Cognitive Gains and Losses of Patients With Alzheimer's Disease During Frequent Practice

Argye Elizabeth Hillis
Johns Hopkins University School of Medicine and The Department of Cognitive Science

Lisa Benzing
Cathy Epstein
Johns Hopkins University Department of Cognitive Science

Constantine Lyketsos
Johns Hopkins University School of Medicine, Department of Psychiatry

This longitudinal study of two patients with Alzheimer's disease (AD), on 12 experimental tasks, was designed to identify a measure of cognitive decline in AD patients that would be sensitive to change over short periods (weeks), resistant to practice (repeated testing over a few weeks), and predictive of change over longer periods on standardized tests. During the periods of frequent testing, both patients improved on most tasks, but performance on the tasks of spelling words and spelling pseudowords declined in both patients during periods when they deteriorated on standardized testing. It is concluded that simple spelling tasks may provide a good baseline of the rate of cognitive decline to identify effects of intervention, at least in some AD patients.

With the aging of the population, clinicians have become increasingly concerned with the management of patients with Alzheimer's disease (AD). This concern, coupled with the recent availability of medications that reportedly slow cognitive decline in patients with AD, presents clinicians with the challenge of measuring the gradual rates of cognitive change. That is, in order to identify the effects of an intervention (e.g., medication, family training), the rate of change prior to intervention must be established. The cognitive batteries currently in use are not designed to be given frequently, as they are subject to practice effects. For example, the Mini-Mental Status Examination (MMSE; Folstein, Folstein, & McHugh, 1975) is a standardized and widely used test that has been shown to reliably document change in global cognitive function in AD when given no more than once every few months. The purpose of this study was to identify a measure of cognitive ability that is resistant to practice effects and reflects decline even when administered twice weekly. This would allow one estimate of the rate of deterioration in a short time period. Such a measure should predict change on the MMSE or other global cognitive measures over longer periods. To date, eight patients with probable AD have begun testing on a battery of 12 candidate measures. This paper presents the preliminary results from the two subjects who have been tested over the course of 1 year.

Method

Subjects

The patients were referred from the Alzheimer's Disease Research Center at Johns Hopkins, where they had been diagnosed with probable AD as defined by the NINCDS-ADRDA research task force (McKhann et al., 1984) 2 years prior to this study. CCS is a 78-year-old, right-handed woman, who is a high school graduate and retired nurse's aide. LAT is a 78-year-old, right-handed male college graduate and retired engineer. Both patients had nonfocal neurological examinations and MRI scans that showed diffuse cortical atrophy and prominent ventricles. LAT's MRI scan also showed some increased signal intensity in the periventricular white matter, but a SPECT scan showed generalized cortical hypoperfusion, especially affecting both temporoparietal regions, which was interpreted as consistent with AD rather than multi-infarct dementia.

Testing

Tasks of language, memory, attention, and visuospatial skills were tested, since these skills have been found to deteriorate in AD patients (Bayles, 1994). Tasks with pseudoword stimuli were included, as these may be sensitive to deterioration in phonological processing in dementia. For each of the 12 tasks, 6 sets of stimuli were developed. The sets were equated in difficulty, as described in Table 1. Stimulus lists are available on request.

Responses were scored independently by 2 observers; interjudge reliability (measured by percent agreement) exceeded 95% in all tasks. These tasks were administered twice weekly over two periods of 18 sessions (9 weeks), with an intervening 9-week maintenance period during which the patient was seen 8 times and engaged in unrelated tasks[1]. With the exception of word pair learning, no feedback was provided during the tasks, other than general encouragement that was not contingent on response accuracy. The MMSE, the logical memory (story retelling) subtest of the Wechsler Memory

[1]CCS discontinued in the midst of the second 9-week series, because she moved to a nursing home where testing could not be carried out in a quiet setting. LAT dropped to one session per week after the maintenance period, so that the second set of 18 sessions stretched to 18 weeks.

 1058-0360/95/0404-0152

TABLE 1. Description of experimental tasks.

