For information:

Sage Publications, Inc.
2455 Teller Road
Thousand Oaks, California 91320
E-mail: order@sagepub.com

Sage Publications Ltd.
1 Oliver's Yard
55 City Road
London EC1Y 1SP
United Kingdom

Sage Publications India Pvt. Ltd.
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road, New Delhi 110 044
India

Sage Publications Asia-Pacific Pte. Ltd.
33 Pekin Street #02-01
Far East Square
Singapore 048763

Printed in the United States of America.

Library of Congress Cataloging-in-Publication Data

Handbook of cognitive aging: interdisciplinary perspectives/edited by Scott M. Hofer, Duane F. Alwin.
p. cm.
Includes bibliographical references and indexes.
ISBN 978-1-4129-4105-1 (cloth: alk. paper)
ISBN 978-1-4129-6028-1 (pbk.: alk. paper)
1. Cognition—Age factors. 2. Aging—Psychological aspects. I. Hofer, Scott M. II. Alwin, Duane F. (Duane Francis), 1944–

BF724.55.C63H36 2008
155.67'13—dc22 2007039241

This book is printed on acid-free paper.

08 09 10 11 12 10 9 8 7 6 5 4 3 2 1

Acquisitions Editor:	Cheri Dellelo
Editorial Assistant:	Lara Grambling
Production Editor:	Tracy Buyan
Copy Editor:	Kathy Anne Savadel
Typesetter:	C&M Digitals (P) Ltd.
Proofreader:	Penelope Sippel
Indexer:	Kathy Paparchontis
Cover Designers:	Ravi Balasuriya, Cheryl McLean
Marketing Manager:	Stephanie Adams

7

CHALLENGES IN ATTENTION

Measures, Methods, and Applications

JOAN M. MCDOWD AND LESA HOFFMAN

The history of gerontology indicates that the study of cognitive aging had its beginnings in applied science. In the late 1920s, the issue of the older worker in industry prompted the study of age differences in sensory and motor abilities. In the 1950s, Welford's work (e.g., Welford, 1958) in human skill and its application to understanding the older worker further contributed to awareness of the importance of understanding cognitive aging as well as to the empirical database documenting cognitive aging. The work of Birren and Botwinick in the 1950s and 1960s (e.g., Birren & Botwinick, 1955) continued the development of knowledge regarding perception and speed of processing in cognitive aging. Then came Canestrari, Eisdorfer, Arenberg, Hulicka, Talland, Craik, and Rabbitt, among others, and the study of cognitive aging was soon growing exponentially (see Riegel, 1977).

Given this history, it is interesting that the executive summary of the National Research Council's (2000) report *The Aging Mind* states that "*Now* [italics added] is a time of great promise for learning more about the aging mind and *turning that knowledge to the advantage of older people* [italics added]" (p. 1). Thus the emphasis is on obtaining information and then translating that information into useful prescriptions for maintaining and improving quality of life among older adults. Indeed, the report goes on to urge the National Institute on Aging to develop "the knowledge needed to design effective technologies . . . to support adaptivity in older people" (p. 3). In this chapter, we address the question of how the study of attentional ability in aging can contribute to this goal as well as outline some more general approaches of measurement that are likely to become increasingly important in the future of cognitive aging research.

ATTENTION AND FUNCTIONAL STATUS

Attention, or control of attention as regulated via executive function, is increasingly recognized as an important factor in the functional status of older adults. This association has been observed in both clinical and nursing home populations (Rochester et al., 2004; Royall, Mahurin, & Gray, 1992; Swanberg, Tractenberg, Mohs, Thal, & Commings, 2004) and generally healthy community-dwelling older adults (e.g., Carlson et al., 1999; Grigsby, Kaye, Baxter, Shetterly, & Hamman, 1998; Royall, Palmer, Chiodo, & Polk, 2004). For example, an association between attentional functioning and functional status was reported by Carlson et al. (1999). They assessed functional ability and what they called "executive attention" in a large sample of women from the Women's Health in Aging Study II. Their functional assessment included both activities of daily living (walking speed, dressing) and instrumental activities of daily living (dial a phone; using a key in a deadbolt lock). Executive attention measures included the Hopkins Attention Screening Test, "measuring temporal sequencing ability" (Carlson et al., 1999, p. S264), which uses a series of lights and tones in a task similar to the children's game Simon Says; the Trail Making Test (A and B); and the Brief Test of Attention, which requires monitoring and counting letters or numbers in an auditory sequence. They found that "executive tests of planning, organization, and flexibility were selectively associated with the performance of IADLs in a physically high-functioning, cognitively intact, community-dwelling sample of older women" (Carlson et al., 1999, p. S268). They further found that the Trail Making Test B accounted for the greatest proportion of variance among the cognitive tests. They concluded that "mental flexibility, rather than fine motor agility, is the attentional component critical to efficiently completing many complex, everyday activities" (Carlson et al., 1999, p. S268).

