a sentence." We have observed just such an effect in an experiment measuring response accuracy to probe questions about sentences containing two to five propositions (Roberts et al. 1997). When subjects are divided into a high-capacity group and a low-capacity group using the reading span measure, both groups are near ceiling with two-clause sentences (99% vs. 92%) and are equally inaccurate with four-clause sentences (34% vs. 31%). However, high-span subjects are significantly more accurate in recalling the propositional content of three-clause sentences (73% vs. 54%). These data provide evidence that the reading span measure divides subjects according to their ability to retain information about the propositional content of sentences, a post-interpretive process. There is currently no measure of the capacity of the interpretive processing system that does not tap post-interpretive processes; designing such a measure remains a goal for further research.

Third, what is the relation between the systems used for interpretive processing and the general verbal working memory system used for post-interpretive processing? C&W claim that the separate language interpretation resource (SLIR) is a separate module within working memory, but what is evidence for this claim? Given that deficit–lesion correlation studies and imaging studies point to different neurological substrates for the two systems, and that C&W have experimentally demonstrated that they do not interact, it seems more plausible to hypothesize that the SLIR and general working memory are entirely distinct. In any event, the fact that the two systems appear to be neurally distinct provides strong evidence against Just, Carpenter and colleagues' view (see, e.g., Just & Carpenter 1992) that the same units are used for interpretative processing and storage within working memory.

What do working-memory tests really measure?

Michael J. Kane,[a] Andrew R. A. Conway,[b] and Randall W. Engle[c]

[a]*Department of Psychology, Georgia State University, Atlanta, GA 30303-3083;* [b]*Department of Psychology, University of Illinois at Chicago, Chicago, IL 30332-0170; and* [c]*School of Psychology, Georgia Institute of Technology, Atlanta, GA 30332-0170.* **mkane@gsu.edu aconway@uic.edu re23@prism.gatech.edu**

Abstract: Individuals may differ in the general-attention executive component or in the subordinate domain-specific "slave" components of working memory. Tasks requiring sustained memory representations across attention shifts are reliable, valid indices of executive abilities. Measures emphasizing specific processing skills may increase reliability within restricted samples but will not reflect the attention component responsible for the broad predictive validity of span tasks.

We thank Caplan & Waters (C&W) for pointing out the weaknesses in the view that language interpretation and parsing are limited by general, domain-free working memory resources (see, e.g., Just & Carpenter 1992). Based largely on their earlier work (Caplan & Waters 1990), Engle and Conway (1998) concluded that there is little evidence to support the claim that the phonological loop or the central executive plays an important role in the early stages of language processing. We have argued that the working memory system is theoretically equivalent to a view of human intelligence proposed by Cattell (1963) and Horn (1980), which distinguishes between a domain-free general fluid intelligence and domain-specific elements of crystallized intelligence, which are peculiar to knowledge, skill, and talent.

Our view, derivative of Baddeley and Hitch (1974), sees working memory as a system of (1) activated traces, which are the output of domain-specific processors, and (2) domain-free controlled attention (Engle et al., in press b; Kane & Engle 1998). Individuals may vary in the quality and skill of their domain-specific processors. They may also vary in their general fluid ability to bring controlled attention to bear on managing information in the temporary storage buffers and on resolving conflict at output (cf. Shallice & Burgess 1993). Thus, variability on working memory tasks can reflect individual differences in the domain-specific processors, the domain-free controlled attention, or both.

We argue that the general capability to sustain an active memory representation in the face of attention shifts or distraction is what drives the relationship between working memory span measures and diverse higher order capabilities. C&W complain that the processing component of working memory span tasks "introduces a dual-task element into these tasks and allows them to be heavily affected by the capacity for divided attention." We argue, however, that the divided-attention component is the strength of working memory span tasks. The divided-attention component is critical to the broad predictive validity of span tasks (and, therefore, of fluid intelligence tests) across task and stimulus domains.