Task	Stimulus	Response	#Stimuli per set/session	Parameters matched across sets
Pseudoword spelling	3–4 phoneme, monosyllabic dictated pseudoword	stimulus written[a]	26	length in phonemes
Word spelling	1–3 syllable word	stimulus word written	34	word frequency[b], word class, #phonemes, #syllables, & #letters
Naming	Line drawing from Snodgrass & Vanderwart (1980)	spoken name	17	name frequency, syllables, object familiarity[c]
Word reading	1–3 syllable printed word in large block print	spoken word	34	word frequency, word class, #phonemes
Pseudoword reading	3–4 phoneme, monosyllabic dictated pseudoword	stimulus spoken[d]	26	length in phonemes
Pseudoword recall	List of 3 pseudowords, spoken slowly	stimuli repeated	8 lists	#phonemes, #syllables
Word recall	List of 3–4 words, spoken slowly	stimuli repeated	10 lists	#phonemes, #syllables, word frequency, & class
Word pair learning	8 word pairs, over 3 trials	2nd word in pair	24 words	word frequency & class
Visual memory	7 targets, mixed with 18 foil line drawings	identifies 7 targets seen previously	7 pictures	name frequency, familiarity, #syllables
"Stroop" test	names of 4 ink colors discordant with word	name of ink color	70	same 70 stimuli, different order on card
Cancellation task	2 arrays of letters	cross out target letter	21 targets	targets in same position in array
Digit-symbol recoding	"key" of 9 digit-letter pairs, list of digits in random order	writes letter corresponding to digit	26	same digits & letters, different pairings

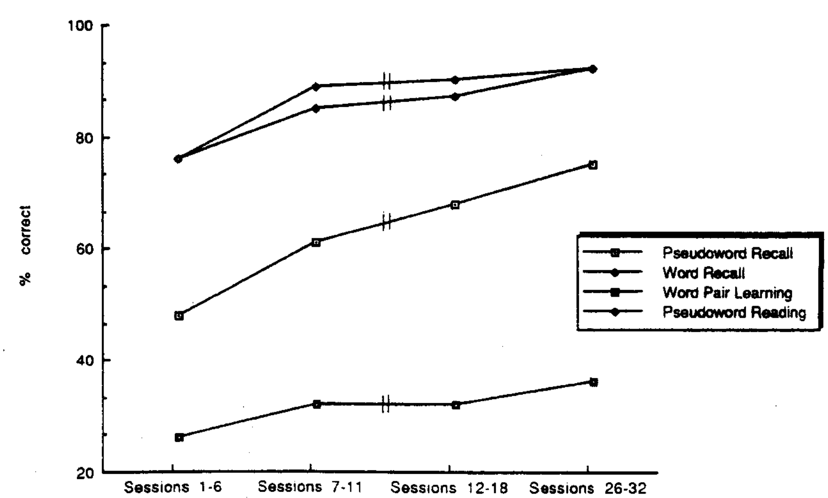
[a]For each phoneme in the string, any letter that corresponds to the stimulus phoneme in English, irrespective of the context, is accepted as correct.
[b]Frequency norms are from Kucera & Frances, 1967.
[c]Familiarity norms are from Snodgrass & Vanderwart, 1980.
[d]For each letter in the string, any phoneme that corresponds to the stimulus letter in English, irrespective of the context, is accepted as correct.

Scale (Wechsler, 1972), and formulation of verbal definitions (in which the patient was simply asked to define abstract and concrete words, given one at a time) were videotaped every 9 weeks (or 18 sessions). A subjective rating of functional cognition by naive judges was based on the videotapes of story retelling and definitions. The tapes of each patient were shown in random order, without information as to the dates or relative timing of the sessions. The judges were asked to rank order the videotapes from "least severe dementia" to "most severe dementia." They were permitted to rank two or more tapes as the same severity of dementia. They were not provided with any definition of dementia or other guidelines on which to base their judgments of dementia severity.

Results

Each patient's scores on each task were collapsed across 6 sessions (with the 6 different sets of stimuli). Stimuli were identical across comparisons. CCS showed steady improvement on all verbal memory tasks and on pseudoword reading, and these gains were maintained even after the period with no practice (Figure 1). The improvement from the first 6 sessions to the last 6 sessions was significant for word list recall ($X^2_1 = 10$; $p < .0001$), pseudoword list recall ($X^2_1 = 21$; $p << .0001$), and pseudoword reading ($X^2_1 = 13$; $p < .001$), but was not significant for word pair learning ($X^2_1 = 3.7$; $p = .05$; Fisher's exact: $p = .07$). For sustained attention tasks (spatial and verbal), performance steadily improved during the period of semiweekly practice, dropped off somewhat during the maintenance period, then began to im-

FIGURE 1. CCS's performance on verbal recall tasks and pseudoword reading.



prove again during practice (Figure 2, top). Gains during semiweekly practice were significant for each task ("Stroop" task: $X^2_1 = 43$, $p << .0001$; letter cancellation: $X^2_1 = 5.4$, $p < .02$; digit-symbol recoding: $X^2_1 = 13.5$, $p < .0002$). CCS's performance on confrontation naming, word reading, and visual recognition memory showed no significant change during the study (Figure 2, bottom). CCS's performance declined on word spelling ($X^2_1 = 4.3$, $p < .04$; Fisher's exact: $p < .05$) and pseudoword spelling ($X^2_1 = 36$; $p << .0001$). The decline in pseudoword spelling was significant between each pair of comparisons (i.e., between each set of 6 sessions and the next set of 6 sessions). The quality of her errors across sessions also reflected deterioration in spelling (see examples in Table 2). Graphs of performance across individual sessions for pseudoword spelling and MMSE scores (both of which deteriorated) contrasted to pseudoword reading are shown in Figure 3. MMSE scores are given as percent correct out of 30 points.