The same sort of association has been observed in clinical populations, such as people with Alzheimer's or Parkinson's disease. For example, Swanberg et al. (2004) assessed attention in patients with Alzheimer's disease using a cancellation task to test "the subject's ability to concentrate and use appropriate search strategies" and a maze task to test "impulse resistance, planning, reasoning, and foresight" (p. 557). They then split their sample of patients with Alzheimer's disease into two groups: (1) those with attention deficits (defined as scores 1.5 *SD* or more below the mean scores of a healthy comparison group) and (2) those without deficits, and compared their functional status using the Alzheimer's Disease Cooperative Study Activities of Daily Living Inventory, which assesses both basic and instrumental activities. They found that the group with attention deficits had significantly worse functional scores than those without attention deficits and concluded that the patients with attention impairments "had poorer everyday and community function" (Swanberg et al., 2004, p. 559). Rochester et al. (2004) examined the role of attention in walking performance of people with Parkinson's disease; they concluded that impaired attentional function increases walking difficulty for people with Parkinson's disease.

Together, these studies indicate a significant role for attention abilities in functional status. Future studies could work toward identifying the specific attentional processes that are responsible for the association and in that way be able to inform intervention and care planning strategies. McDowd, Filion, Pohl, Richards, and Stiers (2003) did some initial work along these lines in the area of stroke recovery. Their project was motivated both by the literature on stroke that was increasingly documenting cognitive deficits and their role in recovery, as well as a series of focus groups conducted with stroke survivors. In those focus groups, the issue of attention and memory problems was raised repeatedly as a significant concern affecting stroke survivors in their everyday lives. In light of this, McDowd et al. designed a project to test a group of stroke survivors who were at least 6 months poststroke, along with a group of older adults without stroke, on a series of experimental tests of attention: divided attention, switching attention, and sustained attention. In addition, they used a stroke-specific measure of everyday function, the Stroke Impact Scale (Duncan et al., 1999), to assess the relation between the cognitive measures and functional status among these individuals.

The findings indicated that, relative to individuals without stroke, stroke survivors were less able to maintain the simultaneous performance of two tasks (divided attention), were more slowed by the requirement to switch attention between stimulus features, and were less able to sustain attention across time. To assess the relevance of these differences observed in the laboratory, McDowd et al. (2003) estimated correlations between the attention measures and the five subscales of the SIS: (1) Physical Functioning, (2) Emotional Control, (3) Communication, (4) Memory, and (5) Social Participation. They observed several moderate associations. Stroke survivors' performance on the memory task under divided-attention conditions was correlated with the Physical Functioning, Memory, and Social Participation subscales. Performance on the alternating-attention task was correlated with the Social Participation subscale, and performance on the sustained-attention task was correlated with scores on the Physical Functioning subscale and the Social Participation subscale. These findings indicate that the level of ability measured in the laboratory was related to people's perceptions of their own everyday abilities. The promise of these data is that measures of attentional ability may be useful in predicting outcome following stroke and may contribute to the planning for accommodations necessary to maximize independence and quality of life. Part of this may involve tailoring rehabilitation interventions to match the needs and abilities of the person receiving services. In addition, these data raise the question of whether attention training might have a positive influence on outcome following stroke. This would seem to be an interesting and important avenue for future research.

Another domain in which attention has been studied relates to the attentional requirements of gait and balance. Investigators interested in balance and gait have frequently used dual-task methodology to understand group differences. In these studies, a secondary task is used to assess the attentional requirements of gait or balance under various conditions. The secondary task typically is of little interest except as a consumer of attention. The primary interest is in what happens to gait or balance under divided-attention conditions. The tables are easily turned, however; Kemper, McDowd, Pohl, Herman, and Jackson (2006) carried out a study of "walking and talking" in which they used the motor task as an attention consumer instead of as a primary task, because they were most interested in language function in the presence of other attentional demands. The participants were a group of 10 stroke survivors who were judged to be highly recovered based on a set of standard assessments including both motor and cognitive items. In particular, none of the participants was deemed aphasic on the basis of their performance on the Aphasia Diagnostic Profile (Goodglass, Kaplan, & Barresi, 2000). However, Kemper et al.'s hypothesis was that dual-task conditions would reveal the presence of significant residual cognitive deficits that were not detected under single-task conditions. Thus, they were predicting that the dual-task assessments would be more sensitive to cognitive deficits than standard assessments against which stroke recovery is judged. Participants were asked to respond to items such as "What is the most important invention in the 20th century?" or "Describe a recent vacation that you enjoyed" for 3 to 5 minutes, while also performing finger-tapping, walking, or no secondary task. Language samples and motor task performance were recorded for off-line analysis.

Motor task performance was significantly affected by dual-task requirements; time on task, tapping, and walking rates were each negatively affected by concurrent talking. In the case of language performance, Kemper et al. (2006) examined indexes of language fluency, complexity, and content. In each case, stroke survivors were more negatively affected by the requirement to divide attention than was a comparison group of older adults without stroke. In some cases the effect was quite dramatic, as either speech or motor performance stopped completely while the participant performed the other task. Thus, although participants had normal language as measured by the Aphasia Diagnostic Profile, they "became aphasic" when task demands were significantly increased. Therefore, the extent of cognitive deficits is not always immediately obvious. Practically speaking, these findings suggest that rehabilitation therapies might be administered in both optimal and suboptimal conditions in order to mimic performance with real life demands. In terms of advancing theory and knowledge about attentional abilities in aging, this study makes obvious that very sensitive measures are required to assess more subtle deficits. Accordingly, further exploration of how individual differences in attention relate to functional outcomes is likely to require new measurement tools that are both theoretically meaningful and psychometrically viable.

