

노인 기억장애 검사(Elderly Memory-disorder Scale)의 검사-재검사에서 변화의 유의성*

김 호 영

최 진 영[†]

서울대학교 심리학과

기억기능은 노인 신경심리평가에서 매우 중요한 영역으로 객관적인 감퇴의 증거를 수집하기 위해 재평가가 이루어지기도 한다. 하지만 기억력 저하는 정상적인 노화과정에서도 흔히 보고되고, 동일한 검사를 반복 측정하면 연습효과도 자주 보고되고 있어 기억검사의 검사-재검사 상황에서는 관찰된 수행변화가 임상적으로 중요한 변화인지 판단할 수 있는 객관적인 기준이 필요하다. 본 연구에서는 정상 노인 111명을 대상으로 노인 기억장애 검사(Elderly Memory disorder Scale: EMS)의 검사-재검사 자료를 수집하여, 기억검사의 반복측정에서 수행 변화에 영향을 미치는 요인들을 탐색하고 변화의 유의성에 대한 기준 자료를 수립하였다. 그 결과, 대부분의 일화적 기억검사에서 기저 점수에 비해 재검사 점수가 유의미하게 향상되었지만 그 크기는 개인의 기저 수행 수준이나 인구 통계학적 특징에 따라 차별적으로 나타났다. 기저 수행이 낮은 사람일수록 수행이 더 많이 향상되는 경향이 있었으며, 기저 수행이 높은 경우 재검사 점수가 오히려 저하되는 경향을 보였다. 또한 기저 수행 수준을 통제했을 때, 교육 수준이 높을수록, 연령이 적을수록 연습효과가 더 큰 것으로 나타났다. 연습효과를 보정한 변화유의지수법(practice-adjusted RCI method; RCIp)과 표준화된 회귀기반법(Standardized Regression Based Change; SRB)에 의해 변화 유의성을 판단할 수 있는 90% 예측구간을 계산하여 제시하였고, EMS의 두 가지 장기 기억지수에 대한 90% 예측구간을 적용하여 감퇴자와 상승자를 각각 분류하였다. RCIp와 SRB 각 방법으로 분류된 집단들 간에 인구통계학적 변인이나 기저 일반 인지기능의 유의미한 차이는 관찰되지 않았다. 하지만 각 방법으로 분류된 집단들의 기저 기억 수행에 대해서는 다른 결과를 보였다. 즉, SRB에 의해 분류된 집단들 간에는 기저 기억 수행에 유의미한 차이가 없었지만, RCIp로 분류한 집단들 중에는 상승자 집단의 기저 기억 수행이 다른 집단들에 비해 유의미하게 낮게 나타났다. 이러한 결과를 볼 때, SRB가 RCIp보다 실제 재검사 자료의 분포를 합리적으로 반영하는 것으로 시사된다.

주요어 : 기억 평가, 연습효과, RCIp, SRB, 예측구간

* 이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No.2010-0008922).

[†] 교신저자: 최진영, 서울대학교 심리학과, 서울시 관악구 관악로 599
Fax : 02-877-6428 / E-mail : jyche@snu.ac.kr

시간 경과에 따른 인지기능의 변화를 측정하는 것은 신경심리평가의 중요한 역할 중 하나이다. 일반적으로는 인지기능 저하에 대한 정확한 평가를 위해 개인의 특성을 고려한 적절한 기준 사용을 권장하지만, 한 기준 집단 내에서도 개인차가 존재하기 때문에 종단적인 평가를 통해 개인 수준의 변화율을 측정하는 것이 보다 정확할 수 있다(Winblad, Palmer, Kivipelto, Jelic, Fratiglioni, Wahlund, Nordberg, Bäckman, Albert, Almkvist, Arai, Basun, Blennow, DeLeon, Decarli, Erkinjuntti, Giacobini, Graff, Hardy, Jack, Jorm, Ritchie, Vanduijn, Visser & Petersen, 2004). 노인기 인지장애를 일으키는 주요 신경정신과적 질환인 치매의 경우, 진단 기준 상 인지기능 감퇴에 대한 객관적인 증거를 필요로 하여(American Psychiatric Association, 1994) 정확한 진단을 위해 기능 변화에 대한 평가가 중요하다. 특히 치매의 핵심증상인 기억 기능의 저하는 정상적인 인지노화 과정에서도 흔히 보고되는 현상이어서 정확한 치매 진단을 위해서는 정상 수준의 기억저하와 병리적인 수준의 기억장애를 감별할 수 있는 판단 기준이 필요하다.

신경심리검사의 재평가에서 검사 점수의 변화를 해석할 때 고려해야 할 일반적인 요인들은 크게 편향(bias)과 오류(error)로 나누어 볼 수 있다(Chelune, 2003). 편향은 실제 수행 능력에서의 체계적인 변화를 반영하는 데 비해, 오류는 측정도구의 불완전함에서 기인하는 점수 변화를 이룬다. 편향에는 치료적 개입, 장애의 진행 혹은 회복으로 인한 수행 변화와 같이 연구자나 임상가의 관심 사건이 검사 수행에 미치는 영향이 포함되지만, 반복 측정으로 인한 연습 효과도 포함된다. 인지기능평가의 연습 효과는 사전경험으로 인한 검사 자극

에 대한 기억, 검사 절차에 대한 학습, 그리고 검사 상황에 대한 친숙성의 영향을 고려해 볼 수 있다. 특히 일화적 기억검사의 경우, 검사 자극에 대한 기억이 직접적인 측정 대상이기 때문에 연습효과에 더욱 취약할 수밖에 없다. 또한 경험으로부터 검사 절차나 상황에 대해 학습하는 정도는 연령, 교육, 성별 등의 인구통계학적 특징과 기저 인지 수행 수준에 따라서 다를 수 있다. 실제 수행 능력의 변화가 없음에도 불구하고 측정 도구의 신뢰도가 불완전하여 발생하는 측정 오류와 기저 수행수준에 따라 재평가 시 수행의 변화가 차별적으로 발생하는 평균으로의 회귀 현상과 같은 통계적인 오류도 검사-재검사 점수 차이를 발생시키는 주요 요인이다. 따라서 어떤 검사가 관심 영역의 실제 변화를 탐지하는 데 유용하게 쓰이려면 연구자가 의도하지 않은 연습 효과와 통계적 오류까지 감안하여 점수 변화의 유의성을 판단할 수 있는 기준이 필요하다. 중추신경계의 두드러진 변화를 경험하지 않은 정상 노인들이라면 동일한 검사에 대한 반복 경험과 인지노화로 인한 체계적 편향과 여러 오류가 검사-재검사 점수 변화의 주요 요인으로 고려될 수 있으며, 이러한 요인을 고려한 후에도 유의미한 변화를 보일 때 비로소 임상적으로 의미있는 변화로 간주할 수 있다.