LAT showed nearly identical patterns of improvement (Figures 4 and 5). Like CCS, he showed steady gains on all verbal memory tasks and on pseudoword reading, and these gains were maintained over the maintenance period. The improvement from the first 6 to the last 6 sessions was significant for pseudoword list recall ($X^2_1 = 8.7$; $p < .01$) and pseudoword reading ($X^2_1 = 4.8$; $p < .03$). As found for CCS, performance on the "Stroop" task (a nonstandardized version developed by the authors; see Table 1) steadily improved during the period of semiweekly practice ($X^2_1 = 97$; $p << .0001$), then dropped off somewhat during the maintenance period, before improving again during the subsequent practice period. Digit-symbol recoding (a nonstandardized task developed by the authors; see Table 1) and letter cancellation were at ceiling levels of accuracy for all sessions, but his time to completion showed the same pattern of improvement. Just as for CCS, LAT's performance on confrontation naming, word reading, and visual recognition memory showed no significant change during the study. But in contrast to CCS, LAT showed no decline in performance on word or pseudoword spelling during the first 9 weeks, the 9-week maintenance period, or the second 9 weeks of practice, during which his MMSE scores were stable. However, his spelling deteriorated during the the final 9 weeks of practice, when his MMSE score also dropped (from 24/30 or 80% to 21/30 or 70%). The accuracy of spelling pseudowords dropped from 98%

FIGURE 2. CCS's performance on the "Stroop" task, cancellation task, and digit-symbol recoding (top) and on visual confrontation naming, word reading, and visual recognition memory (bottom).

