is not known, although this study suggests that the indirect effect may be substantial.

A third class of studies that has addressed the role of PI in WM span are those that have introduced manipulations to the WM span task that are designed to reduce PI. These manipulations tend to reduce individual differences in WM span. The crucial issue for each of these studies is the validity of the manipulations as affecting PI.

May et al. (1999; see also Emery, 2006; Lustig et al., 2001) introduced two types of manipulations: they administered a working memory span task in descending difficulty order (i.e., greatest item length to least) instead of the more typical ascending order, and they introduced breaks between items involving semantically unrelated cognitive tasks. They claimed that these manipulations reduced the differences in WM span scores between a group of younger adults and a group of older adults. However, their results were substantially more complicated, with neither manipulation alone producing an effect for younger adults and some manipulation conditions yielding decreased performance. Therefore, it is unclear how valid the manipulations were in terms of affecting PI. Furthermore, it is also plausible that the manipulations affected strategy production, as the breaks may have allowed extra time for production.

Bunting (2006) introduced a release-from-PI manipulation, in which the information to be remembered changed, theoretically yielding at least partial release from PI. He implemented two versions of the manipulation, one in which a change from digits to words occurred within each single item (all items were of length 6) and one in which the change between words and digits occurred after every third item of the 12 item task, starting with words. Performance under these manipulations was compared to a control version consisting entirely of words to be remembered. Performance was better for both manipulations. Two aspects of this study are of importance. First, the intraitem manipulation may have affected primacy and recency effects, so the interpretation of the manipulation as purely affecting PI may be incorrect. Second, the intertask manipulation did not appear to counteract an intratask decline, but rather facilitate an intratask increase compared to stable performance for the control task. Thus, it is unclear whether PI was an important aspect of performance on the control task. Instead, the manipulation may have facilitated strategy production, perhaps particularly for the digits which were not part of the control task.

Thus, for each of the studies that have emphasized the role of PI in WM span, strategy production may have been an equally important or more important aspect of performance. This has important implications for understanding the role of WM in other forms of cognition. Consider, for example, the relation between WM span and reasoning ability. Two recent studies addressed this relation and found that items with the most PI had the largest correlation with reasoning ability (Bunting, 2006; Emery, 2006). Items interpreted as having high PI may instead have had the greatest level of strategy production, and strategy production would seem to be an important aspect of reasoning ability. Thus, the correlational findings are readily explainable in terms of strategy production. Further research is needed to confirm this hypothesis.

Definition of working memory span

An interesting issue that this dissertation brings up is what is meant by WM span. These results indicate that performance on tasks designed to measure WM span is not static. Instead, the ability to respond correctly changes over the course of the task. Therefore, WM span defined as the ability to respond correctly cannot be a stable trait of an individual, despite being generally conceptualized as a static capacity. Individual differences in capacity are clearly relevant in these results, as there were substantial individual differences, including age-related declines, in both the initial and asymptotic trait levels. Either of these could be the more accurate reflection of the construct of WM span as psychologists conceptualize it. WM span as an individual differences concept may be an initial level concept, in that it is a trait that reflects WM capacity independent of non-capacity effects such as strategy production or PI. Alternatively, WM span may be an asymptotic level concept, in that it reflects individual differences in WM capacity remaining after all non-capacity effects have reached their full impact. No answer to this issue is apparent in psychological theory. Future theoretical considerations of WM should be more precise on what is meant by WM span, or the idea of WM span as a static concept should be eliminated in favor of multiple concepts of individual differences in dynamic aspects of performance on WM tasks.

Future Studies

This research raises three major concerns for future studies. First, the generalizability of the sample should be examined, by comparing these findings with those from previous research studies. In this study, I was not able to deal completely with selection biases; the sample shared some characteristics with typical laboratory samples, but did not share other characteristics. It will be useful to replicate this study in a typical laboratory setting.

Second, the task design could be adjusted to more fully assess how well participants maintained experimental protocols. One common practice in Internet-based studies is to collect and save the computer (IP) addresses used by the participants in order to determine whether participants have taken the task more than once. If more than one participants uses the same IP address, then it is possible that those participants are in reality the same person participating multiple times. This issue was partially addressed in this study with the pretest question asking whether about previous participation. However, that response requires honesty and accuracy from the participant, whereas the

IP address is automatically saved without participant intervention. The IP address is not foolproof, as multiple participants could use the same computer or a single participant could use multiple computers,. The IP address provides a second layer of control, but other mechanisms need to be worked out.

A third issue is the question on whether the participant was able to concentrate for the entire experiment. This question was designed to address whether participants were able to release proactive interference during the task. However, the results suggest that strategy production is a more important aspect of performance than PI. Furthermore, the results did not differ substantially when I excluded participants who said they were not able to concentrate (21% of the full sample). This suggests that concentration as indicated by the posttest question does not appear to be an accurate gauge of experimental protocol maintenance, as strategy production seems to have occurred regardless of self-reported concentration. A revised posttest questionnaire with one or more rephrased questions on concentration may more effectively differentiate between participants who were able to produce effective strategies despite a lack of full concentration from those who concentrated so little that the task was not a relevant measure of WM span and strategy production.