기억 영역의 재평가 과정에서 연습 효과는 다시 동일한 과제 수행 절차에 반복적으로 노출되면서 검사 수행 전략을 발달시키는 검사 특장적 연습(test specific practice)과 같은 정보를 반복적으로 기억하게 함으로써 발생하는 문항 특장적 연습(item specific practice)으로 구분하여 생각할 수 있다(Anastasi, 1988). Benedict와 Zgajardic(1998)은 절차는 동일하지만 문항은 다른 동형검사 사용함으로써 기억기능의 검사-

재검사에서 연습효과를 약화시킬 수 있는 지 살펴보았다. 그 결과, 언어적 기억검사는 동형검사의 사용으로 연습 효과를 유의미하지 않은 수준으로 낮출 수 있었지만, 비언어적 기억검사에서는 동형검사 사용에도 불구하고 여전히 연습 효과가 유의미하게 나타났다. 언어적 자극은 비언어적 자극에 비해 그 내용에 대한 기억이 용이하여 문항 특정적 연습 효과에 취약할 수 있다. 이 연구 결과는 동형검사의 사용으로 문항 특정적 연습 효과는 어느 정도 감소시킬 수 있지만 과제 수행 전략과 같은 절차 학습이나 전체 연습 효과에 기여하는 다양한 요인들을 모두 통제하지는 못함을 시사한다(Bush, Chelune, & Suchy, 2006). 이 때문에 신경심리검사의 재평가 시 동형검사를 사용해도 연습 효과가 여전히 유의미하게 나타나는 것이다(Beglinger, Gaydos, Tangphao-Daniels, Duff, Kareken, Crawford, Fastenau, & Siemers, 2005). 또한 동형검사로는 신뢰도, 측정오류와 같은 심리 측정적 문제를 다룰 수 없어, 검사-재검사 시 개입되는 편향과 오류를 고려하여 변화의 유의성을 평가할 수 있는 여러 방법들이 개발되었다.

신경심리평가의 검사-재검사에서 변화의 유의성을 평가하는 방법으로 다양한 방법들이 제안되고 있지만 여러 경험적인 연구들에서 다음의 두 방법이 가장 신뢰할 만한 방법으로 추천되고 있다(Fridrichs & Tuokko, 2005; Lineweaver & Chelune, 2003; Temkin, Heaton, Grant, & Dikmen, 1999): 연습효과를 보정한 변화유의지수법(practice-adjusted Reliable Change Index method; 이하 RCIP, Chelune, Naugle, Luders, Sedlak, & Awad, 1993), 표준화 회귀 기반법(Standardized Regression Based Change; 이하 SRB, McSweeney, Naugle, Chelune, & Luders,

1993). RCIP는 기준 집단의 검사-재검사 자료에 근거하여 해당 영역의 실제적인 변화가 없음에도 불구하고 나타날 수 있는 검사-재검사 점수 변화의 분포를 추정하고, 이 분포를 이용하여 관찰된 검사-재검사 변화가 임상적으로 중요한 사건이 발생하지 않은 때에도 일어날 확률을 계산하여 변화의 유의성을 판단할 수 있게 해준다. 연습효과로 인한 전반적인 수행 향상을 보정하기 위해 초기 점수에다 기준 집단에서 계산된 점수 변화의 평균값을 더하여 재검사 점수를 예측하고, 이 예측값과 실제 관찰된 재검사 점수의 차이가 검사-재검사 점수 차이의 표준오차를 고려할 때 유의미한 수준인지를 판단한다(Chelune et al., 1993). 이 방법은 전체 집단의 평균적인 연습 효과와 측정오류는 고려할 수 있지만, 평균으로의 회귀 현상과 연습 효과 크기에 대한 인구 통계학적 요인의 조절효과는 고려하지 못한다.

SRB는 RCIP의 이러한 제한점을 보완하기 위해 McSweeney 등(1993)이 제안한 방법으로, 검사-재검사 기준 자료를 이용하여 해당 검사의 기저 점수와 인구 통계학적 변수들을 예측 변수로 재검사 점수를 예측하는 다중회귀 방정식을 구하고 이 식에 따라 개인의 재검사 점수를 예측한다. 이러한 절차를 통해 점수 변화에 대한 인구 통계학적 변수의 조절 효과와 평균으로의 회귀 현상을 동시에 고려할 수 있다. 회귀식에 따라 추정된 재검사 점수와 실제 관찰된 재검사 점수 간 차이의 유의성은 그 회귀식의 추정의 표준 오차를 고려하여 판단한다. 연구방법의 자료분석 부분에서 각 방법에 대한 구체적인 공식을 제시하였다(공식 [가], [나]).

Lineweaver와 Chelune(2003)은 정상인 281명을 대상으로 Wechsler Memory Scale-III(WMS-III, the

Psychological Corporation, 1997)의 검사-재검사에 따른 수행 변화를 SRB 방법을 적용하여 분석하였다. 그 결과, 모든 기억 요인에서 기저 점수가 재검사 수행의 분산을 가장 많이 설명하였으며, 기저점수에 따라 재검사에 따른 수행 변화가 다르게 나타났다. 또한 시각 기억과 청각 기억에서는 즉각회상과 지연회상에서 모두 전체 지능이 높을수록 반복 측정에 따른 수행 향상이 더 크게 나타났고, 시각 기억의 경우 연령이 높을수록 수행 향상이 덜 나타나는 경향을 보였다. 한편 검사-재검사 수행 간의 간격은 WMS-III의 재검사에 점수에 유의미한 영향을 미치지 않는 것으로 나타났는데, 다른 연구들에서도 검사-재검사 간격은 재검사 수행에 큰 영향을 미치지 않았다(Basso, Carona, Lowery, & Axelrod, 2002; McSweeney et al., 1993).

Frerichs와 Tuokko(2005)는 노인 기억평가를 중심으로 변화의 유의성을 평가하는 여러 가지 방법들의 유용성을 비교분석한 바 있다. 이들은 정상노인을 대상으로 수집된 검사-재검사 자료에 입각하여 각 분석방법을 이용하여 변화에 대한 90% 예측구간을 계산하였다. 이를 적용할 때, 정상 집단에서 유의미하게 상승, 혹은 감퇴한 것으로 분류되는 비율을 통계적 기대치와 비교분석하고 나아가 이 예측구간이 정상에서 치매로 이환된 집단을 감퇴 집단으로 정확히 분류할 수 있는지를 탐색하였다. 세부 기억검사에 따라 다소 차이가 있었지만, RCIP와 SRB를 적용하여 유의미하게 상승 혹은 감퇴한 것으로 분류되는 정상인의 비율은 대체로 통계적 기대치와 유의미한 차이를 보이지 않았다. 또한 SRB가 정상에서 치매로 이환된 노인들과 치매로 이환되지 않은 노인들을 가장 잘 구분해주는 방법으로 확인

되었다.