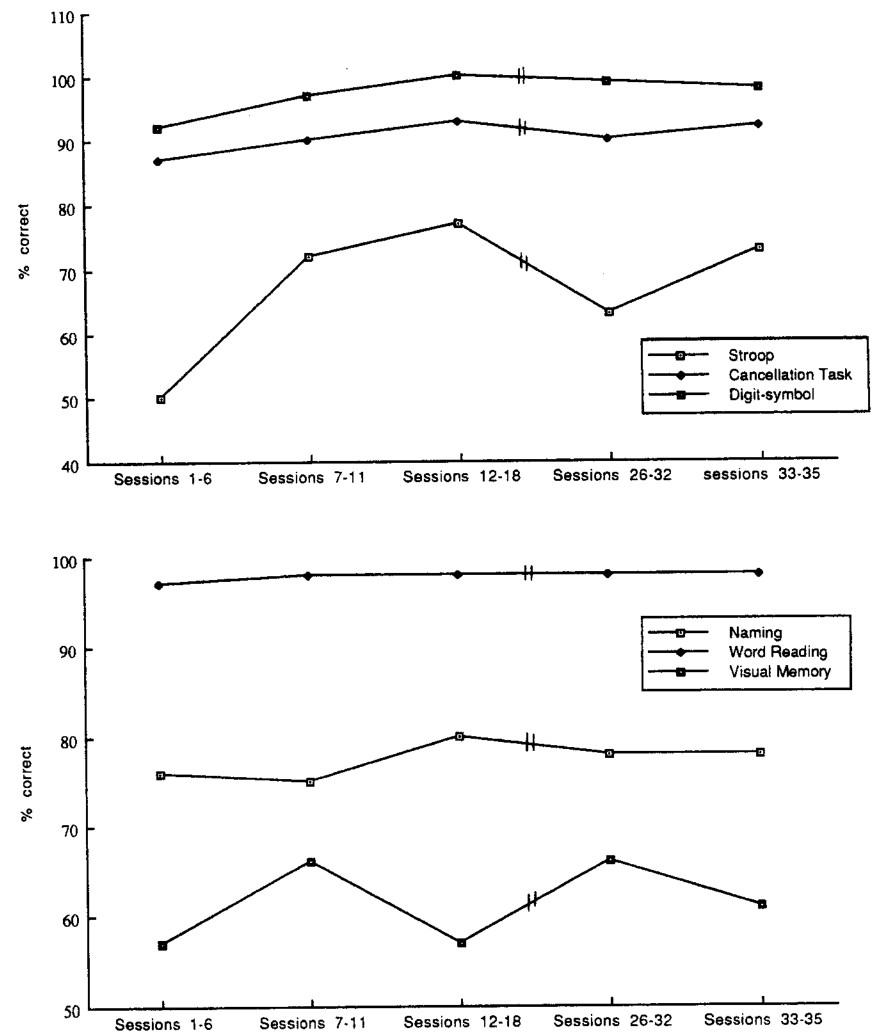
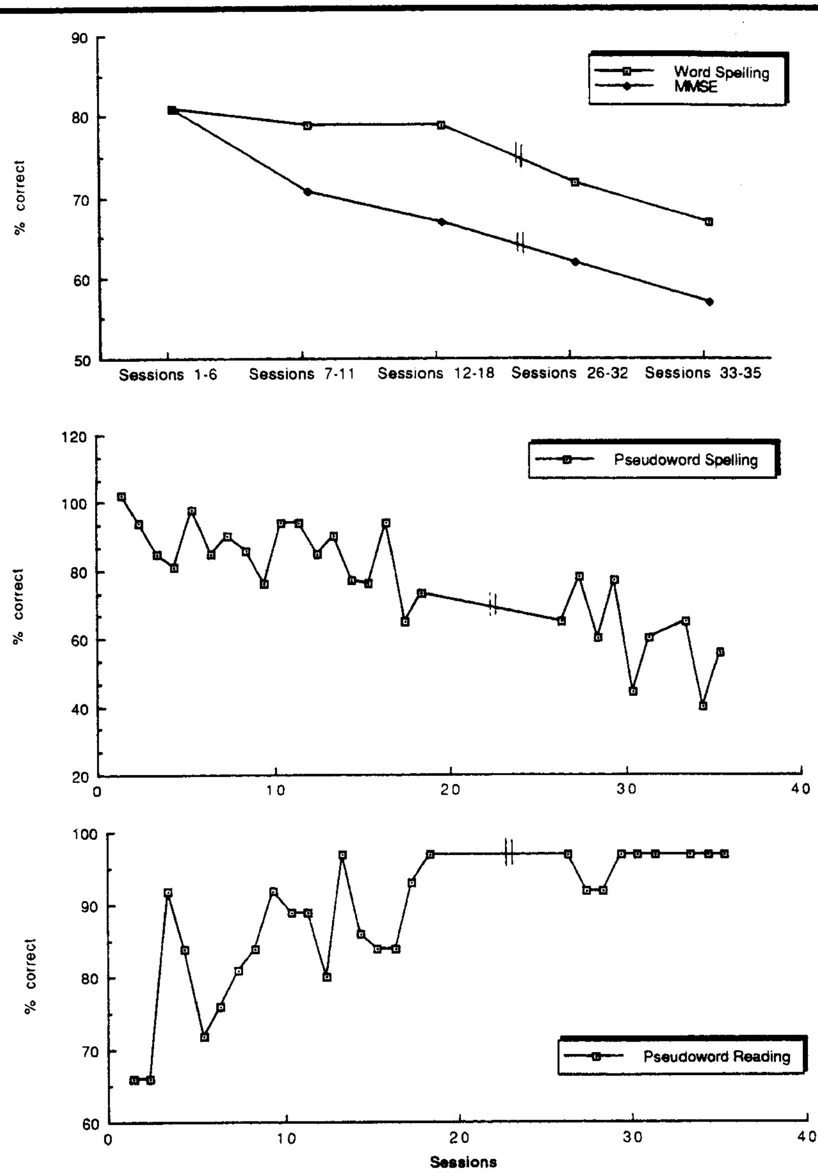


TABLE 2. Examples of CCS's spelling errors across practice periods.

	Stimulus	Response			
		Experimental Sessions			
		1–6	6–12	13–18	26–32
Words	pumpkin	pumkin	pumkin	pumpink	bupmink
	iron	iron	iorn	iorn	iorm
	exposure	exposer	expourse	exspore	exporise
	ritual	richuall	richall	richell	richella
	hybrid	hybred	hyberid	hyber	hibart
	screw	screw	skew	swew	swerex
Pseudowords	zeet /zit/	zeet	zeet	zet	jest
	pled /plɛd/	pled	plad	plade	blub

FIGURE 3. CCS's accuracy on the MMSE, spelling, and pseudoword reading.



to 92% between the penultimate and the ultimate series of 6 sessions ($X^2_1 = 4.8$; $p < .03$, Fisher's Exact: $p = .05$). The decrease from the first to the last series of 6 sessions was significant for both words ($X^2_1 = 3.7$; $p = .05$, Fisher's Exact: $p = .05$) and pseudowords ($X^2_1 = 8.8$; $p < .003$). Graphs of performance across individual sessions for word and pseudoword spelling and MMSE scores, contrasted to pseudoword reading, are shown in Figure 6.