All studies that use individual or group differences to make inferences about working memory involve the measurement of constructs and the partitioning of variance in those constructs. If the range of general ability in a sample of subjects is sufficiently broad, then measures of working memory capacity, such as the reading span task (Daneman & Carpenter 1980) and the operation span task (Turner & Engle 1989), will be reliable. Furthermore, they will reflect the ability of both the domain-free central executive and the domain-specific processors necessary to perform the tasks because no working memory task is a pure test of either the central executive or the slave systems (Engle et al., in press a). If, however, a sample's range of general ability is narrow, as it often is at prestigious universities, there will be little variability in the general fluid ability construct. Therefore, the variability in performance for the restricted sample will reflect primarily differences in the skill of the domain-specific processors necessary to perform the task.

To the extent that the "processing" component is entered into the working memory span score as C&W suggest, reliability may increase. However, the construct measured by that score is more likely to reflect the domain-specific processors required for that task and not the general controlled attention–central executive construct. In fact, our own work (e.g., Conway & Engle 1996; Conway et al. 1998) suggests that processing skills within span tasks can actually suppress the relationship between working memory scores and higher order cognition. Thus, the "innovation" proposed by C&W may increase reliability in a restricted sample, but it will be a good test only if the goal is to measure specific processing skills. It will not be a good test if the goal is to measure general domain-free controlled attention capacity. If the goal is to measure the domain-free executive component of working memory, then we recommend administering a battery of different working memory tasks that share the dual-task quality mentioned above but differ in the domain-specific processes required to perform the task.

C&W questioned the reliability of the reading span task, but, as we pointed out above, low reliability can result from sampling a restricted range of individuals. Kitty Klein and William Fiss of North Carolina State University recently conducted an extensive analysis of the operation span task (Turner & Engle 1989). They tested 33 subjects at three different times, the second time 3 weeks after the first and the third time 6–7 weeks after the second. The corrected reliability was .88 and the stability scores ranged from .76 to .92.

As for the "fractionation" of working memory, we would argue that C&W's measures reflect quite specific skills at parsing and processing syntactically complex sentences. These processes occur at a pre-interpretative level and thus do not generally make demands on the central executive. The two different components they propose reflect the outputs from specific processors that do their work with little need for limited-capacity, controlled attention. For example, adding a secondary working memory load to a syntactic processing task has no detrimental effect. Because syntactic processing may proceed unhindered by a load, it may be performed relatively automatically, without much controlled attention. However, secondary load tasks do interfere with syntactic processing when the load-task stimuli are interleaved with the syntax task, demonstrating that, when syntactic information must be sustained across an attention shift, it suffers significantly.

Rather than assign syntactic processing to a specialized component of working memory, then, we suggest that it operates independently of the central executive. That is, working memory capacity is needed only under attention-demanding circumstances, and, insofar as syntactic processing appears to be immune to divided-attention conditions, it likely occurs relatively automatically. Caplan and Waters (1990) argued that the phonological loop may be required in some syntactic parsing circumstances, such as when subjects are "garden-pathed" or when many words must be maintained in active memory. Why not use that interpretation for the present work?

The age invariance of working memory measures and noninvariance of producing complex syntax

Susan Kemper and Karen A. Kemtes

Department of Psychology and Gerontology Center, University of Kansas, Lawrence, KS 66045. **skemper@ukans.edu**
www.lsi.ukans.edu/clang/sklab.htm **kkemtes@ukans.edu**

Abstract: In challenging current conceptions of the role of working memory in sentence processing, Caplan & Waters consider studies comparing young and older adults on sentence processing. This commentary raises two challenges to Caplan & Waters's conclusions: first, working memory tasks appear to be age invariant. Second, the production of complex syntactic constructions appears not to be age invariant.

Caplan & Waters (C&W) raise a number of questions about the role of working memory during sentence processing. They consider evidence from studies comparing young and older adults on sentence processing tasks relevant to this issue because older adults are typically found to exhibit working memory deficits on a variety of tasks. This commentary raises two challenges to C&W's conclusions. First, working memory measures appear to be age-invariant. Second, the production of complex syntactic constructions appears not to be age-invariant.

Age invariance. C&W question whether the tasks commonly used to measure individual differences in working memory are reliable and stable. An additional concern, not considered by C&W, is whether the tasks are age invariant. As Stine (1995) notes, even when younger and older adults are matched based on working memory scores, investigators might inevitably find that older adults have processing difficulties that are seemingly unexplained by their performance on the working memory tasks.