Measuring Attention

In our view, an important and productive avenue for developing psychometrically and theoretically sound measures of attention utilizes the collaboration of experimental and psychometric approaches. In a typical experimental study of attention, differences between younger and older adults in performance of carefully controlled tasks are evaluated at the group level. Accordingly, any individual differences observed within experimental studies are regarded as error—a nuisance to be eliminated as much as possible. However, it is just these individual differences in attentional abilities that are likely to be relevant in evaluating many different cognitive and functional outcomes in aging individuals. The removal of driving privileges, placement into assisted-care facilities, and other similar concerns are serious decisions that require the utmost precision of measurement in order to minimize misdiagnosis and mislabeling. Thus, individual differences in attention represent an important factor to be conceptualized, measured, and explored in the context of both theoretical examinations of cognition and in everyday life.

Therefore, in addition to experimental approaches, the study of attention could also be informed by employing correlational approaches from the other side of the methodological spectrum. Large-scale individual-differences studies of persons varying in age are common within the study of cognitive aging, often with an emphasis on the extent to which age-related differences are common or unique across perceptual and cognitive abilities (e.g., Anstey, Hofer, & Luszcz, 2003; Anstey, Luszcz, & Sanchez, 2001; Baltes & Lindenberger, 1997; Salthouse & Czaja, 2000; Salthouse & Ferrer-Caja, 2003). In contrast to other primary abilities, such as working memory or processing speed, however, examination of individual differences in attentional ability is exceedingly rare in correlational studies. The most likely reason for this lack of attention to the construct of attention is a lack of measurement tools with which attentional abilities can be assessed.

Investigators seeking to understand relationships among individual differences in cognitive processes often administer existing instruments used in neuropsychology (e.g., Wisconsin Card Sorting Test; Heaton, 1981) or the study of intelligence (e.g., the Wechsler Adult Intelligence Scale subscales; Wechsler, 1997). Although such measures may not be ideal, they are convenient for the study of individual differences in that they can be administered quickly and often without specialized equipment. Further, because they are commonly used, direct comparisons can be made with findings from other studies as well as with published norms. In comparison, investigators interested in relating individual differences in attention with those of other constructs frequently have resorted to ad hoc measures of individual mean response times from study-specific attention-based experimental tasks (e.g., D'Aloisio & Klein, 1990; Pringle, Kramer, & Irwin, 2004; Stankov, 1988). The use of mean response times or response time differences as direct indicators of individual ability may be appropriate within a single study, in which only the rank order of participants is of interest as it relates to their rank order on another outcome. However, the direct use of experimental tasks as psychometric instruments falls short of many goals of measurement. Despite their intuitive appeal, the limitations of such approaches preclude any meaningful interpretation of individual differences across samples or across time, and they offer little evidence of construct validity of the task as a measure of individual differences in attentional ability. The reasons for these limitations, as well as the solutions proposed by the use of explanatory latent trait modeling, are presented next.

Benefits of Explanatory Latent Trait Methods in the Measurement of Attention

The scoring of many popular cognitive tests has its roots in classical test theory (i.e., true score theory; Gulliksen, 1950), in which the same group of functionally identical items or trials are typically administered to each person, and an aggregate score across items (e.g., percentage correct, mean across items) is thought to provide the best of estimate of each person's true ability. The core assumption in aggregating across items is that the items are equivalent to each other in terms of their *difficulty*, or in their location along the latent continuum of ability to be measured, and in terms of their *discrimination*, or in the strength of their relation to the underlying ability measured by the test. Item aggregation has two undesirable consequences. First, because the overall score depends on the exact set of items administered, scores from different forms of the instrument (i.e., with items added or removed) cannot be directly compared. Second, the individual scores have meaning only in relation to the other scores from the sample. The same aggregate score could place a given person in the 90th or 40th percentile depending on the scores of the persons to whom he or she was being referenced. Large norming samples are thus needed to interpret the score of a given person, but such norm-based scores are still somewhat arbitrary in that they have little direct relation to what a given person can actually do (i.e., how he or she is likely to answer a given item).

A fundamentally different perspective is taken within latent trait theory (i.e., item response and Rasch models; De Boeck & Wilson, 2004; Embretson & Reise, 2000). For ease of exposition, we consider models in which accuracy is the response. The measurement model for a latent trait includes differences in the abilities of the persons being measured as well as differences in the difficulty of the items being answered. This separation of persons and items (within Rasch models in particular) is called *specific objectivity*: Person ability estimates do not depend on the particular items administered, and item difficulty estimates do not depend on the particular persons responding to the items. Person abilities and item difficulties are located on a common underlying or latent continuum. A person's ability is the latent trait level that is most likely to have given rise to the observed responses and is the item difficulty location at which he or she has a 50% probability of a correct response. A latent trait estimate is directly informative as to the most likely response to an item of any given difficulty. It is important to note that because latent trait estimates are interpreted relative to the items rather than to the sample, they are directly comparable across different combinations of items (i.e., test versions).