본 연구에서는 정상 노인을 대상으로 노인기 기억기능을 평가하는 도구 중 하나인 노인 기억장애 검사(Elderly Memory disorder Scale, 최진영, 2007)의 검사-재검사 수행 자료를 수집하여 RCIP와 SRB에 의한 변화의 유의성 규준을 각각 수립하고 두 방법에 따라 분류된 결과를 비교하고자 한다. 먼저 정상 노인들의 기억점수 변화에 영향을 미치는 요인과 각 기억검사들에서 나타나는 변화 양상을 탐색한 후, RCIP와 SRB에 따라 변화유의성 판단할 수 있는 예측 구간을 계산하고 각각의 방법으로 극단적인 상승 혹은 감퇴를 보이는 사람들을 적절히 구분할 수 있는지 그 가능성을 살펴보았다.

방 법

연구대상

본 연구는 서울대학교 임상신경과학연구소에서 2004년에서 2008년까지 서울 및 수도권 지역사회 노인을 대상으로 수행한 인지노화 종단연구에 2회 이상 참여하였고, 다음의 배제 조건에 해당하지 않는 60세 이상의 정상 노인을 대상으로 수행되었다. 초기 검사에는 총 214명이 참여하였으나 다음 배제조건에 의해 49명이 배제되었다: 1) 검사 당시 치매를 포함한 신경과적 병력이나 주요 정신장애의 병력이 있는 자, 2) 두부 외상이나 뇌수술과 같은 신경외과적 병력이 있는 자, 3) 약물로도 조절이 안 되는 심한 당뇨와 고혈압 병력이 있는 자, 4) 안경이나 보청기와 같은 보조기 사용으로도 검사 실시가 불가능할 정도의 심

한 감각 기능 손상이 있는 자. 두 번째 검사 시에 파킨슨씨 병, 뇌혈관 질환, 뇌졸중, 뇌종양 등의 신경과적 문제나 정신과적 문제가 추가로 확인된 자와 재검사 후 1년 이내에 신경학적 문제가 발생한 자도 추가로 배제되었다. 이러한 사유로 추가 배제된 사람, 두 번째 검사에 참여하지 않은 사람, 그리고 두 검사 시점 사이에 일부 검사에 노출되어 자료가 오염된 사람은 모두 54명으로 이들을 제외한 최종 분석 대상은 총 111명이었다. 최종 분석 표본의 인구 통계학적 특징과 검사 재검사 간격은 표 1과 같다. 신경심리 평가는 매년 실시되었

으며, 참여자에 사정에 따라 실시 간격이 1년 단위로 통일되지 않아 검사 간격을 개월 단위로 계산하였다. 본 연구는 서울대학교 생명윤리심의 위원회의 심의를 거쳐 수행되었으며, 모든 참여자는 연구 참여에 앞서 연구에 대한 연구자의 설명을 듣고 서면 동의서를 통해 참여를 수락하였다.

측정도구

한국판 치매평가 검사(Korean-Dementia Rating Scale; K-DRS, 최진영, 1998)

K-DRS는 한국 노인을 위해 DRS(Mattis, 1988)를 국내 사정에 맞게 수정하여 표준화한 치매 평가도구이다. 이 검사는 만 55세 이상 성인의 인지기능을 평가하는데 타당하고 신뢰로운 검사로 입증되었으며(Chey, Na, Park, Park, & Lee, 1999), 주의, 관리, 개념화, 구성, 기억 소검사로 구성되어 비교적 다양한 인지 기능을 평가한다. K-DRS 총점은 144점 만점으로 본 연구에서는 노인들의 일반 인지기능을 측정하기 위해 사용되었다.

노인용 기억장애 검사(Elderly Memory disorder Scale; EMS, 최진영, 2007)

EMS는 노인기 기억장애를 정밀하게 평가하기 위해 국내에서 개발된 기억검사로 국내 노인들의 인지 및 기억 특성을 고려하였으며, 기억을 가장 효율적으로 측정할 수 있는 기억 평가 패러다임에 기초하여 개발되었다. 이 검사는 일화적 기억능력 평가를 주목적으로 하며, 언어적, 비언어적 기억 검사가 주축이 되고 그 밖에 작업기억과 시공간 능력, 개념화 능력, 언어적 기능을 간단히 평가할 수 있는 검사들을 포함하고 있다. 언어적 기억 검사에

표 1. 전체 연구 참여자의 인구 통계학적 정보 (N=111)

변인	명(%)	M (SD)
초기 검사 연령		69.09 (4.80)
60 - 69세	67 (60.4)	
70 - 79세	39 (35.1)	
80세 이상	5 (4.5)	
교육		9.45 (4.32)
무학	3 (2.7)	
1-6년	33 (29.7)	
7-12년	54 (48.7)	
12년 이상	21 (18.9)	
성별		
남	14 (12.6)	
여	97 (87.4)	
검사-재검사 간격		18.16 (11.05)
7-15개월	77 (69.4)	
22-25개월	18 (16.2)	
34-42개월	7 (6.3)	
43-50개월	9 (8.1)	

는 노인 언어학습검사(Elderly Verbal Learning Test; EVLT)와 이야기 회상검사(Story Recall Test; SRT)가 포함되었고, 시공간 구성능력 및 비언어적 기억 평가를 위해 단순 Rey 도형검사(Simple Rey Figure Test; SRFT)가 포함되어 있으며, 각 기억 검사는 즉각회상, 지연회상, 재인 능력을 평가할 수 있는 측정치들로 구성된다. 또한 작업기억 및 단기기억을 측정하기 위해 숫자폭 검사(Digit Span Test; DST)와 시공간폭 검사(Spatial Span Test; SST)가 포함되어 있고, 시계그리기검사(Clock Drawing Test; CDT)와 단축형 보스턴 명명검사(Korean Boston Naming Test: K-BNT)도 포함되어 있다. 또한 여러 기억 검사 결과를 통합하여 장기기억 능력을 평가할 수 있는 장기 회상지수와 장기 재인지수를 계산할 수 있다. 본 연구에서는 근사값 사용으로 인한 계산오차의 영향을 축소하기 위해, 노인 기억장애 검사 실시요강에 제시된 공식으로 계산된 장기기억 지수에 100을 곱한 값으로 분석하였다.

자료분석

본 연구에서는 전체 종단 자료 중 검사-재검사 간격에 상관없이 처음 2회의 신경심리평가 자료가 분석되었으며, 검사-재검사 간격의 범위는 7-50개월이었다. 모든 통계적 분석은 윈도우용 SPSS 12.0으로 실행되었다.