Horn and McArdle (1992) explain that the latent variable approach to constructing measurement permits the test of two common concepts that receive little attention in intergenerational research: construct validity and construct invariance, across samples such as age groups. Horn and McArdle define measurement invariance as factorial invariance: the characteristic that measurements are composed of linear composites, that is, item scores summed to form a total score, is stable across measurement occasions of same or different samples. the authors present four criteria that are sufficient to determine factor invariance: (1) Factor patterns are equivalent across groups, a necessary condition for factor invariance to hold. (2) Factor variances are equivalent across groups. (3) Factor covariances are equivalent across groups. (4) Factor averages of composite scores are equivalent across groups. A fifth criterion they impose is whether cross-products are equivalent across groups. If the answer is "yes," then the answer to questions (1–4) is necessarily "yes"; if the answer is "no," invariance is not necessarily falsified.

To date, few studies have used a latent variable approach to assess the age invariance of working memory measures. As an example of this approach, Kemtes and Kemper (1998) assessed working memory performance for 248 adults (young and old) in four separate studies using a variety of verbal working memory tasks. With the latent variable approach, they tested whether a simple within-studies factor analytic structure with a single latent variable of working memory was reliable across age groups. For three of the studies, this single-latent-factor working memory model was validated across age groups, suggesting that a common construct was measured. Kemtes and Kemper's finding of factor invariance across these different samples strongly supports the conclusions that working memory, as measured by the digit spans, reading span, and listening span, is age invariant.

Age noninvariance. C&W also review the existing literature on the effects of aging on sentence processing and conclude that there are few age differences in on-line sentence processing although post-interpretative processes of question answering and text recall may be affected by aging. However, they overlook an additional body of research on age differences in production. Older adults show a reduction in their production of complex syntactic constructions such as those involving subordinate and embedded clauses (Kemper 1987; 1988; Kemper et al. 1989; 1992; Kynette & Kemper 1986). The age-related decline in syntactic production is somewhat greater for left-branching constructions (e.g., *The gal who runs a nursery school for our church is awfully young*) than for coordinate or right-branching constructions (e.g., *She's awfully young to be running a nursery school for our children*). This asymmetry in the production of right-branching constructions versus left-branching constructions provides strong evidence for the effects of working memory limitations. Left-branching constructions, including center-embedded object relative clauses such as *The dog that the man that the cat bit chased escaped*, are typically considered more complex than right-branching constructions such as *The cat bit the man that chased the dog that escaped* (Gibson et al. 1996; Lewis 1996b).

Cheung and Kemper (1992) investigated the relationship between age, working memory, and production using a number of different ways of measuring linguistic complexity including: mean length of utterance (MLU; Chapman & Miller 1984), developmental sentence scoring (DSS; Lee 1974), developmental level (DLEVEL; Rosenberg & Abbeduto 1987), two alternative ways of measuring Yngve depth (Yngve 1960), and two variants of Frazier's (1985) node count. In addition, propositional density (PDENSITY), based on Kintsch and Keenan's (1973) analyses of text difficulty, was computed in order to assess whether semantic content covaries with grammatical complexity. Cheung and Kemper (1993) applied structural equation modeling to these linguistic complexity metrics using language samples from younger and older adults. The best-fitting model fit the data by specifying two correlated factors, verbal ability and working memory. Age was negatively associated with working memory, leading to a decline in digit span with advancing age, and was somewhat positively associated with verbal ability, reflecting a slight improvement in vocabulary with advancing age. Working memory was related to three syntactic factors: length, measured by MLU; the amount of embedding, measured by MCU; and the type of embedding, measured by DSS and DLEVEL as well as by both Yngve depth metrics and both Frazier counts. Finally, verbal ability predicted another linguistic factor, semantic content, measured by PDENSITY, which was not correlated with the syntactic factors.

Working memory limitations associated with aging affect older adults' production of complex syntactic structures. A key determinant of syntactic complexity, affecting the DSS, DLEVEL, Yngve, and Frazier metrics, is whether embeddings occur in the main-clause subject of left-branching constructions such as "*Going to the St. Louis World Fair* was a major undertaking" or in the