Conclusions

Although many psychological theories imply intratask change, few statistical models have been developed to directly analyze intratask change. These few intratask change models have a long history in psychology, particularly in learning research, but unrealistic assumptions and challenges in estimation have minimized their usefulness despite their importance for psychological research. In this dissertation, I developed intratask change item response models (ICIRMs), a family of item response models that incorporate change functions, so that the shape, direction, and magnitude of intratask change can be assessed, as well as individual differences in the change parameters. Simulations indicated that these models can provide an effective means to analyze intratask change.

I applied these models to test a theory that predicts that, due to proactive interference, the ability to respond correctly to a working memory span task should go down over the course of the task. Furthermore, intratask change is predicted to be stronger, i.e., more negative, for older adults, accounting for at least some of the age-related decline in WM span. Because of the need for randomized order of presentation in order to separate intratask change from change in item difficulty, I collected new working memory span data over the internet. The analysis of these data indicated that the intratask change followed an exponential change model, and that, contrary to predictions, the average intratask change was positive. Older adults had greater total change from initial level, but combined with a nonsignificant but slower growth rate, there were no apparent age difference in intratask change. However, the simulations, which indicated that the individual differences in change parameters were not estimated well, and the sampling characteristics together suggest that the results on the age relations should be considered tentative pending replication.

The findings suggest that strategy production offers a more powerful explanation for intratask change in working memory span tasks than proactive interference. Strategy production refers to the within task development and use of strategies for maximizing working memory capabilities. These results, together with previous results on the relation between age and working memory, suggest that age differences in strategy production may account for some of the age-related decline in working memory span. I conclude by noting that this dissertation highlights the importance of models for intratask change to gather fresh and more valid insights into the nature of working memory and the role of strategy production, and that these models should be considered for many psychological theories that imply intratask change.

References

Ackerman, T., & Spray, J. A. (1987). *A general model for item dependency* [ACT Research Report 87-9]. Iowa City: ACT.

Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Holt, Rinehart, and Winston.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csáki (Eds.), *Proceedings of the 2nd International Symposium on Information Theory* (pp. 267-281). Budapest, Hungary: Akadémiai Kiadó.

Anderson, R. B., & Tweney, R. D. (1997). Artifactual power curves in forgetting. *Memory and Cognition, 25,* 724-730.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89-95). New York: Academic Press.

Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-89). New York: Academic Press.

Bailey, R. D., Foote, W. E., & Throckmorton, B. (2000). Human sexual behavior: A comparison of college and Internet surveys. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 141-168). New York: Academic Press.

Bieman-Copland, S., & Charness, N. (1994). Memory knowledge and memory monitoring in adulthood. *Psychology and Aging, 9,* 287-302.

Birnbaum, M. H. (2000a). Decision making in the lab and on the web. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 3-34). New York: Academic Press.

Birnbaum, M. H. (Ed.). (2000b). *Psychological experiments on the Internet*. New York: Academic Press.

Blatt-Eisengart, I., & Lachman, M. E. (2004). Attributions for memory performance in adulthood: Age differences and mediation effects. *Aging, Neuropsychology, and Cognition*, *11*, 68-79.

Böckenholt, U. (1993). A latent class regression approach for the analysis of recurrent choice data. *British Journal of Mathematical and Statistical Psychology, 46,* 95-118.

Böckenholt, U. (2005). A latent Markov model for the analysis of longitudinal data collected in continuous time: States, durations, and transitions. *Psychological Methods, 10,* 65-83.

Bond, T. G. & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences.* Mahwah, NJ: Erlbaum.

Bonney, G. E. (1987). Logistic regression for dependent binary observations. *Biometrics, 43,* 951-973.

Bowles, R.P. & Grimm, K.J. (2006). *Local person dependence and the measurement of change.* Manuscript in preparation.

Bowles, R. P., & Salthouse, T. A. (2003a, August). *Age relations and processing components of synonyms and antonyms*. Poster presented at the Annual Meeting of the American Psychological Association, Toronto.

Bowles, R. P., & Salthouse, T. A. (2003b). Assessing the age-related effects of proactive interference on working memory using the Rasch model. *Psychology and Aging, 18,* 608-615.

Bunting, M. F. (2006). Proactive interference and item similarity in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 183-196.

Burke, D. M. (1997). Language, aging, and inhibitory deficits: Evaluation of a theory. *Journal of Gerontology: Psychological Sciences, 52,* P254-264.

Bush, R. R., & Mosteller, F. (1951). A mathematical model for simple learning. *Psychological Review, 58,* 313-323.

Bush, R. R., & Mosteller, F. (1955). *Stochastic models for learning.* New York: Wiley.

Bush, R. R., Mosteller, F., & Thompson, G. L. (1954). A formal structure for multiple-choice situations. In R. M. Thral, C. H. Coombs, & R. L. Davis (Eds.), *Decision processes* (pp. 99-126). New York: Wiley

Chen, W.H. & Thissen, D. (1997). Local dependence indices for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, *22*, 265-289.

Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika, 40,* 5-32.

Collins, L. M., & Sayer, A. G. (Eds.). (2001). *New methods for the analysis of change.* Washington, DC: American Psychological Association.

Conway, A. R. A., & Engle, R. W. (1994). Working memory and retrieval: A resource-dependent inhibition model. *Journal of Experimental Psychology: General, 123,* 354-373.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin and Review, 12,* 769-786.

Cowan, N. (1995). *Attention and memory: An integrated framework.* New York: Oxford University Press.

Cudeck, R., & Klebe, K. J. (2002). Multiphase mixed-effects models for repeated measures data. *Psychological Methods, 7,* 41-63.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450-466.

Daneman, M., & Merikle, P. M. (1996). Working memory and comprehension: A meta-analysis. *Psychonomic Bulletin and Review, 3,* 422-433.

Dawson, T. L. (2000). Moral and evaluative reasoning across the life-span. *Journal of Applied Measurement, 1,* 346-371.

Dunlosky, J., & Hertzog, C. (1998). Aging and deficits in associative memory: What is the role of strategy production? *Psychology and Aging, 13,* 597-607.

Dunlosky, J., & Hertzog, C. (2000). Updating knowledge about encoding strategies: A componential analysis of learning about strategy effectiveness from task experience. *Psychology and Aging, 15,* 462-474.

Dunlosky, J., & Kane, M. J. (2006). *The contributions of strategy use to working memory span: A comparison of strategy-assessment methods.* Manuscript submitted for publication.

Embretson, S. E. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika, 56,* 495-516.

Embretson, S. E. (1995). A measurement model for linking individual change to processes and knowledge: Application to mathematical learning. *Journal of Educational Measurement, 32,* 277-294.

Embretson, S. E., & Schmidt McCollam, K. M. (2000). A multicomponent Rasch model for measuring covert processes: Application to lifespan ability changes. In M. Wilson & J. Englehard, George (Eds.), *Objective measurement: Theory into practice* (Vol. 5, pp. 203-218). Norwood, NJ: Ablex.

Emery, L. (2006, April). *Individual and age differences in proactive interference and working memory.* Poster presented at the Cognitive Aging Conference, Atlanta, GA.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19,* 972-992.

Engle, R. W., Carullo, J. J., & Collins, K. W. (1991). Individual differences in working memory for comprehension and following directions. *Journal of Educational Research, 84*, 253-262.

Engle, R. W., Conway, A. R. A., Tuholski, S. W., & Shisler, R. J. (1995). A resource account of inhibition. *Psychological Science, 6,* 122-125.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General, 128,* 309-331.

Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological Review, 57,* 84-107.

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin, 53*, 134-140.

Fahrmeir, L., & Tutz, G. (2001). *Multivariate statistical modeling based on generalized linear models* (2nd ed.). New York: Springer.

Ferrer, E., & McArdle, J. J. (2004). An experimental analysis of dynamic hypotheses about cognitive abilities and achievement from childhood to early adulthood. *Developmental Psychology, 40,* 935-952.

Ferrer, E., Salthouse, T. A., Stewart, W. F., & Schwartz, B. S. (2004). Modeling age and retest processes in longitudinal studies of cognitive abilities. *Psychology and Aging, 19,* 243-259.

Finlay, F., Hitch, G. J., & Meudell, P. R. (2000). Mutual inhibition in collaborative recall: Evidence for a retrieval-based account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1556-1567.

Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica, 37*, 359-374.

Fischer, G. H. (1976). Some probabilistic models for measuring change. In D. N. M. Dr Gruijter & L. J. T. Van der Kemp (Eds.), *Advances in psychological and educational measurement* (pp. 97-110). New York: Wiley.

Fischer, G. H. (1977). Some probabilistic models for the description of attitudinal and behavioral changes under the influence of mass communication. In W. F. Kempf & B. H. Repp (Eds.), *Mathematical models for social psychology* (pp. 102-151). Berne: Huber.

Fischer, G. H. (1989). An IRT-based model for dichotomous longitudinal data. *Psychometrika, 54,* 599-624.

Fischer, G. H. (1995). Linear logistic models for change. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 157-180). New York: Springer.

Frankel, M. S., & Siang, S. (1999, November). *Ethical and legal aspects on human subjects research on the Internet*. Paper presented at the Workshop report for the American Association for the Advancement of Science, Washington, DC.

Friedman, N. P., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language, 51,* 136-158.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 6,* 721-741.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? *American Psychologist, 59,* 93-104.

Hambrick, D. Z., & Engle, R. W. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-is-power hypothesis. *Cognitive Psychology, 44*, 339-387.

Hasher, L., & Zacks, R. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 193-226). New York: Academic Press.

Hasher, L., Zacks, R., & May, C. P. (1999). Inhibitory control, circadian arousal, and age. In D. Gopher (Ed.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 653-675). Cambridge, MA: MIT Press.