Three naive judges viewed the videotapes of the story retelling and verbal definitions by CCS and LAT. They each independently rated CCS's dementia as least severe in test 1 and most severe in test 3; and rated LAT's dementia as unchanged across the first 3 tests and worse in the last test.

Conclusions and Implications

The only scores on tasks that showed decline over short periods and predicted overall deterioration as measured by the MMSE and subjective impression of dementia severity were spelling scores (especially pseudoword spelling). How do we account for the decline in pseudoword spelling despite practice and despite improvement in other cognitive tasks that were practiced? To answer this question, the cognitive mechanisms that underlie pseudoword spelling will be considered. Pseudoword spelling depends on sublexical spelling mechanisms: (a) phonological analysis, (b) converting sound to print (via "phoneme to grapheme conversion" mechanisms), (c) holding the yielded string of graphemes in a "graphemic buffer" while output processes are implemented, (d) converting from the abstract letter identity or grapheme to a particular letter shape ("allographic conversion"), and (e) motor writing processes (see Hillis & Caramazza, 1987). When damage to lexical-semantic mechanisms prevents access to the stored spelling of words, the patient may rely on sublexical mechanisms to spell not only pseudowords, but also regular words. Reliance on sublexical mechanisms would result in misspelling of irregular words, producing a response that is phonologically plausible (e.g., *scheme* spelled *skeam*). Platel et al. (1993) found that 22 AD patients they studied made phonologically plausible spelling errors on words and spelled pseudowords accurately on initial testing, indicating spared sublexical mechanisms but impaired lexical-semantic mechanisms (see also Rapcsak, Arthur, Biklen, & Rubens, 1989). When the same patients were retested later, Platel and coworkers found an increase in phonologically implausible errors and significantly more errors in spelling pseudowords, indicating a late decline in sublexical mechanisms. The authors lamented that they could not determine whether these later errors resulted from a deterioration in phoneme-to-grapheme conversion, the graphemic buffer, or allographic conversion. The progressively less plausible nature of errors by CCS (see Table 2) is also consistent with deterioration in any of these three mechanisms. However, further analyses of performance by CCS and LAT allowed us to distinguish among these possible causes of late-appearing phonologically implausible spellings and errors on pseudowords.

The possible accounts will be considered in turn. First, a decline in phonological analysis would predict a decline in recall of series of pseudowords, which was not demonstrated. Progressive damage to the allographic conversion would predict better oral than written spelling, since oral spelling does not require conversion from a grapheme to an allograph. But both patients showed identical accuracy rates in oral and written spelling. Decay of allographic conversion would also predict equal decline in spelling pseudowords and all types of words (since each grapheme must be converted to an allograph, irrespective of the type of stimulus). In contrast to this prediction, CCS and LAT both showed more rapid decline in spelling pseudowords than words and in spelling abstract words as opposed to concrete words. In fact, they both improved slightly in spelling concrete words (from 82% to 86% for CCS and 83% to 87% for LAT; NS by chi square), while they declined in spelling abstract words. The decline in spelling abstract words from 74% to 68% by CCS from the first to the last series was not significant; but the decline from 87% to 76% by LAT from the penultimate to the ultimate series of 6 sessions was statistically significant ($X^2_1 = 5.3$; $p < .03$; Fisher's Exact: $p < .03$). Such lexical effects cannot be accounted for by progressive impairment in phonological analysis, allographic conversion, or motor writing.

Deterioration of the graphemic buffer would also predict equal decline in spelling pseudowords and words and equal decline for abstract and concrete words of a given length (since the string of graphemes must be held in the buffer during output, whether it is a pseudoword, abstract word, or concrete word). As shown above, this prediction was not borne out in the performance by CCS or LAT. Further, a degeneration of the graphemic buffer would predict more rapid decline in spelling longer words than in spelling shorter words, since such a short term storage system is sensitive only to the number of items to be stored (Hillis & Caramazza, 1987). That is, longer words would have to be held longer in the buffer and would thus be subject to greater degradation in the face of dysfunction of the buffer. In contrast to this prediction, both patients showed a more rapid decline in spelling shorter words. CCS's spelling deteriorated on 4–5 letter words, remained stable on 6–7 letter words, and improved from 63% to 72% (NS) on 8–9 letter words. Similarly, an increase in errors on 4–5 letter words accounted for the majority (56%) of LAT's increased error rate over the final series. Finally, deterioration of the graphemic buffer would also predict a decline in *reading* pseudowords, since the graphemic buffer is needed not only for spelling words and pseudowords, but also for reading pseudowords (see Caramazza, Capasso, & Miceli, 1994). Pseudoword reading actually improved, rather than declined, in CCS and LAT. Thus, deterioration of the function of the graphemic buffer cannot account for the evolution of dysgraphia in these patients.