우선 K-DRS 총점과 EMS 각 검사에서 기저 수행과 재검사 수행 간의 차이의 유의성을 대응표본 *t* 검정을 통해 살펴보았다. 또한 EMS 각 검사에서 차이값을 종속변수로 할 때, 이에 영향을 미치는 변수를 단계적 회귀 절차를 통해 탐색하였다. 단계적 회귀 분석에서 예측 변수로 투입된 변수는 EMS 각 검사의 기저

점수, 검사-재검사 간격, K-DRS 총점, 그리고 연령, 교육 연한, 성별의 인구 통계학적 변수들이다. 각 단계에서 변수를 선택할 때 유의수준은 .05, 제거할 때 유의수준은 .10으로 지정하였다.

변화 유의성 평가

기저 검사와 재검사에서 변화의 유의성을 평가하는 방법으로는 연습효과를 보정한 변화 유의지수법(practice-adjusted RCI method; RCIp, Chelune et al., 1993)과 표준화 회귀기반법(SRB; McSweeney et al., 1993)을 사용하였다.

RCI_p는 Temkin 등(1999)이 사용한 방식에 따라 계산되었으며, 그 공식은 다음과 같다(공식 [가]).

$$RCI_p = \frac{T_2 - (T_1 + D)}{SD_{diff}} \quad [가]$$

(단, T_1 =기저수행, T_2 =재검사 수행, $D=T_2-T_1$ 의 평균값, $SD_{diff}=T_2-T_1$ 의 표준편차)

검사-재검사 점수의 차이값이 정규 분포한다고 가정할 때, 추정되는 재검사 점수의 예측구간을 평가자가 설정한 기준에 따라 계산할 수 있다. 예를 들어 90% 예측구간은 재검사 점수의 기대값(T_1+D)에서 $<1.645 \times SD_{diff}>$ 이내의 구간으로 계산할 수 있다. 90% 예측구간을 적용하면 임상적으로 중요한 변화가 발생하지 않은 정상 집단에서 재검사 점수가 예측구간의 하한선 이하이거나 상한선 이상일 가능성은 각각 5% 이하이다. 본 연구에서도 90% 예측구간을 적용하여 분석하였는데, 이 경우 상승자나 감퇴자를 분류할 때 정상 집단에서 예상되는 유형 1 오류는 각각 5%이다. 개인의 관찰된 재검사 점수(T_2)가 이 예측 구

간에 포함되지 않는다면 그 개인은 재검사에서 유의미한 수행 변화를 보였다고 해석할 수 있다. 임상가가 변화 유의성을 판단하는 기준을 보다 엄격하게 설정하고자 한다면 SD_{diff} 에 1.645보다 더 큰 값을 곱하여 예측 구간을 구할 수 있다. 예를 들어 1.645대신 1.960을 곱하면 95% 예측구간을 구할 수 있다.

SRB 모델에서는 실제 재검사 점수(T_2)와 회귀 방정식에 의해 예측된 재검사 점수(T'_2)의 차이값을 회귀 모형의 추정의 표준오차(SE_{est})로 나눈으로써 표준화하고, 이 값을 통해 변화의 유의성을 해석한다. 이를 공식화 하면 다음과 같다(공식 [나]).

$$SRB = \frac{T'_2 - T_2}{SE_{est}} \quad [나]$$

(단, T'_2 =추정된 재검사 수행, SE_{est} =추정의 표준오차)

본 연구에서는 단계적 회귀 절차를 이용하여 재검사 점수를 예측하는 회귀 방정식을 도출하였으며, 이때 연령, 교육 연한, 성별의 인구통계학적 변수들과 검사-재검사 간격, 그리고 선행 연구 결과에 따라 각 검사의 기저 점수, K-DRS 총점이 후보 예측 변수로 포함되었다. 예측 변수들은 단계적 회귀 절차를 통해 선택되었으며, 선택할 때 유의수준은 .05, 제거할 때 유의수준은 .10으로 지정하였다. 이렇게 도출된 회귀방정을 이용하여 예측된 재검사 점수(T'_2)에서 $<1.645 \times SE_{est}>$ 이내의 구간으로 90% 예측구간을 계산할 수 있으며, RCIp에서와 마찬가지로 재검사 점수가 이 예측구간에 포함되지 않는다면 재검사 시 기능이 유의미하게 향상 혹은 감퇴된 것으로 해석할 수 있다. 위 공식들로 계산된 RCIp값과 SRB 값은

모두 표준화된 값으로 z점수와 같이 해석할 수 있다.

결 과

검사-재검사 시 수행 변화에 영향을 미치는 요인

검사-재검사간 점수 변화의 유의성을 확인하기 위해 두 대응표본 평균의 차이에 대한 t 검정을 한 결과는 표 2와 같다. K-DRS 총점은 기저점수에 비해 재검사에서 유의미한 상승을 보였다($p < .01$). EVLT와 SRFT에서는 즉각회상과 지연자유회상에서는 유의미한 상승을 보였으나 재인 과제에서는 유의미한 향상을 보이지 않았다($p_s > .05$). 반면 SRT는 재인 점수에서만 유의미하게 향상되었고($p < .001$) 즉각회상과 지연회상에서는 유의미한 향상을 보이지 않았다($p_s > .05$). 즉각회상과 자유회상 점수를 이야기 단위 점수와 주제단위 점수를 세분화하여 분석하면 즉각회상의 이야기 단위 점수에서만 유의미하게 상승하였다($p < .05$). EVLT의 학습목록 즉각회상은 5차례의 즉각회상 시행 중 1, 2차 시행만 유의미한 향상을 보였으며($p < .05$), 앞 시행일 수록 연습효과가 크고 후반부로 갈 수록 연습효과가 감소하는 양상이었다. DST와 SST의 시행점수는 바로따라하기와 거꾸로 따라하기에서 모두 유의미한 상승이 관찰되지 않았으며, DST 바로 따라하기에서는 오히려 유의미하게 수행이 감소하였다. CDT 총점은 유의미하게 수행이 향상되었고, SRFT의 복사시행과 단축형 K-BNT는 검사-재검사에 따른 수행의 변화가 거의 없었다($p_s > .20$). 한편 장기 기억 지수에서는 두 지수 모두에서 재검사에

표 2. 검사-재검사 결과에 대한 기술 통계량(N=111)