Heathcote, A., Brown, S. & Mewhort, D. J. K. (2000) The Power Law repealed: The case for an exponential law of practice. *Psychonomic Bulletin and Review,7*, 185-207.

Hebb, D. O. (1949). *The organization of behavior*. Wiley, New York.

Hoskens, M., & De Boeck, P. (1997). A parametric model for local dependence among test items. *Psychological Methods, 2*, 261-277.

Humphreys, K. (1998). The latent Markov chain with multivariate random effects. *Sociological Methods and Research, 26,* 269-299.

Jannarone, R. J. (1986). Conjunctive item response theory kernels. *Psychometrika, 51,* 357-373.

Jansen, M. G. H. (1997). Applications of Rasch's Poisson counts model to longitudinal count data. In J. Rost & R. Langeheine (Eds.), *Applications of latent trait and latent class models in the social sciences* (pp. 380-389). Munster, Germany: Verlag.

Kamata, A. (2001). Item analysis by the hierarchical generalized linear model. *Journal of Educational Measurement, 38,* 79-93.

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 336-358.

Kempf, W. F. (1977). A dynamic test model and its use in the microevaluation of instructional material. In H. Spada & W. F. Kempf (Eds.), *Structural models for thinking and learning* (pp. 295-318). Bern: Huber.

Keppel, G., Postman, L., & Zavortink, B. (1968). Studies of learning to learn: VIII. The influence of massive amounts of training upon the learning and retention of paired-associate lists. *Journal of Verbal Learning and Verbal Behavior, 7*, 790-796.

Keppel, G., & Underwood, B. J. (1962). Proactive interference in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior, 1*, 153-161.

Klauer, K. C., & Sydow, H. (2001). Modeling learning in short-term learning tests. In A. Boomsma, M. A. J. van Duijn, & T. A. B. Snijders (Eds.), *Essays on item response theory*. New York: Springer.

Kliegl, R., Smith, J., & Baltes, P. B. (1990). On the locus and process of magnification of age differences during mnemonic training. *Development Psychology, 26,* 894-904.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working memory capacity?! *Intelligence, 14*, 389-433.

Lachman, M. E., & Andreoletti, C. (2006). Strategy use mediates the relationship between control beliefs and memory performance for middle-aged and older adults. *Journal of Gerontology: Psychological Sciences, 61,* P88-P94.

Lee, O. K. (2003). Rasch simultaneous vertical equating for measuring reading growth. *Journal of Applied Measurement, 4*, 10-23.

Lenhart, A., Horrigan, J., Rainie, L., Allen, K., Boyce, A., Madden, M., & O'Grady, E. (2003). *The ever-shifting Internet population.* Retrieved May 1, 2006 from Pew Internet and American Life Project Web site: www.pewinternet.org/pdfs/PIP_Shifting_Net_Pop_Report.pdf

Liang, K.-Y., & Zeger, S. L. (1989). A class of logistic regression models for multivariabe binary time series. *Journal of the American Statistical Association, 84,* 447-451.

Light, L. L. (1991). Memory and aging: Four hypotheses in search of data. *Annual Review of Psychology, 42,* 333-376.

Long, D. L., & Prat, C. S. (2002). Working memory and Stroop interference: An individual differences inverstigation. *Memory and Cognition, 32,* 592-601.

Luce, R. D. (1959). *Individual choice behavior.* New York: Wiley.

Lustig, C., & Hasher, L. (2002). Working memory span: The effect of prior learning. *American Journal of Psychology, 115*, 89-101.

Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General, 130*, 199-207.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 60,* 523-547.

May, C. P., Hasher, L., & Kane, M. J. (1999). The role of interference in memory span. *Memory and Cognition, 27*, 759-767.

McArdle, J. J. (1986). Latent growth within behavior genetic models. *Behavior Genetics, 16,* 163-200.

McArdle, J. J., & Anderson, E. (1990). Latent variable growth models for research on aging. In J. E. Birrwn & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (3rd ed., ppp. 21-43). New York: Plenum.

McArdle, J. J. & Epstein, D. B. (1987). Latent growth curves within developmental structural equation models. *Child Development, 58,* 110-133.

McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology, 38*, 115-142.

McArdle, J. J., Grimm, K. J., Hamagami, F., Bowles, R. P., & Meredith, W. (2006). *Modeling latent growth curves using longitudinal data with non-repeated measurements*. Manuscript submitted for publication.

McArdle, J. J., & Hamagami, F. (2004). Methods for dynamic change hypotheses. In K. van Montfort, J. Oud, & A. Satorra (Eds.), Recent developments in structural equation models (pp. 295-335). Dordrecht: Kluwer.

McArdle, J. J., & Nesselroade, J. R. (2002). Growth curve analysis in comtemporary psychological research. In J. Schinka & W. Velicer (Eds.), *Comprehensive handbook of psychology, Volume Two: Research methods in psychology.* New York: Wiley.

McArdle, J. J., & Woodcock, J. R. (1997). Expanding test-retest designs to include developmental time-lag components. *Psychological Methods, 2* 403-435.