In contrast, the hypothesis of decay in phoneme-to-grapheme conversion would

FIGURE 4. LAT's performance on verbal recall tasks and pseudoword reading.

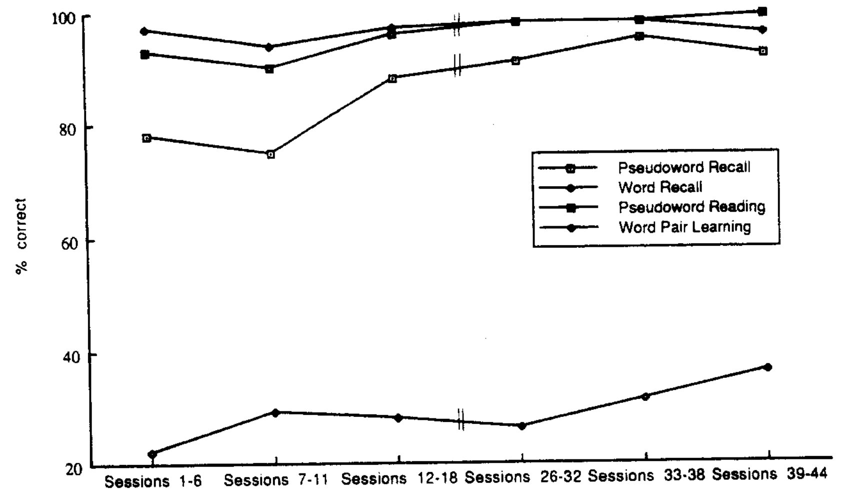
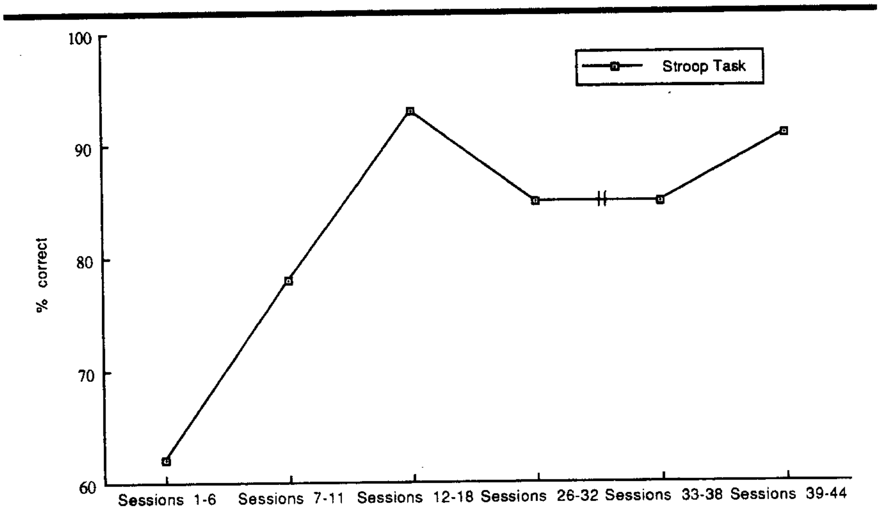


FIGURE 5. LAT's performance on the "Stroop" task (top) and visual confrontation naming, word reading, and visual recognition memory (bottom).