An additional issue concerns the measurement properties of ability scores within each framework. Latent trait estimates of ability have interval measurement properties; in other words, the latent trait metric is continuous, with unlimited points, and differences in latent abilities have the same meaning across the latent metric. On the other hand, because aggregate observed scores from classical test theory are likely to be compressed at the extreme ability levels, they are likely to be nonlinearly related to the underlying latent trait and thus ordinal rather than interval (Maxwell & Delaney, 1985). Although common, the use of ordinal outcomes in statistical analyses that assume interval measurement has been shown to result in greater bias and Type I error rates for estimates of group mean differences, group interactions, and regression coefficients than when analyzing interval-level latent ability outcomes. These problems worsen with a greater mismatch between the distributions of person ability and item difficulty, as is probable when groups that differ widely on ability are administered the same test (i.e., younger vs. older adults; Embretson, 1996; Kang & Waller, 2005). Although the extent of this mismatch can be evaluated within latent trait models, there is no basis for its evaluation within classical test theory.

The two frameworks also differ in terms of assessing reliability of measurement (see Smith, 2001). Reliability in classical test theory is assumed to be a static property that applies equally across all persons in the sample. Yet because items are often selected in order to be maximally sensitive at the mid-range of ability (i.e., a proportion passing of 50%), the range over which ability can be measured reliably is likely to be compressed. As a result, the obtained estimate of reliability may not be applicable for persons of extreme abilities, for whom many items may be too easy or too difficult and thus for whom little information is available with which to measure their abilities. In evaluating persons with attentional impairments, this lack of sensitivity at extreme ability levels can be problematic.

In contrast, in latent trait theory reliability differs across persons, such that the precision with which persons at a given ability level are measured depends directly on the availability of items of comparable difficulty. Measurement precision at each ability level can be evaluated explicitly; to the extent that the distributions of person ability and item difficulty are well matched (i.e., an adequate number of items for each ability level), then reliability will be high. It is thus desirable to include items that differ broadly in difficulty in order to achieve good coverage of the range of abilities in the population to be measured. One way to achieve a broad range of item difficulty is to design items that differ in their levels of the features believed to relate to the process being measured. Items can be designed systematically with features that produce targeted levels of difficulty, filling in any gaps along the ability–difficulty continuum and thus improving measurement precision. This advantage is significant when measuring change in aging individuals, because floor effects can be avoided by adding easier items at each occasion. Yet given precalibration of the properties of the potential pool of items, individual abilities can still be compared across time because they are in reference to the underlying common metric of item difficulty and person ability, rather than an aggregate score without a direct interpretation outside the sample at hand.

In addition to providing a means with which to improve measurement precision, decomposition of item difficulty also provides evidence for task validity. Specifically, the extent to which item difficulty relates to these item design features provides evidence for *construct representation*, or the underlying meaning of the construct measured by the test (see Embretson, 1983, 1998). For example, what do we mean by *attention*? How can one observe attentional processes, or how can one observe their limits? It is in this step that the rich experimental tradition can be most helpful. Items or tasks can be constructed to tap theoretically motivated aspects of attentional processing, and multiple dimensions of processing can be assessed independently through careful instrument design. Furthermore, as substantive theories of attentional processing undergo revision, so too can an instrument measuring attention (e.g., by incorporating new sources of difficulty). This represents a most exciting possibility—that individual-differences measures of attention (and of other processes as well) can become not only a tool with which to conduct research on cognitive aging but also a continually evolving reflection of that research.

A final issue concerns how *nomothetic span*, or a test's utility in measuring individual differences, can be evaluated within each framework. In addition to static estimates of reliability, psychometric evaluations in classical test theory are almost exclusively based on evidence of predictive and concurrent validity, or the extent to which the scores on the test relate to the scores from other tests in an anticipated direction. Such relationships do not unequivocally support construct representation, and they are not informative as to the sensitivity of the test in distinguishing individuals of different ability levels or detecting changes in ability. As discussed previously, precision of measurement across ability levels can be explicitly evaluated and potentially improved through the use of latent trait modeling, and this improvement in reliability is then likely in turn to improve the power with which relations of individual differences across constructs can be detected.

To summarize, the decomposition of difficulty within explanatory latent trait modeling provides the basis for developing psychometric instruments that are theoretically informed; sensitive over a wide range of abilities; and, most important, that are modifiable as needed while retaining comparability across samples or occasions. These modeling techniques provide important guidance for developing new instruments with which to assess individual differences in attention, as well as in evaluating current measures of attention.

Current Approaches in Measuring Attention

One such measure is the Useful Field of View Test (UFOV), developed by Ball, Owsley, and colleagues (Ball & Owsley, 1993; Owsley, Ball, Sloane, Roenker, & Bruni, 1991) on the basis of earlier work (Mackworth, 1965; Sanders, 1970). The UFOV has three subtests designed to assess the spatial extent of the attentional window in older adults in the context of predicting automobile accident risk and is sold commercially (by Harcourt Assessment) for that purpose. The Processing Speed subtest requires the discrimination of a central car or truck subtending 3° × 5° visual angle. The Divided Attention subtest pairs the same central discrimination task with a concurrent peripheral localization task, in which a car is also presented on one of eight radial spokes at a fixed eccentricity of 30°. In the Selective Attention subtest the central discrimination and peripheral localization tasks are performed in the presence of 47 triangles across the 30° visual field. Thus, in both the Divided Attention and the Selective Attention subtests, attention must be distributed in order to perform both tasks correctly, whereas attention must only be allocated centrally in the Processing Speed subtest.