McCollam (Schmidt), K. M. (1997). *The modifiability of age differences in spatial visualization*. Unpublished doctoral dissertation, University of Kansas.

McDowd, J. M. (1997). Inhibition in attention and aging. *Journal of Gerontology: Psychological Sciences, 52,* P265-P273.

McGraw, K. O., Tew, M. D., & Williams, J. E. (2000). PsychExps: An on-line psychology laboratory. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 219-234) New York: Academic Press.

McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. *Memory and Cognition, 29,* 10-17.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55,* 107-122.

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control.* New York: Cambridge University Press.

Mogle, J. (2006). *Cognitive research over the Internet: A validation of performance-based measures.* Unpublished master's thesis, Syracuse University.

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary research and perspectives, 2*, 201-218.

Molenberghs, G., & Verbeke, G. (2004). An introduction to (generalized (non)linear mixed models. In P. De Boeck & M. Wilson (Eds.), *Explanatory item response models* (pp. 111-153). New York: Springer.

Moser, E. B. (2006). MVNGen.sas - Generate multivariate normal data [Computer software]. Retrieved from http://www.stat.lsu.edu/faculty/moser/exst7037/MVNGen.html

Moskowitz, D. S., & Hershberger, S. L. (Eds.). (2002). *Modeling intraindividual variaibility with repeated measures data.* Mahwah, NJ: Erlbaum.

Musch, J., & Reips, U.-D. (2000). A brief history of web experimenting. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 61-88). New York: Academic Press.

Muthén, B. O. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49,* 115-132.

Muthén, B. O. (1996). Growth modeling with binary responses. In A. von Eye & C. C. Clogg (Eds.), *Categorical variables in developmental research.* New York: Academic Press.

Muthén, B. O., & Muthén, L. K. (2004). Mplus (Version 3.01). Los Angeles: Muthén & Muthén.

Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology, 53,* 53-81.

Nesselroade, J. R., & Molenaar, P. C. M. (1999). Pooling lagged covariance structures based on short, multivariate time-series for dynamic factor analysis. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 223-250). Newbury Park, CA: Sage Publications.

Nesselroade, J. R., & Schmidt McCollam, K. M. (2000). Putting the process in developmental processes. *International Journal of Behavioral Development, 24,* 295-300.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1-55). Hillsdale, NJ: Erlbaum.

Nosek, B., Banaji, M. R., & Greenwald, A. G. (2002a). E-Research: Ethics, security, design, and control in psychological research on the Internet. *Journal of Social Issues, 58*, 161-176.

Nosek, B., Banaji, M. R., & Greenwald, A. G. (2002b). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practive, 6*, 101-115.

Patz, R. J., & Junker, B. W. (1999). Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses. *Journal of Educational and Behavioral Statistics, 24,* 342-366.

Ram, N., Chow, S.-M., Bowles, R. P., Wang, L., Grimm, K. J., Fujita, F., & Nesselroade, J. R. (2005). Examining interindividual differences in cyclicity of pleasant and unpleasant affect using spectral analysis and item response modeling. *Psychometrika, 70,* 773-790.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainments tests* (expanded edition). Chicago: University of Chicago Press. (Original work published 1960).

Reips, U.-D. (2000). The web experiment method: Advantages, disadvantages, and solutions. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 89-120). New York: Academic Press.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64-99). New York: Appleton-Century-Crofts.

Rijmen, F., De Boeck, P., & van der Maas, H. L. J. (2005). An IRT model with a parameter-driven process for change. *Psychometrika, 70,* 651-669.

Rogers, W. A., Hertzog, C., & Fisk, A. D. (2000). Age-related differences in associative learning: An individual differences analysis of ability and strategy influences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 359-394.

Rosen, V. M., & Engle, R. W. (1998). Working memory capacity and suppression. *Journal of Memory and Language, 39*, 418-436.

Rost, J. (2001). The growing family of Rasch models. In A. Boomsma, M. A. J. van Duijn, & T. A. B. Snijders (Eds.), *Essays on item response theory* (pp. 2-42). New York: Springer.

Salthouse, T. A. (1993a). Influence of working memory on adult age differences in matrix reasoning. *British Journal of Psychology, 84*, 171-199.

Salthouse, T. A. (1993b). Speed and knowledge as determinants of adult age differences in verbal tasks. *Journal of Gerontology: Psychological Sciences, 48*, P29-P36.

Salthouse, T. A. (1994a). The aging of working memory. *Neuropsychology, 8*, 535-543.

Salthouse, T. A. (1994b). How many causes are there of aging-related decrements in cognitive functioning? *Developmental Review, 14*, 413-437.

Salthouse, T. A. (1996). The process-speed theory of adult age differences in cognition. *Psychological Review, 103,* 403-428.

Salthouse, T. A., Babcock, R. L., Mitchell, D. R. D., Palmon, R., & Skovronek, E. (1990). Sources of individual differences in spatial visualization ability. *Intelligence, 14*, 187-230.

Salthouse, T.A., Schroeder, D.H., & Ferrer, E. (2004). Estimating retest effects in longitudinal assessments of cognitive functioning in adults between 18 and 60 years of age. *Developmental Psychology, 40,* 813-822.