검사	기저 검사(T_1)		재검사(T_2)		$T_2 - T_1$		t	df	p -value ^a
	M	SD	M	SD	M	SD			
K-DRS	131.45	6.54	132.86	7.34	1.41	5.89	2.53	110	0.01
EVLТ									
즉각총점	31.41	4.84	32.90	4.94	1.50	4.59	3.43	110	0.00
단기지연	6.10	1.80	6.72	1.43	0.62	1.56	4.21	110	0.00
장기지연	6.17	2.00	6.84	1.56	0.67	1.59	4.43	110	0.00
재 인	27.63	2.02	27.95	1.82	0.32	1.71	1.95	110	0.05
SRFT									
복 사	14.79	1.32	14.91	1.12	0.11	1.36	0.87	110	0.38
즉각회상	11.36	2.81	12.55	2.59	1.19	2.63	4.78	110	0.00
지연회상	11.33	2.70	12.45	2.37	1.12	2.62	4.51	110	0.00
재 인	17.33	1.69	17.64	1.80	0.31	2.09	1.53	106	0.13
SRT									
즉각회상	15.58	5.40	16.30	4.76	0.73	4.21	1.81	110	0.07
지연회상	14.15	5.62	14.66	5.05	0.51	5.05	1.07	110	0.29
재 인	7.59	1.38	8.07	1.19	0.48	1.48	3.41	109	0.00
DST									
바 로	9.27	2.17	8.45	2.17	-0.82	2.09	-4.10	109	0.00
거 꾸 로	4.73	1.98	4.91	1.86	0.18	1.56	1.17	107	0.24
SST									
바 로	6.58	1.56	6.79	1.70	0.22	1.77	1.26	105	0.21
거 꾸 로	5.03	1.77	5.33	2.11	0.30	1.83	1.66	103	0.10
CDT	18.40	2.90	19.31	1.63	0.91	2.34	4.07	108	0.00
K-BNT	11.49	2.22	11.68	2.12	0.19	1.68	1.19	109	0.24
장기회상지수 ^b	186.56	42.83	202.69	33.20	16.13	30.19	5.63	110	0.00
장기재인지수 ^b	254.26	21.13	261.88	17.21	7.63	19.59	4.01	105	0.00

주. 모든 표에서 검사명의 축약어는 다음과 같이 통일됨: KDRS=한국치매평가검사(Korean Dementia Rating Scale), EVLT=노인언어학습검사(Elderly Verbal Learning Test), SRT=이야기 회상검사(Story Recall Test), SRFT=단순 Rey 도형검사(Simple Rey Figure Test), DST=숫자폭 검사(Digit Span Test), SST=시공간폭 검사(Spatial Span Test), CDT=시계그리기검사(Clock Drawing Test), K-BNT=단축형 보스턴 명명검사(Korean Boston Naming Test), EVLT 즉각총점은 학습목록 즉각회상 1-5의 총합이고 장기지연과 단기지연은 모두 자유회상, DST와 SST는 시행점수
^a양측 검정 시 p 값 ^b장기회상지수와 장기재인지수는 EMS 실시요강에서 제시한 공식에 100을 곱한 값

표 3. 재검사에 따른 점수 변화량(T2-T1)에 대한 단계적 회귀분석 결과(N=111)

검사	R ²	R ² 변화량 ^a					검사간격
		검사1	KDRS1	교육	연령	성별 ^b	
EVLT							
즉각총점	.321	.205***	.042*		.074**		
단기지연	.413	.413***					
장기지연	.455	.417***				.039**	
재 인	.287	.287***					
SRFT							
복 사	.463	.427***		.036**			
즉 각	.332	.300***			.032*		
지 연	.399	.365***			.034*		
재 인	.317	.317***					
SRT							
즉각회상	.326	.264***			.042*		
지연회상	.349	.309***		.040*			
재 인	.497	.427***	.070***				
DST							
바 로	.328	.232***	.056**				.040*
거 꾸 로	.222	.222***					
SST							
바 로	.315	.235***	.080**				
거 꾸 로	.221	.097***	.124***				
CDT	.778	.686***	.071***		.009*		.012*
K-BNT	.245	.195***	.050**				
장기회상지수	.441	.404***			.037**		
장기재인지수	.474	.416***	.058**				

주. ^a 해당 값의 크기 순서로 회귀모형에 변수가 추가될 때 설명력의 증가분. 단, SST 거꾸로 따라하기의 경우 검사1, KDRS1의 순서로 추가될 때의 증가분

^b 남성=1, 여성=0

* $p < .05$, ** $p < .01$, *** $p < .001$, 모든 F 값의 $ps < .0001$

다른 수행 향상이 유의미하였다(장기 회상지수, 장기 재인지수 모두 $p_s < .001$).

단계적 회귀 분석을 통해 EMS 각 검사의 변화량에 영향을 미치는 요인을 탐색한 결과(표 3), 기저 수행 수준은 모든 검사에서 점수 변화를 유의미하게 예측하는 변수로 확인되었는데, 초기 검사에서 수행 수준이 낮을수록 수행 향상이 더 큰 것으로 나타났다. 또한 SST 거꾸로 따라하기 시행점수를 제외한 모든 검사에서 기저 점수가 재검사에 따른 점수변화량을 예측하는데 있어 가장 큰 설명량을 가졌다(R^2 는 .20-.69, $p_s < .001$). 기저 검사의 K-DRS 총점은 EVLT 즉각총점, SRT 재인, DST 바로따라하기, SST의 바로 따라하기와 거꾸로 따라하기, 단축형 K-BNT와 두 장기기억 지수의 변화량을 유의미하게 예측하였는데($p_s < .001$), K-DRS 총점이 클수록 재검사 시 더 많은 향상을 보였다. SST 거꾸로 따라하기의 경우, 기저 수행 수준에 K-DRS 총점이 추가될 때 예측력의 증가분(R^2 변화량=.124)이 기저 수행 수준(R^2 변화량=.097)이 설명하는 총 변량보다도 더 크게 나타났다. 교육은 SRFT 복사과 SRT 재인 수행의 변화량에 대해서만 유의미한 예측변수로 나타났는데, 교육 수준이 높을수록 변화량도 더 큰 것으로 나타났다. 연령은 EVLT와 SRFT, SRT의 즉각회상과 SRFT의 지연회상, 그리고 장기 회상지수에서 유의미한 예측 변수로 나타났으며, 연령이 높을수록 수행이 덜 향상되는 편이었다. 하지만 CDT 총점에서는 연령이 높을수록 오히려 수행이 향상되는 경향을 보였다. 검사-재검사 간격은 DST의 바로 따라하기와 CDT의 변화량만 유의미하게 예측하였으며, 검사 간격이 길어질수록 연습 효과가 작은 경향을 보였다. 한편 성별은 EVLT 장기지연 자유회상에서의 연습 효과를

유의미하게 예측하는 것으로 나타났는데, 기저 수행 수준을 통제해도 남자가 여자에 비해 연습효과가 작은 편이었다.

검사-재검사에 따른 변화 유의성 기준

변화 유의지수는 앞서 기술하였던 RCIP와 SRB 방법에 따라 계산되었으며, 각각의 기준에서 유의미하게 상승한 집단, 유의미하게 감퇴한 집단, 유의미한 변화를 보이지 않은 집단이 점수 변화량 이외의 다른 변수들에서 어떤 차이를 보이는지를 비교분석하였다.