Presentation time is increased after incorrect responses and decreased after correct responses, and the time needed to achieve 75% accuracy is the ability score for each subtest. Although the subtests have distinct attentional requirements, their thresholds are nevertheless combined into a single score, interpreted as the percentage reduction in the visual field. In the PC version of the UFOV as described, the eccentricity of the peripheral target is fixed, and thus "shrinkage" can be evaluated at only two levels, "near" or "far." In the original version of the UFOV, however, the eccentricity of the peripheral target was varied instead of presentation time. Scores between the original and PC versions have shown adequate correspondence (attenuation-corrected $rs \approx .7$), and each have shown adequate test–retest reliability over short intervals (corrected $rs \approx .7$ after < 1 year; Edwards et al., 2005). However, psychometric evaluation of the UFOV has mostly been limited to its utility in predicting accident risk. The UFOV has shown good sensitivity and specificity in predicting accident risk when participants are heavily oversampled for visual impairment and history of previous accidents (Ball & Owsley, 1993; Owsley et al., 1991). However, comparable levels of sensitivity and specificity have not been reported when participants are not oversampled for history of previous accidents (Brown, Greaney, & Mitchel, 1993, cited in Harris, 1999; Hennessey, 1995, cited in Staplin, Lococo, Stewart, & Decina, 1999; Hoffman, McDowd, Atchley, & Dubinsky, 2005).

A second measure of attention, the Attention Network Test (ANT; Fan, McCandliss, Sommer, Raz, & Posner, 2002; Fan, Wu, Fossella, & Posner, 2001; Fossella, Posner, Fan, Swanson, & Pfaff, 2002), was developed for use with children, adults, or animals and is available online. The ANT was derived on the basis of neurological evidence for the existence of three distinct functions of attention: (1) Alerting (achieving and maintaining an alert state), (2) Orienting (selection of information from sensory input), and (3) Executive Control (resolution of conflict among responses). Ability estimates in each function are obtained through the subtraction of mean response times in different conditions of a single task, a fixed-choice discrimination of the direction of central target (left or right arrow) flanked by response-neutral, incompatible, or compatible distracter arrows. Before each target is either no cue, an alerting cue (i.e., single or double nondirectional cue), or a valid spatial cue. For Alerting, double-alerting-cue mean response time is subtracted from no-cue mean response time. For Orienting, spatial-cue mean response time is subtracted from single-alerting-cue mean response time. For Executive Control, congruent-response mean response time is subtracted from incongruent-response mean response time. Subtraction scores were uncorrelated across functions in 40 adults ages 20 through 44, suggesting orthogonal functions of attention, although the effect of incongruent responses was slightly greater in the alerting-cue condition. One-hour test–retest correlations for the subtraction scores ranged from .5 to .7 (Fan et al., 2002).

A third measure of attention, DriverScan, illustrates how experimental findings can be translated into components of difficulty within a psychometric instrument for measuring attentional ability in aging individuals (Hoffman et al., 2005; Hoffman, Yang, Bovaird, & Embretson, 2006). DriverScan uses a change detection task (Rensink, O'Regan, & Clark, 1997) in which original (A) and modified (A′) digital photographs of driving scenes are presented for 280 milliseconds, while blank screens are interspersed for 80 milliseconds as follows: A, blank, A, blank, A′, blank, A′, blank . . . , for 45 seconds or until the change is detected, whichever comes first. With this method of presentation, search for a change between repeated presentations of an otherwise-identical scene must be conducted through controlled processing, given that local luminance cues at the change location are unable to direct attention in the presence of a global luminance change (the blank screen). It often takes considerably longer to notice even large, salient changes than when such changes are presented without interruption (i.e., *change blindness*). The primary measure of performance is a combination of accuracy and response time to detect the change, including categories of *immediate response* (< 8 seconds), *delayed response* (8–45 seconds), or *no response* (a time out of the trial), and a Rasch version of a graded response model was used to estimate latent traits as an index of attentional search ability.

Previous research has suggested that both endogenous (goal-directed) and exogenous (stimulus-driven) orienting can be utilized to facilitate change detection (e.g., Hollingworth & Henderson, 2000; Scholl, 2000; Werner & Thies, 2000; Williams & Simons, 2000). The DriverScan items were designed to vary along three dimensions in order to tap into these types of search processes, and subjective ratings for each dimension for each item were obtained during test development. Stimulus-driven attentional search was represented through the dimensions of *visual clutter* (i.e., the overall level of congestion of the scene) and *change brightness* (i.e., the overall conspicuity of the change), whereas goal-directed attention search was represented through *change relevance*, or how important the change would be to the driver in the scene. Each dimension was shown to significantly predict item difficulty, such that items were more difficult when they included a scene with a greater amount of visual clutter and featured a change of less brightness and less relevance to driving. Thus, a deficit in any one of these areas would be related to less efficient search performance in the task (Hoffman et al., 2006). Furthermore, DriverScan performance was shown to be related to the Divided and Selection Attention subtests of the UFOV and was shown to independently predict simulated driving performance in older adults over and above the contributions of visual functioning and the UFOV subtests (Hoffman et al., 2005).

Evaluating Current Approaches to Measuring Attention

The principles of latent trait modeling can serve as a guide for evaluating the strengths and weaknesses of each of aforementioned measures of attention. Furthermore, these principles can also inform the development of new measures of attention through the integration of experimental and psychometric approaches.