Schneider, W., & Detweiller, M. (1987). A connectionist/control architecture for working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 21, pp. 53-119). New York: Academic Press.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461-464.

Singer, J. D., & Willett, J. R. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* New York: Oxford University Press.

Singer, T., Lindenberger, U., & Baltes, P. B. (2003). Plasticity of memory for new learning in very old age: A story of major loss? *Psychology and Aging, 18,* 306-317.

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis.* New York: Appleton- Century.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling.* New York: Chapman and Hall.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B. 64,* 583-630.

Spiegelhalter, D. J., Thomas, A., Best, N. G., & Lunn, D. (2003). WinBUGS version 1.4.1 [Computer software manual]. Retrieved May 24, 2005 from http://www.mrc-bsu.cam.ac.uk/bugs.

Sternberg, S. H. (1963). Stochastic learning theory. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology, Vol. II* (pp. 1-120). New York: Wiley.

Stoltzfus, E. R., Hasher, L., & Zacks, R. (1996). Working memory and aging: Current status of the inhibitory view. In J. T. E. Richardson & R. W. Engle & L. Hasher & R. H. Logie & E. R. Stoltzfus & R. Zacks (Eds.), *Working memory and human cognition* (pp. 66-88). New York: Oxford University Press.

Süβ, H., Oberauer, K., Wittmann, W., W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability- and a little bit more. *Intelligence, 30*, 261-288.

Takane, Y., & de Leeuw (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika, 52,* 393-408.

Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *Psychological Review Monograph Supplement, 2*.

Thorndike, E. L. (1932). *The fundamentals of learning*. New York: Teachers College Press.

Tuerlinckx, F., & De Boeck, P. (2004). Models for residual dependencies. In P. De Boeck & M. Wilson (Eds.), *Explanatory item response models* (pp. 289-316). New York: Springer.

Turley-Ames, K. J., & Whitfield, M. M. (2003). Strategy training and working memory task performance, *Journal of Memory and Language, 49,* 446-468.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28,* 127-154.

U. S. Census Bureau (2000). *Census 2000 summary file 1*. Retrieved May 25, 2005 from factfinder.census.gov

Underwood, B. J. (1957). Interference and forgetting. *Psychological Review, 64*, 49-60.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37,* 498-505.

Verguts, T., & De Boeck, P. (2000). A Rasch model for detecting learning while solving an intelligence test. *Applied Psychological Measurement, 24,* 151-162.

Verhaeghen, P., Marcoen, P., & Goossens, L. (1992). Improving memory performance in the aged through mnemonic training: A meta-analytic study. *Psychology and Aging, 7,* 242-251.

Verhaeghen, P., & Marcoen, A. (1996). On the mechanisms of plasticity in young and older adults after instruction in the method of loci: Evidence for an amplification model. *Psychology and Aging*, *11*, 164-178.

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

Verhaeghen, P., & Salthouse, T. A. (1997). Meta-analyses of age-cognition relations in adulthood: Estimates of linear and nonlinear age effects and structural models. *Psychological Bulletin, 122*, 231-249.

Verhelst, N. D., & Glas, C. A. W. (1993). A dynamic generalization of the Rasch model. *Psychometrika, 58,* 395-425.

Verhelst, N. D., & Glas, C. A. W. (1995). Dynamic generalizations of the Rasch model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 181-201). New York: Springer.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika, 54,* 427-450.

Waters, G. S., & Caplan, D. (1996). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review, 103,* 761-772.

Watson, J. B. (1914). *Behavior: An introduction to comparative psychology*. New York: Holt.

Whitney, P., Arnett, P. A., Driver, A., & Budd, D. (2001). Measuring central executive functioning: What's in a reading span? *Brain and Cognition, 45*, 1-14.

Wickens, D. D., Born, D. G., & Allen, C. K. (1963). Proactive inhibition and item similarity in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 2*, 440-445.

Wilson, M. (1989). Saltus: A psychometric model of discontinuity in cognitive development. *Psychological Bulletin, 105,* 276-289.

Wilson, M., & De Boeck, P. (2004). Descriptive and explanatory item response models. In P. De Boeck & M. Wilson (Eds.), *Explanatory item response models* (pp. 43-74). New York: Springer.

Wixted, J. T., & Rohrer, D. (1993). Proactive interference and the dynamics of free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1024-1039.

Zacks, R. T., Hasher, L., & Li, K. Z. H. (2000). Human memory. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed., pp. 293-358). Mahwah, NJ: Erlbaum.

Zhang, Z., McArdle, J. J., Wang, L., & Hamagami, F. (2006). *Using WinBUGS from SAS for Bayesian analysis: With an example of the growth curve model.* Manuscript in preparation.