RCIP 모델에 따른 기준

EMS 각 검사와 기억지수에서 검사-재검사 간 신뢰도, RCIP 모델에 따른 90% 예측구간은 표 4에 제시하였다. 검사-재검사 신뢰도는 .29-.71까지로 나타났으며, EVLT, SRFT, SRT에서 지연자유회상($p_s < .001$)에 비해 재인의 신뢰도($p_s < .01$)가 더 낮게 나타났다. SRFT의 복사도 신뢰도가 낮은 편이었다($p < .001$). 개인의 점수 변화량($T_2 - T_1$)이 평균적인 차이 값에서 $1.645 \times SD_{diff}$ 이내(90% 예측 구간 = [RC-, RC+])의 범위에 포함된다면 그 개인은 두 검사 사이의 기간 동안 유의미한 변화를 보이지 않은 것으로 해석할 수 있다.

SRB 모델에 따른 기준

EMS 각 검사와 장기기억 지수의 재검사 수행을 예측하는 회귀 방정식을 도출하기 위해, 각 검사의 기저 점수와 기저 시점의 K-DRS 총점(KDRS1), 연령, 교육 수준, 성별 중 재검사 점수를 유의미하게 예측하는 변수를 단계적 회귀절차에 따라 선택하였다. 그 결과, 각 검사의 기저 점수는 모든 측정치에서 재검사와

표 4. 검사-재검사 신뢰도와 RCIp의 예측구간 (N=111)

검사	검사-재검사 신뢰도	90% 예측구간	RC-	RC+
EVLТ				
즉각총점	0.56	±7.55	-6.05	9.05
단기지연	0.56	±2.57	-1.95	3.19
장기지연	0.63	±2.62	-1.95	3.29
재인	0.61	±2.81	-2.49	3.13
SRFT				
복사	0.39	±2.24	-2.13	2.35
즉각	0.53	±4.33	-3.14	5.52
지연	0.47	±4.31	-3.19	5.43
재인	0.29	±3.44	-3.13	3.75
SRT				
즉각회상	0.66	±6.93	-6.20	7.66
지연회상	0.56	±8.31	-7.80	8.82
재인	0.34	±2.43	-1.95	2.91
DST				
바로	0.54	±3.44	-4.26	2.62
거꾸로	0.67	±2.57	-2.39	2.75
SST				
바로	0.41	±2.91	-2.69	3.13
거꾸로	0.57	±3.01	-2.71	3.31
CDT	0.59	±3.85	-2.94	4.76
K-BNT	0.70	±2.76	-2.57	2.95
장기회상지수	0.71	±49.66	-33.53	65.79
장기재인지수	0.49	±32.23	-24.60	39.86

주. EVLT 단기지연, 장기지연 모두 자유 회상
DST, SST 바로, 거꾸로 모두 시행점수

점수를 유의미하게 예측할 뿐만 아니라 재검사 점수의 분산을 가장 많이 설명하는 변수로 나타났으며, 각 검사의 차이 점수에 대한 회귀 분석에서와 동일한 변수들이 유의미한 예측변수로 나타났다(표 5). 잔차들의 표준 편차를 이용하여 구한 90% 예측구간은 표 5 맨 우측 열에 제시한 바와 같으며, 관찰된 재검사 점수가 회귀방정식에 의해 추정된 개인의 점수에서 위, 아래로 이 구간 이상 벗어나면 유의미하게 향상된 혹은 감퇴한 것으로 해석할 수 있다.

장기회상지수와 장기재인지수에서 RCIp와 SRB의 비교

RCIp의 90% 예측 구간에 의해 상승자와 감퇴자로 분류된 사람들은 장기 회상지수와 장기 재인지수에서 다음과 같았다. 장기 회상지수는 상승자 6.3%, 감퇴자 4.5%로 나타났고, 장기 재인지수는 상승자 6.6%, 감퇴자가 5.7%로 나타났다. 한편, SRB의 90% 예측 구간에 의해 상승자와 감퇴자로 분류된 사람들은 장기 회상지수에서는 상승자 5.4%, 감퇴자 6.3%였으며, 장기 재인지수는 상승자 6.6%, 감퇴자 4.7%로 나타났다. 두 방법 모두 통계적 기대치인 5%에서 크게 벗어나지 않는 비율로 상승자와 감퇴자를 분류했으나, 공통되게 상승자 혹은 감퇴자로 분류된 사람들의 비율은 상당히 낮았다: 장기 회상지수 상승자 0.9%, 감퇴자 3.6%; 장기 재인지수 상승자 2.8%, 감퇴자 0.9%.

각 방법에 의해 상승집단, 감퇴집단, 유의미한 변화가 없는 집단을 나누어 초기 검사 시 연령, 교육 연한, 검사-재검사 간격, 초기검사에서의 점수와 K-DRS 점수에서 집단간 차이가 있는지 일원분산분석(one-way ANOVA)을 실

표 5. SRB에 의한 유의미한 변화 통계량(N=111)

검사	F	R ²	절편	회귀계수					SE	90% PI
				검사1	KDRS1	교육	연령	성별 ^a		
EVLТ										
즉각총점	25.07	.41	12.48	.48***	.15*		-.22**		3.84	±6.32
단기지연	49.01	.31	4.01	.44***					1.20	±1.97
장기지연	41.73	.44	4.15	.46***				-.96**	1.18	±1.94
재 인	64.64	.37	12.79	.55***					1.45	±2.39
SRFT										
복 사	14.01	.21	10.81	.24**		.07**			1.00	±1.65
즉 각	24.49	.31	14.13	.46***			-.10*		2.17	±3.57
지 연	19.31	.26	15.31	.38***			-.10*		2.05	±3.37
재 인	9.31**	.08	12.36	.30**					1.73	±2.85
SRT										
즉각회상	48.25	.47	20.62	.55***			-.19*		3.49	±5.74
지연회상	29.09	.35	5.88	.46***		.24*			4.11	±6.76
재 인	15.53	.23	-1.80	.20*	.06***				1.06	±1.74
DST										
바 로	21.34	.38	-4.23	.48***	.07*			-.04*	1.74	±2.86
거 꾸 로	86.07	.45	1.94	.63***					1.39	±2.29
SST										
바 로	17.63	.26	-5.91	.40***	.08**				1.48	±2.43
거 꾸 로	35.57	.41	-11.60	.48***	.11***				1.63	±2.68
CDT	30.76	.54	-2.35	.29***	.10***		.05*	-.02*	1.12	±1.84
K-BNT	58.38	.52	-3.32	.60***	.06**				1.45	±2.39
장기회상지수	62.88	.54	193.16	.52***			-1.26**		22.78	±37.47
장기재인지수	24.11	.32	78.34	.31***	.81**				14.34	±23.59