The first consideration of instrument design is the response to be modeled. In the ANT, the outcome is the difference in mean response time across trials that have different functional requirements. Because only response times from correct trials are included, to the extent that accuracy differs across items or across persons, response time will be misleading as the sole indicator of ability. The response times of persons with low ability may also be less reliable because fewer trials were answered correctly to be included in their means. Furthermore, large response times may be obtained from persons who lack the ability needed to solve the items of from persons with sufficient ability but who deliberately respond more slowly.

A different problem with the response outcome is encountered in the UFOV, in which the outcome for each subtest is the presentation time needed to achieve 75% accuracy. In the Divided Attention and Selective Attention subtests, if the central task is performed incorrectly, then that trial is not counted, and another trial is administered. Thus, observers are differentially receiving practice on these tasks, or "speed of processing training" (e.g., Ball et al., 2002), while being measured. As a result, persons of lower abilities may have artificially improved scores relative to persons who perform the central task more accurately to begin with.

Yet another problem with the response outcome is encountered in DriverScan. Because response times were censored by accuracy, the response time distribution was divided at admittedly arbitrary points to approximate an ordinal variable that best describes the combination of speed and accuracy. This is not an ideal solution, because there is no empirical basis with which to evaluate the quality of the cut-points.

In short, the problem of speed–accuracy dependency is likely to arise whenever response times are the primary outcome and accuracy is not at ceiling, or when accuracy is the primary outcome but response time is still informative as to individual differences in the process under study. A possible resolution lies in the conversion of time to an independent variable, or as another design feature along which items can vary (e.g., Verhaeghen, 2000). Accuracy for a given presentation time would then be the sole outcome measure. The manipulation of time as a design feature has been applied in many experimental paradigms and in the UFOV test and is also a natural manner in which the range of difficulty covered in a test can be extended to measure extreme levels of ability more reliably.

A second consideration of instrument design is variation in the difficulty levels of the test items and the resulting precision of measurement across ability levels. To the extent that the factors manipulated to produce trials of differential difficulty have very few distinct levels, the capacity of the test to reliably distinguish individuals of different abilities will be severely limited. For example, in the ANT, Alerting is measured by contrasting trials with two types of precues; differences of distractor type across trials are ignored. Executive Control is measured by contrasting trials with two types of distractors; differences of precue type across trials are ignored. In the UFOV, although subtraction of thresholds is not explicitly conducted as in the ANT, Selective Attention is essentially measured by two types of trials: with or without distractors, and Divided Attention ability is also measured by two types of trials: "near" or "far" targets. Even though the incremental varying of presentation time should help matters, just as instructors would not expect to obtain reliable estimates of achievement from a test with two questions, a researcher cannot expect to obtain reliable estimates of attentional ability from a task with limited variation in its trials. Instead, the levels of the factors thought to relate to task difficulty should vary more continuously from one another rather than discretely. Although completely contrary to the traditional experimental mindset, factors that vary more continuously will be helpful in obtaining the range of item difficulty necessary for precise measurement of individual differences. However, it is important that the levels of the factors be independently verifiable and meaningful, rather than based on subjective ratings (i.e., as was used in DriverScan).

Finally, the process of instrument design must consider the generalizability of performance in experimental tasks to real world abilities. Research on attention has primarily relied on tightly controlled and thus necessarily contrived experimental tasks, yet accurate, responsible assessment of attentional functioning requires that we move beyond artificial laboratory manipulations into a more realistic and ultimately applicable depiction of what the result of attentional declines may be for an older person. Because attention is thought to support higher-order cognitive processing, it is essential to assess attentional abilities in the contexts in which they operate. Previous research has demonstrated that older adults are able to make use of environmental supports to guide attention, yet neither the UFOV nor the ANT include a real world context in which to perform their tasks. Thus, the extent to which any deficits observed in these tasks would be evidenced in more contextually driven and ecologically valid tasks is an open question, one with significant implications for real world decisions. Kramer and Kray (2006) also noted the "relatively sparse literature" comparing performance under laboratory and "real-world" situations (pp. 65–66), observing that such comparisons will be an important part of future research.

Although DriverScan does feature a real world context (i.e., of driving), the use of real world scenes directly as test items presents a new problem, in that the items cannot be meaningfully readministered at later points in time (i.e., because they are likely to be recognized). One solution for this problem within longitudinal studies is to create an item bank of natural scenes and manipulated scene changes and administer only some of the items at a given occasion. This solution may not be practically feasible, however, given that natural scenes will vary in multiple unknown dimensions and thus would not be directly comparable, even after controlling for key design features. A better solution would be to make use of the identified sources of difficulty to computer-generate new items as needed online. As discussed earlier, these design features could then provide the structure for an infinite number of items that vary along the dimensions that underlie item difficulty.

Conclusion

Although the studies described here have operationalized attentional performance in different ways, they are by no means exhaustive. Indeed, one of the thorniest problems researchers interested in studying and measuring attention must face is the multifaceted nature of the construct. The integral nature of attention in cognitive processing makes it difficult to construct what is ultimately needed: an independent definition and conceptualization of attention that could be the basis of one or more psychometric instruments. Continued development of measures and methods for examining individual differences in attention is critical to the future success of theoretical investigations of cognitive aging as well as real world assessments of ability and disability.