Appendix A

Estimation of ICIRMs using WinBUGS

ICIRMs for which the change process can be expressed as a linear combination of parameters, which include the no change, linear change, and quadratic change models, are members of the family of generalized linear models (Fahrmeir & Tutz, 2001; Skrondal & Rabe-Hesketh, 2004). As a result, these models can be estimated effectively with a number of estimation techniques and programs (Molenberghs & Verbeke, 2004). More complex nonlinear change models, however, may not be estimable with many programs. Therefore, a general estimation program that allows for a wide variety of complex models is needed.

For all analyses, I employed WinBUGS (Spiegelhalter, Thomas, Best, & Lunn, 2003), an open source freeware general estimation program available on-line. WinBUGS implements Monte Carlo Markov chain (MCMC) estimation using Gibbs sampling (Geman & Geman, 1984), a Bayesian approach. With Bayesian estimation, a prior distribution on the parameters must be given. However, when non-informative priors are selected, the estimation is virtually equivalent to maximum likelihood. The estimation works as follows:

1. Starting values for all parameters are selected, $\Theta^{[0]} = [\theta_1^{[0]}, \theta_2^{[0]}, \theta_3^{[0]}, ..., \theta_k^{[0]}]$, where the superscript indicates the iteration.
2. Select a value for $\Theta^{[1]}$ at random from the k conditional distributions:

   $\theta_1^{[1]} \sim f(\theta_1 \mid \theta_2^{[0]}, \theta_3^{[0]}, ..., \theta_k^{[0]})$

$\theta_2^{[1]} \sim f(\theta_2 \mid \theta_1^{[1]}, \theta_3^{[0]}, ..., \theta_k^{[0]})$

…

$\theta_{k-1}^{[1]} \sim f(\theta_{k-1} \mid \theta_2^{[1]}, \theta_2^{[1]}, \theta_3^{[1]}, ..., \theta_k^{[0]})$

$\theta_k^{[1]} \sim f(\theta_k \mid \theta_2^{[1]}, \theta_2^{[1]}, \theta_3^{[1]}, ..., \theta_{k-1}^{[1]})$

3. Repeat step 2 until convergence is achieved.
4. After convergence is reached, repeat step 2 to get an estimate of the posterior distribution of Θ, with the mean of the posterior distributions used as estimates of the individual parameters.

MCMC estimation has been used for a number of IRMs (Patz & Junker, 1999), including dynamic across-task IRMs (Ram et al., 2005).

Appendix B

WinBUGS Script for Exponential ICIRM

```
model{
	for (n in 1:260) {
		for (t in 1:15) {
p[n,t]<-phi(1/1.7*(thetaL[n]+thetaS[n]*exp(thetar[n]*(t-1))-betaa*itema[n,t]-
betab*itemb[n,t]-betac*itemc[n,t]-betad*itemd[n,t]-betaf*itemf[n,t]-betag*itemg[n,t]))
			itemscorec[n,t] ~ dbern(p[n,t])
			}
		thetaL[n] ~ dnorm(mu_thetaL,tau_thetaL)
		thetaS[n] ~ dnorm(mu_thetaS,tau_thetaS)
		thetar[n] ~ dnorm(mu_thetar,tau_thetar)
		}

#priors

	betaa ~ dnorm(0,1.0E-6)
	betab ~ dnorm(0,1.0E-6)
	betac ~ dnorm(0,1.0E-6)
	betad ~ dnorm(0,1.0E-6)
	betaf ~ dnorm(0,1.0E-6)
	betag ~ dnorm(0,1.0E-6)

	mu_thetaL ~ dnorm(0,1.0E-6)
	tau_thetaL ~ dgamma(0.01,0.01)
	mu_thetaS ~ dnorm(0,1.0E-6)
	tau_thetaS ~ dgamma(0.01,0.01)
	mu_thetar ~ dnorm(0,1.0E-6)
	tau_thetar ~ dgamma(0.01,0.01)
	sig_thetaL <-1/tau_thetaL
	sig_thetaS <-1/tau_thetaS
	sig_thetar <-1/tau_thetar
}
```