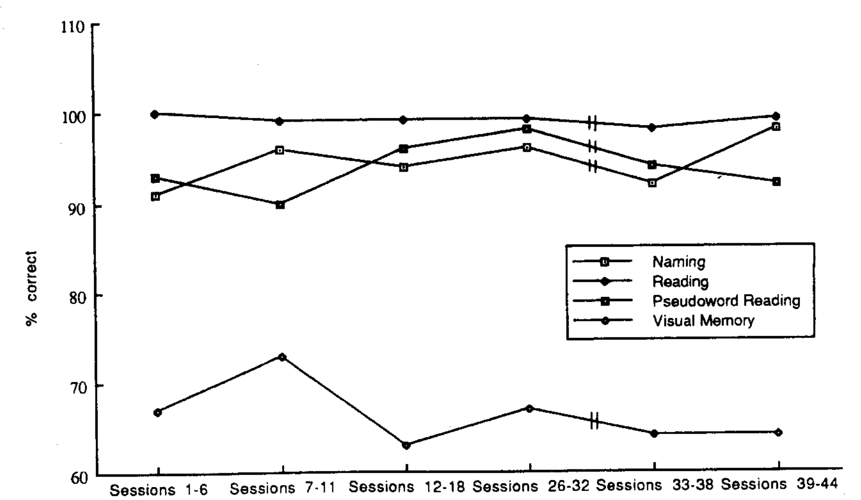
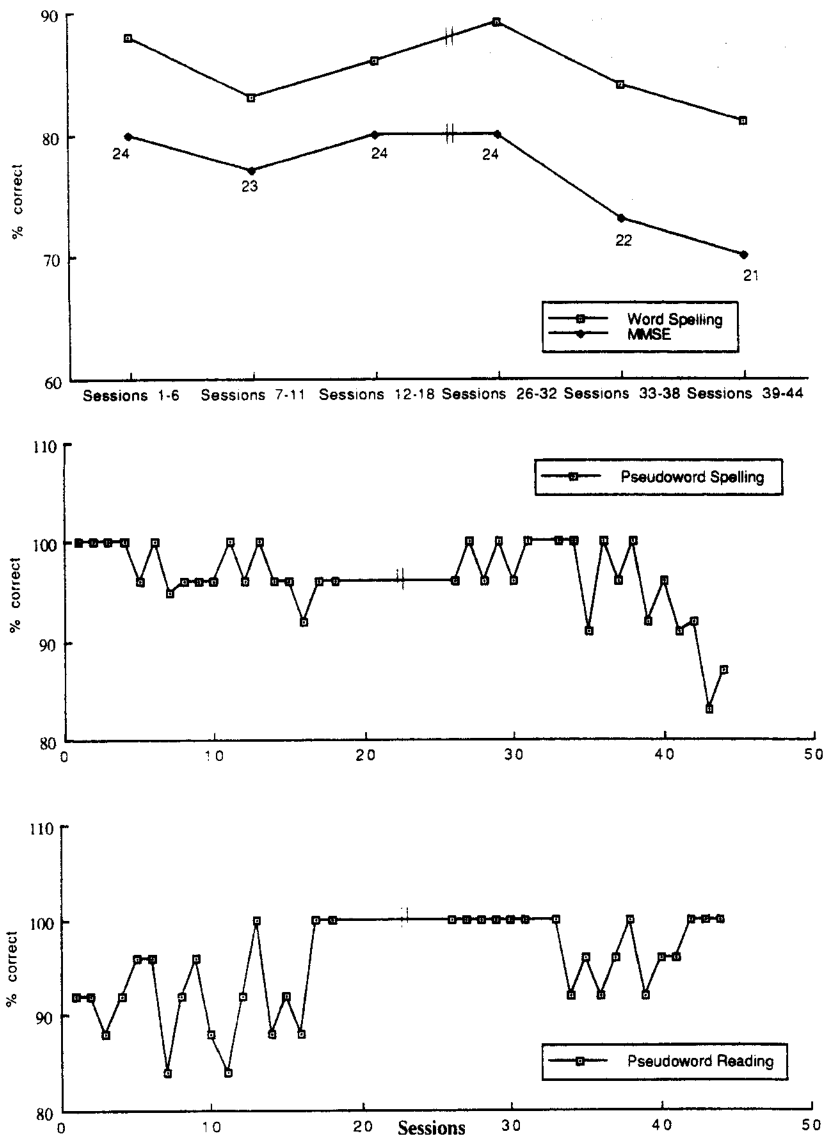


FIGURE 6. LAT's accuracy on the MMSE, spelling, and pseudoword reading.



account for the more rapid decline in pseudoword spelling than in word spelling, since only pseudoword spelling normally depends on sublexical, phoneme-to-grapheme conversion. Deterioration in phoneme-to-grapheme conversion can also account for a slow decline in word spelling in the presence of a partially impaired semantics, since access to the lexical representation (the stored spelling of the word) may depend on summation of partial information from the impaired semantic system and information from the phoneme-to-grapheme conversion mechanism (Hillis & Caramazza, 1991). On this hypothesis, words that cannot be accessed via the semantic system would be vulnerable to partial impairment of phoneme-to-grapheme conversion. Therefore, a deterioration of phoneme-to-grapheme conversion can alone account for the pattern of decline by CCS and LAT in spelling words and pseudowords despite practice.

In conclusion, results of this study and previous studies converge in support of the hypothesis that, at least in some AD patients, deterioration of phoneme-to-grapheme conversion occurs only relatively late in the disease. Since decline in other aspects of cognitive function is most rapid late in AD (Drachman, O'Donnell, Lew, & Swerer, 1990), it is plausible that a relatively late-appearing deficit such as deterioration in pseudoword spelling would be a sensitive marker of the beginning of more rapid deterioration in other cognitive functions.