Research on attention thus far has been largely based on experimental designs, with very little emphasis on individual differences. Yet the measurement of attention and other cognitive processes could be well informed by traditions common in the educational testing fields, namely, the use of model-based measurement procedures. Specifically, we believe a latent trait approach as presented here shows great promise for developing and evaluating meaningful and reliable psychometric instruments with which to assess ability and change in ability. Indeed, by carefully considering the properties of both items and persons, we can construct measures that are most efficient and most accurate and yet retain the possibility of comparing across samples or time points within an individual. Furthermore, as substantive theories of attentional processing in aging are continually refined, through an understanding of the origins of item difficulty, so too can measures of individual differences.

The study of attentional processes in older age has had a rich history, a tradition that will no doubt continue. But the time has come to ask more of our research in order to answer the challenges of the world outside the laboratory. Meeting these challenges will likely require research teams who bring varied methodological and disciplinary expertise to the problem of understanding attention in aging and age-related neurological disease. This "problem" is an important one, having implications for cognitive aging theory as well as laying the groundwork for translating empirical findings into assessment tools that can be used to maximize quality of life among older adults.

REFERENCES

Anstey, K. J., Hofer, S. M., & Luszcz, M. A. (2003). Cross-sectional and longitudinal patterns of dedifferentiation in late-life cognitive and sensory function: The effects of age, ability, attrition, and occasion of measurement. *Journal of Experimental Psychology: General, 132,* 470–487.

Anstey, K. J., Luszcz, M. A., & Sanchez, L. (2001). A reevaluation of the common factor theory of shared variance among age, sensory function, and cognitive function in older adults. *Journal of Gerontology: Psychological Sciences, 56B,* P3–P11.

Ball, K., Berch, D. B., Helmers, K. F., Jobe, J. B., Leveck, M. D., Marsiske, M., et al. (2002). Effects of cognitive training interventions with older adults: A randomized controlled trial. *Journal of the American Medical Association, 288,* 2271–2281.

Ball, K. K., & Owsley, C. (1993). The Useful Field of View Test: A new technique for evaluating age-related declines in visual function. *Journal of the American Optomological Association, 63,* 71–79.

Baltes, P. B., & Lindenberger, U. (1997). Emergence of a powerful connection between sensory and cognitive functions across the adult life span: A new window to the study of cognitive aging? *Psychology and Aging, 12,* 12–21.

Birren, J. E., & Botwinick, J. (1955). Age differences in finger, jaw, and foot reaction time to auditory stimuli. *Journal of Gerontology, 10,* 429–432.

Carlson, M., Fried, L., Xue, Q., Bandeen-Roche, K., Zeger, S., & Brandt, J. (1999). Association between executive attention and physical functioning performance in community-dwelling older women. *Journal of Gerontology: Social Sciences, 54B,* 262–270.

D'Aloisio, A., & Klein, R. M. (1990). Aging and the deployment of visual attention. In J. T. Enns (Ed.), *The development of attention: Research and theory* (pp. 447–465). New York: Elsevier North-Holland.

De Boeck, P., & Wilson, M. (2004). *Explanatory item response models.* New York: Springer.

Duncan, P. W., Wallace, D., Lai, S. M., Johnson, D., Embretson, S., & Laster, L. J. (1999). The Stroke Impact Scale version 2.0: Evaluation of reliability, validity, and sensitivity to change. *Stroke, 30,* 2131–2140.

Edwards, J. D., Vance, D. E., Wadley, V. G., Cissell, G. M., Roenker, D. L., & Ball, K. K. (2005). Reliability and validity of Useful Field of View Test scores as administered by personal computer. *Journal of Clinical and Experimental Neuropsychology, 27,* 529–543.

Embretson, S. E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin, 93,* 179–197.

Embretson, S. E. (1996). Item response theory models and spurious interaction effects in factorial ANOVA designs. *Applied Psychological Measurement, 20,* 201–212.

Embretson, S. E. (1998). A cognitive design system approach to generating valid tests: Application to abstract reasoning. *Psychological Methods, 3,* 380–396.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Lawrence Erlbaum.

Fan, J. I., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience, 14,* 340–347.

Fan, J., Wu, Y., Fossella, J., & Posner, M. I. (2001). Assessing the heritability of attentional networks. *BioMed Central Neuroscience, 2,* 14.

Fossella, J., Posner, M. I., Fan, J., Swanson, J. M., & Pfaff, D. W. (2002). Attentional phenotypes for the analysis of higher mental function. *Scientific World, 2,* 217–223.

Goodglass, H., Kaplan, E., & Barresi, B. (2000). *Boston Diagnostic Aphasia Examination* (3rd ed.). Austin, TX: Pro-Ed.

Grigsby, J., Kaye, K., Baxter, J., Shetterly, S., & Hamman, R. (1998). Executive cognitive abilities and functional status among community-dwelling older persons in the San Luis Valley Health and Aging Study. *Journal of the American Geriatrics Society, 46,* 590–596.

Gulliksen, H. (1950). *Theory of mental tests.* New York: Wiley.

Harris, P. (1999). Vision and driving in the elderly. *Journal of Optometric Vision Development, 30,* 188–197.

Heaton, R. K. (1981). *A manual for the Wisconsin Card Sorting Test.* Odessa, FL: Psychological Assessment Resources.

Hoffman, L., McDowd, J. M., Atchley, P., & Dubinsky, R. A. (2005). The role of visual attention in predicting driving impairment in older adults. *Psychology and Aging, 20,* 610–622.