주. ^a 남성=1, 여성=0

SE = Standard Error of estimate, PI= prediction interval

* p<.05, ** p<.01, ***p<.001, 모든 F값의 ps<.0001

시하였다. 그 결과, 두 장기 기억지수에 대해 시 연령, 각 기억 지수 값, K-DRS 점수와 교
 서 SRB에 의해 분류된 집단 간에는 초기 검사 육 수준, 검사-재검사 간격에서의 차이가 유의

미하지 않았다(장기 회상지수 $F_s < 2.10$, $p > .13$; 장기 재인지수 $F_s < 1.32$, $p > .27$). 반면, RCIp에 의해 분류된 집단 간에는 각 장기 기억 지수의 초기 수행에서만 집단 차가 유의미하게 나타났다. 즉, 장기 회상지수의 상승집단, 감퇴집단, 무변화집단 간의 초기 장기 회상지수의 차이가 유의미한 것으로 나타났으며($F(2, 108) = 9.10$, $p < .001$), 사후분석결과 재검사 시 유의미한 상승을 보인 사람들은 그렇지 않은 사람들에 비해 초기 검사의 장기 회상지수가 낮은 것으로 나타났다(상승집단의 평균 = 132.67 ± 36.30 vs. 무변화집단 평균 = 188.29 ± 40.85 , 감퇴집단 = 227.75 ± 18.21). 장기 재인지수의 집단간에도 초기 검사 시 재인지수의 차이가 유의미한 것으로 나타났으며($F(2, 103) = 13.63$, $p < .001$), 사후분석결과 재검사 시 유의미한 상승을 보인 사람들은 그렇지 않은 사람들에 비해 초기 검사 시 장기재인지수가 낮은 것으로 나타났다(상승집단의 평균 = 221.67 ± 16.72 vs. 무변화집단 평균 = 255.47 ± 19.20 , 감퇴집단 = 273.61 ± 17.40). 즉, RCIp는 초기검사 시 기억 수행 수준이 낮은 사람들에 편중하여 상승자로 분류하는 경향을 보였다.

논 의

이 연구에서는 정상 노인 대상으로 EMS의 검사-재검사에 따른 수행의 변화 양상과 이에 영향을 미치는 요인들을 탐색하고 RCIp와 SRB에 따른 재검사 점수의 기준을 수립하였다. 또한 각 기준의 90% 예측구간을 적용하여 정상 노인들을 분류한 결과를 비교분석하여 변화 유의성을 판단하는 데 보다 유용한 방법이

무엇인지 살펴보았다. 기억평가는 노인 환자의 신경심리 평가에서 매우 큰 비중을 차지하는 영역으로 본 연구는 노인 기억평가에서 반복측정에 따른 수행 변화를 체계적으로 탐색하고 기준 자료를 제시한 국내 최초의 연구이다.

재검사 시 EMS의 주요 일화기억 측정치에서 모두 유의미한 점수 변화가 관찰되었으며 특히 첫 번째 자극 제시 후 즉각회상의 수행 향상이 두드러졌다. 이에 비해 시공간 구성 능력이나 언어 기능 검사에서는 반복 측정에도 불구하고 수행 향상을 보이지 않았다. EMS의 일화기억 검사들은 즉각회상, 지연회상, 재인 절차를 기본으로 한다. 즉각회상은 정보입력 직후 자유 회상을 하도록 단계로 정보 입력의 효율성에 영향을 크게 받는 단계이며, 지연 회상은 즉각회상 후 일정 과제를 수행하다가 다시 입력된 정보를 기억해내게 하는 단계로 일단 입력된 정보를 보유하고 인출하는 능력의 영향을 받는다. 검사 문항에 대한 사전노출로 그 내용을 기억하게 되는 문항 특정적 연습 효과는 재검사 시 일화 기억검사의 모든 절차에 공통적으로 영향을 미칠 수 있다. 이에 비해 검사 절차에 대한 경험으로 인한 검사 특정적 연습효과는 효율적인 부호화 전략을 구사하고 보다 과제에 집중하도록 함으로써 특히 정보 입력 단계에 큰 영향을 미칠 수 있다. 이 때문에 지연회상이나 재인이 비해 즉각 회상의 수행 향상이 두드러지게 나타난 것으로 보인다. 이러한 양상은 SRT에서 분명하게 나타났는데, 자유 회상 측정치 중에는 즉각회상의 이야기 단위에서만 유의미한 향상을 보였고(평균 차이점수 = $.71 \pm 3.22$, $t(110) = 232$, $p = .02$) 재인점수도 향상되었다(평균 차이점수 = $.48 \pm 1.48$, $t(109) = 341$, $p < .01$). SRT의

재인은 ‘예’, ‘아니오’로 대답하는 EVLT와 SRFT의 재인과 달리 구체적인 정보를 담고 있는 세 가지 선택지 중에 주어진 질문의 정답을 선택하도록 하는 형식을 띠고 있어 자극에 대한 친숙성을 넘어서 세부 정보에 대한 정확한 변별력을 필요로 한다. 기저 검사에서 SRT의 재인 문항을 경험한 후, 재검사에서는 정보 입력단계에서부터 세부 사항을 기억하려는 노력을 기울이게 되어 즉각회상의 이야기 점수가 향상하게 되었을 가능성이 있다. SRT 지연회상의 이야기 단위 점수에는 유의한 상승을 보이지 않는 것으로 보아 즉각회상의 이야기 단위 점수 향상에 이야기 내용에 자체에 대한 장기기억의 효과보다 수행 전략에 대한 학습 효과가 더 큰 영향을 미친 것으로 보인다.

EMS의 모든 검사에서 검사-재검사에 따른 검사 수행의 변화는 기저 수행 수준의 영향을 가장 크게 받았는데, 기저 수행 수준이 낮을수록 재검사 시 수행 향상이 더 크게 나타났다. WMS-III에서도 각 소검사의 기저 점수가 검사-재검사 점수 차이를 가장 유의미하게 예측하는 변수로 확인된 바 있다(Lineweaver & Chelune, 2003). 기저 수행 수준이 낮을수록 재검사 시 수행이 더 많이 향상된다면, 수행이 낮았던 사람에 비해 수행이 높았던 사람의 점수 증가폭이 작아서 재검사에서는 둘 간의 차이가 줄어들 것이다. 기억검사를 재검사할 때 기저 검사에 비해 재검사에서 수행의 차이가 감소하는 평균으로의 회귀 현상이 광범위하게 나타나는 것으로 시사된다.