Appendix C

SAS script for simulations with exponential ICIRM

```
libname sim 'C:\Documents and Settings\rpb3b\My Documents\dissertation\simulations';

%macro exp_simulation(numiterations);
	%global A;

	data sim.exp_simulation_results2;
	run;

	%do iteration=1 %to &numiterations;

proc IML;
	Reset NoLog NoPrint;
	n=260;
	R={1 -.36 .20,-.36 1 -.21,.20 -.21 1};
	Ds=Diag({1.14 1.07 .31});
	S=Ds*R*Ds;
	T=Root(S);
	d=NRow(S);
	X=J(n,d,0);
	X=Rannor(X);
	Y=X*T;
	Create NormalData From Y;
	Append From Y;
	Close NormalData;
quit;

data exp_sim;
	set NormalData;
	array itema{15} itema1-itema15;
	array itemb{15} itemb1-itemb15;
	array itemc{15} itemc1-itemc15;
	array itemd{15} itemd1-itemd15;
	array iteme{15} iteme1-iteme15;
	array itemf{15} itemf1-itemf15;
	array itemg{15} itemg1-itemg15;
	array itemscorec{15} itemscorec1-itemscorec15;

	thetaL=col1+1.15;
	thetaS=col2-.71;
```

```
          thetar=col3+.38;

          betaa=-3.2;
          betab=-2.3;
          betac=-1.9;
          betad=-1.1;
          betae=0;
          betaf=1.0;
          betag=2.3;

   *generate raw data;
   do time=1 to 15;
                    efftheta=thetaL+thetaS*exp(-1*thetar*(time-1));
                    check=ranuni(0);
                    itema{time}=0; if check GE 0 and check LT 1/7 then itema{time}=1;
                    itemb{time}=0; if check GE 1/7 and check LT 2/7 then itemb{time}=1;
                    itemc{time}=0; if check GE 2/7 and check LT 3/7 then itemc{time}=1;
                    itemd{time}=0; if check GE 3/7 and check LT 4/7 then itemd{time}=1;
                    iteme{time}=0; if check GE 4/7 and check LT 5/7 then iteme{time}=1;
                    itemf{time}=0; if check GE 5/7 and check LT 6/7 then itemf{time}=1;
                    itemg{time}=0; if check GE 6/7 and check LT 1 then itemg{time}=1;
                    expon=efftheta-
(betaa*itema{time}+betab*itemb{time}+betac*itemc{time}+betad*itemd{time}+betae*i
teme{time}+betaf*itemf{time}+betag*itemg{time});
                    p=exp(expon)/(1+exp(expon));
                    check2=ranuni(0);
                    if check2<p then itemscorec{time}=1;
                    else itemscorec{time}=0;
                    keep thetaL thetaS thetar itema1-itema15 itemb1-itemb15 itemc1-itemc15
itemd1-itemd15 iteme1-iteme15 itemf1-itemf15 itemg1-itemg15 itemscorec1-
itemscorec15;
          end;
          output;
run;

*data;
%_sexport(data=exp_sim,
          file='C:/Documents and Settings/rpb3b/My
Documents/dissertation/simulations/itemscorec_data.txt',
          var=itema1-itema15 itemb1-itemb15 itemc1-itemc15 itemd1-itemd15 itemf1-
itemf15 itemg1-itemg15 itemscorec1-itemscorec15);

*starting values;
data _NULL_;
          file 'C:/Documents and Settings/rpb3b/My
Documents/dissertation/simulations/exp_parameterstartvalues.txt';
```

```
	put 'list(mu_thetaL=1, mu_thetaS=-1, mu_thetar=-.5, tau_thetaL=1, tau_thetaS=1,
tau_thetar=5, betaa=-3.3, betab=-2.4, betac=-2.0, betad=-1.1, betaf=1.1, betag=2.3)';
run;

*batch file;
data _NULL_;
	filename script 'C:\Program Files\WinBUGS14-2\exp_sim_batch.txt';
	file script;
	put // @@
	#1 "display('log')"
	#2 "check('C:/Documents and Settings/rpb3b/My
Documents/dissertation/simulations/exp_script.txt')"
	#3 "data('C:/Documents and Settings/rpb3b/My
Documents/dissertation/simulations/itemscorec_data.txt')"
	#4 "compile(1)"
	#5 "inits(1,'C:/Documents and Settings/rpb3b/My
Documents/dissertation/simulations/exp_parameterstartvalues.txt')"
	#6 "gen.inits()"
	#7 "update(5000)"
	#8 "set(betaa)"
	#9 "set(betab)"
	#10 "set(betac)"
	#11 "set(betad)"
	#12 "set(betaf)"
	#13 "set(betag)"
	#14 "set(mu_thetaL)"
	#15 "set(sig_thetaL)"
	#16 "set(mu_thetaS)"
	#17 "set(sig_thetaS)"
	#18 "set(mu_thetar)"
	#19 "set(sig_thetar)"
	#40 "dic.set()"
	#41 "update(10000)"
	#42 "dic.stats()"
	#43 "coda(*,'C:/Documents and Settings/rpb3b/My
Documents/dissertation/simulations/exp_output')"
	#44 "save('C:/Documents and Settings/rpb3b/My
Documents/dissertation/simulations/exp_bugslog.txt')"
	#45 "quit()"
	;
run;

*run script;
data _NULL_;
	file 'C:\run.bat';
	put //@@
```

```
        #1 '"C:\Program Files\WinBUGS14-2\winbugs14.exe" /PAR exp_sim_batch.txt'
        #2 'exit'
        ;
run;

data _NULL_;
        X "C:\run.bat";
run;

%coda2sas(out=exp_results,
infile='C:\Documents and Settings\rpb3b\My
Documents\dissertation\simulations\exp_outputIndex.txt',
chain='C:\Documents and Settings\rpb3b\My
Documents\dissertation\simulations\exp_output1.txt');
quit;

*means;
proc means data=exp_results;
        output out=exp_means;
run;

data sim.exp_simulation_results2;
        set sim.exp_simulation_results2 exp_means;
        if _STAT_ NE "MEAN" then delete;
run;

%end;
%mend;

%exp_simulation(25);

proc means data=sim.exp_simulation_results2;
run;
```