Although previous authors have reported that writing impairments are associated with disease severity in AD, spelling itself is associated with severity in only a subset of patients (Henderson, Buckwalter, Sobel, Freed, & Diz, 1992; Horner, Heyman, Dawson, & Rogers, 1988). It follows that CCS and LAT might represent a subclass of AD patients (e.g., a group with more prominent language disturbance, as described by Binetti et al., 1993). Patients with prominent language impairment would be likely to show a decrease in MMSE scores, as shown by our patients, since the MMSE is heavily language-weighted. Therefore, a decline in spelling may be a sensitive predictor of more general decline only in some AD patients. Nevertheless, any documented decline in the simple test of pseudoword spelling over the course of a few weeks in an AD patient should provide a reasonable baseline to identify the effects of intervention (e.g., medical or behavioral treatment) in that patient.

An important adventitious finding was that both patients improved on all tasks of sustained attention during practice, although gains were not maintained without practice. Moreover, both patients improved on all tasks of verbal short-term memory and learning (and pseudoword reading), and the gains in these tasks were maintained even during the 9 weeks without practice. Notably, the patients actually improved, and did not simply maintain performance, on the practiced attention and recall tasks. CCS showed significant gains in these practiced tasks, despite the fact that she showed continued decline in unpracticed tasks (e.g., the MMSE). Therefore, there may be a role for therapy to improve performance on specific functional tasks that require sustained attention, short-term memory, and/or verbal learning even in AD patients during periods of progressive cognitive decline.

Acknowledgments

The research reported in this paper was supported by NIH grant (NINCD) RO1 19330 and by a Medical Student Scholarship from the Hartford Foundation/ American Federation on Aging Research. The authors gratefully acknowledge this support. We are also grateful to CCS and LAT for their cheerful participation in our studies and to two anonymous reviewers for helpful comments on an earlier version of the paper.

References

Bayles, K. A. (1994). Management of neurogenic communication disorders associated with dementia. In R. Chapey (Ed.), *Language intervention strategies in adult aphasia*. Baltimore: Williams and Wilkens, pp. 535–545.

Binetti, G., Magni, E., Padovani, A., Cappa, S., Bianchetti, A., & Trabucchi, M. (1993). Neuropsychological heterogeneity in mild Alzheimer's disease. *Dementia, 4*, 321–326.

Caramazza, A., Capasso, R., & Miceli, G. (1994). The role of the graphemic buffer in reading. *Reports of the Cognitive Neuropsychology Laboratory*. Dartmouth College.

Drachman, D., O'Donnell, B., Lew, R., & Swerer, J. (1990). The prognosis in Alzheimer's disease: 'How far' rather than 'how fast' predicts the course. *Archives of Neurology, 47*, 851–856.

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

Henderson, V. W., Buckwalter, J. G., Sobel, E., Freed, D. M., & Diz, M. M. (1992). The agraphia of Alzheimer's disease. *Neurology, 42*, 776–784.

Hillis, A. E., & Caramazza, A. (1987). Model-driven treatment of dysgraphia. *Clinical Aphasiology, 17*, 84–105.

Hillis, A. E., & Caramazza, A. (1991). Mechanisms for accessing lexical representations for output: Evidence from a category-specific semantic deficit. *Brain and Language, 40*, 106–144.

Horner, J., Heyman, A., Dawson, D., & Rogers, H. (1988). The relationship of agraphia to the severity of dementia in Alzheimer's disease. *Archives of Neurology, 45*, 760–763.

Kucera, H., & Francis, W. N. (1967). *A computational analysis of present-day American English*. Providence, Rhode Island: Brown University Press.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA workgroup under the auspices of Health and Human Services Task Force on Alzheimer's Disease. *Neurology (Cleveland), 34*, 939–944.

Platel, H., Lambert, J., Eustache, F., Cadet, B., Dary, M., Viado, F., & Lechevalier, B. (1993). Characteristics and evolution of writing impairment in Alzheimer's disease. *Neuropsychologia, 31*, 1147–1158.

Rapcsak, S. Z., Arthur, S. A., Biklen, D. A., & Rubens, A. B. (1989). Lexical agraphia in Alzheimer's disease. *Archives of Neurology, 46*, 65–68.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardised set of 260 pictures: Norms for name agreement, image agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 174–214.

Wechsler, D. (1972). *The Wechsler Memory Scale*. New York: The Psychological Corp.

Contact author: Argye Elizabeth Hillis, MD, Department of Medicine, Physicians' Pavilion East, Suite 203, Greater Baltimore Medical Center, 6565 North Charles Street, Baltimore, MD 21204

Key Words: dysgraphia, Alzheimer's disease, dementia