Hoffman, L., Yang, X., Bovaird, J. A., & Embretson, S. E. (2006). Measuring attention in older adults: Development and psychometric evaluation of DriverScan. *Educational and Psychological Measurement, 66,* 984–1000.

Hollingworth, A., & Henderson, J. M. (2000). Semantic informativeness indicates the detection of changes in natural scenes. *Visual Cognition, 7*(1/2/3), 213–235.

Kang, S. M., & Waller, N. G. (2005). Moderated multiple regression, spurious interaction effects, and IRT. *Applied Psychological Measurement, 29,* 87–105.

Kemper, S., McDowd, J., Pohl, P., Herman, R., & Jackson, S. (2006). Revealing language deficits after stroke: The costs of doing two things at once. *Aging, Neuropsychology, and Cognition, 13,* 113–139.

Kramer, A. F., & Kray, J. (2006). Aging and attention. In E. Bialystock & F. I. M. Craik (Eds.), *Lifespan cognition: Mechanisms of change* (pp. 57–69). Oxford, UK: Oxford University Press.

Mackworth, N. H. (1965). Visual noise causes tunnel vision. *Psychonomic Science, 3,* 67–68.

Maxwell, S. E., & Delaney, H. D. (1985). Measurement and statistics: An examination of construct validity. *Psychological Bulletin, 97,* 85–93.

McDowd, J., Filion, D., Pohl, P., Richards, L., & Stiers, W. (2003). Attentional abilities and functional outcome following stroke. *Journal of Gerontology: Psychological Sciences, 58B,* P45–P53.

National Research Council. (2000). *The aging mind: Opportunities in cognitive research.* Washington, DC: National Academy Press.

Owsley, C., Ball, K. K., Sloane, M. E., Roenker, D. L., & Bruni, J. R. (1991). Visual/cognitive correlates of vehicle accidents in older drivers. *Psychology and Aging, 6,* 403–415.

Pringle, H. L., Kramer, A. F., & Irwin, D. E. (2004). Individual differences in the visual representation of scenes. In D. T. Levin (Ed.), *Thinking and seeing: Visual metacognition in adults and children* (pp. 165–186). Westport, CT: Greenwood/Praeger.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8,* 368–373.

Riegel, K. F. (1977). History of psychological gerontology. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 70–102). New York: Van Nostrand Reinhold.

Rochester, L., Hetherington, V., Jones, D., Nieuwboer, A., Willems, A., Kwakkel, G., et al. (2004). Attending to the task: Interference effects of functional tasks on walking in Parkinson's disease and the roles of cognition, depression, fatigue, and balance. *Archives of Physical Medicine and Rehabilitation, 85,* 1578–1585.

Royall, D. R., Mahurin, R. K., & Gray, K. (1992). Bedside assessment of executive impairment: The Executive Interview (EXIT). *Journal of the American Geriatrics Society, 40,* 1221–1226.

Royall, D. R., Palmer, R., Chiodo, L., & Polk, M. (2004). Declining executive control in normal aging predicts change in functional status: The Freedom House study. *Journal of the American Geriatrics Society, 52,* 346–352.

Salthouse, T. A., & Czaja, S. J. (2000). Structural constraints on process explanations in cognitive aging. *Psychology and Aging, 15,* 44–55.

Salthouse, T. A., & Ferrer-Caja, E. (2003). What needs to be explained to account for age-related effects on multiple cognitive variables? *Psychology and Aging, 18,* 91–110.

Sanders, A. F. (1970). Some aspects of the selective process in the functional visual field. *Ergonomics, 13,* 101–117.

Scholl, B. J. (2000). Attenuated change blindness for exogenously attended items in a flicker paradigm. *Visual Cognition, 7*(1/2/3), 377–396.

Smith, E. V., Jr. (2001). Evidence for the reliability of measures and validity of measure interpretation: A Rasch measurement perspective. *Journal of Applied Measurement, 2,* 281–311.

Stankov, L. (1988). Aging, attention, and intelligence. *Psychology and Aging, 3,* 59–74.

Staplin, L. S., Lococo, K. H., Stewart, J., & Decina, L. E. (1999). *Safe mobility for older people notebook* (DOT HS 808-853). Washington, DC: National Highway Traffic Safety Administration. Retrieved September 8, 2006, from http://www.nhtsa.dot.gov/people/injury/olddrive/safe/safe-toc.htm

Swanberg, M. M., Tractenberg, R. E., Mohs, R., Thal, L. J., & Cummings, J. L. (2004). Executive dysfunction in Alzheimer disease. *Archives of Neurology, 61,* 556–560.

Verhaeghen, P. (2000). The parallels in beauty's brow: Time–accuracy functions and their implications for cognitive aging theories. In T. J. Perfect & E. A. Maylor (Eds.), *Models of cognitive aging* (pp. 50–85). Oxford, UK: Oxford University Press.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale—Third edition.* San Antonio, TX: Harcourt Assessment.

Welford, A. T. (1958). *Aging and human skill.* London: Methuen.

Werner, S., & Thies, B. (2000). Is "change blindness" attenuated by domain-specific expertise? An expert–novice comparison of change detection in football images. *Visual Cognition, 7*(1/2/3), 163–173.

Williams, P., & Simons, D. J. (2000). Detecting changes in novel, complex three-dimensional objects. *Visual Cognition, 7*(1/2/3), 297–322.