기억 검사는 일반 인지기능을 측정하는 검사들에 비해 기저점수가 점수변화에 미치는 영향이 더욱 큰 것으로 보인다. 다음 공식 [다], [라]에서 알 수 있듯이, 추정된 검사-재검사 차이값(T'_2-T_1)은 다른 변수들을 통제했을

때, a-1의 절대값이 클수록, 다시 말해 회귀계수 a가 1에서 멀어 질수록, 기저점수(T_1)의 영향을 크게 받는다.

$$T'_2 = a * T_1 + b_1 * X_1 + \dots + b_n * X_n + c \quad [다]$$

$$T'_2 - T_1 = (a-1) * T_1 + b_1 * X_1 + \dots + b_n * X_n + c \quad [라]$$

(단, a=기저 수행에 대한 회귀 계수, X_1, \dots, X_n =수행변화를 조절변수들, b_1, \dots, b_n = 조절변수들의 회귀 계수들)

그런데 McSweeny 등(1993)의 연구 결과, 재검사를 예측할 때 기저 점수의 회귀 계수가 WAIS로 측정된 지능 지수들에서는 .94-1.09이고, WMS-R의 기억 지수들은 기저 점수에 대한 회귀계수가 .78-.88로 나타났다. 본 연구에서는 EMS 각 검사들의 재검사 점수를 예측할 때 기저 점수의 회귀계수가 .20-.63으로 나타났다. K-DRS의 검사-재검사 기준 연구(김호영, 석정서, 최진영, 2010)에서는 일반 인지기능을 측정하는 K-DRS 총점에 대한 기저 수행의 회귀계수가 .82로 나타났다. 즉, 일반 인지기능을 측정하는 검사들에 비해 기억검사들에서 재검사 점수를 추정하는 기저 점수의 회귀계수가 1에서 멀게 나타났다. 이는 다른 인구통계학적 변수를 통제할 때, 일반 인지기능을 측정하는 검사들에 비해 기억검사들에서 점수 변화량이 기저 수행 수준의 영향을 더 크게 받음을 시사한다. 따라서 기억 검사의 점수 변화량을 해석할 때에는 반드시 기저 수행 수준을 고려해야 한다.

신경심리평가에서 기저수행이 낮은 정상 노인들이 재검사 시 수행 향상이 크게 나타난다는 특징은 진단적으로도 유용한 정보가 될 수 있다. 국내 노인들은 정규 교육을 받지 않은

무학 노인들이 60세 이상 노인인구의 26%나 차지하고 있는데(통계청, 2007), 이들은 인지장애가 없음에도 불구하고 신경심리평가에서 낮은 수행을 보이는 경향이 있다. 따라서 저교육자들이 신경심리평가에서 낮은 인지 수행을 보일 때, 원래 인지기능은 낮은 것인지 치매가 진행되고 있는 중인지를 구분하는 것이 중요하다. 그러나 일회적인 평가만으로는 이러한 판단을 내리기 어려운 경우가 많다. 기저수행이 낮은 사람일수록 연습 효과가 더 크게 나타나는 경향을 고려하여 재검사 시 기저수행에 따라 기대되는 연습효과가 나타나는지를 살펴보면 보다 정확한 해석을 내리는데 도움이 될 것이다. 최근 한 연구에서 기억검사의 재평가에서 정상노인들은 수행이 향상되었지만 초기 알츠하이머성 환자들은 연습효과를 보이지 않았다는 결과를 보고했듯이(Zehnder, Bläsi, Berres, Spiegel, & Monsch, 2007), 정상적으로 기대되는 연습효과가 나타나지 않으면 치매 진행을 의심해 볼 수 있다. 특히 기저 검사에서 낮은 수행을 보인 사람들은 많은 수행 향상이 기대되므로 재검사 수행이 기저수행보다 저하되지 않았다 하더라도 기대되는 수준의 연습 효과가 나타나지 않는다면 인지장애가 진행되고 있을 가능성을 고려해야 한다.

K-DRS 총점으로 측정된, 기저 시점의 일반 인지기능 수준은 EMS의 여러 검사 중 단기 기억 및 작업기억 검사, 언어 재인기억 검사의 검사-재검사 차이점수에 유의미한 영향을 미쳤는데, 전반적인 인지기능 수준이 높을수록 연습효과도 더 크게 나타났다. Lineweaver와 Chelune(2003)도 본 연구와 일치하는 결과를 보고한 바 있다. 이들의 연구에서는 WAIS-III의 전체 지능 지수가 높을수록 WMS-III의 주요 기억지수에서 연습효과가 더 크게 나타났다. 특

정 검사의 기저 수행 수준을 통제하여 평균으로의 회귀 현상의 영향을 제거했을 때, 기억 검사에서 기저 인지기능이 우수한 사람들은 인지기능이 낮은 사람들에 비해 연습 효과가 더 크며, 특히 작업기억과 재인기억에서 그러한 경향이 두드러지는 것으로 시사된다.

연령, 교육, 성별과 같은 인구통계학적 변인들도 EMS의 일부 검사에서 재검사에 따른 수행 향상 정도에 유의미한 영향을 미쳤다. 기저 검사 시점의 연령은 언어적 기억검사의 즉각회상, 비언어적 기억검사의 즉각 및 지연회상, 그리고 장기 회상지수에서의 향상을 유의미하게 예언하였는데, 연령이 높을수록 점수가 덜 향상되는 경향을 보였다. WMS-III에서도 본 연구에서와 마찬가지로 연령이 높을수록 재검사에 따른 연습효과가 작게 나타나(Lineweaver & Chelune, 2003), 연령이 낮을수록 일화적 기억검사의 사전 경험으로부터 더 많은 이득을 얻는 것으로 보인다. 기억 이외의 다른 신경심리 검사의 재검사에 대한 선행 연구들에서도 이 결과와 일관되게 연령이 높을수록 연습효과가 작은 경향을 보고하고 있다(Temkin et al., 1999; 김호영 등, 2010). 이와는 상반되게 일화적 기억 기능의 영향이 크지 않은 CDT에서는 오히려 연령이 높을수록 연습효과가 크게 나타났다. 하지만 기저 시점의 연령과 CDT 점수 변화간의 단순 상관이 유의미하지 않았고($r=.07, p=.45$), 연령만을 예측변수로 하여 CDT 점수변화를 예측하는 단순 회귀모형도 유의미하지 않았고($F(1, 107)=.57, p=.45$). 따라서 본 연구에서 CDT 점수변화에 대한 연령의 효과는 주어진 표본 자료로 단계적 회귀 절차에 따라 최적화된 회귀식을 도출해내는 과정에서 나온 결과일 뿐 일반적인 현상으로 해석하기 어렵겠다. 한편 시공간 구성

능력과 이야기 회상 검사의 지연회상에서는 교육 수준이 높을수록 수행이 더 많이 향상되는 것으로 나타났는데, Temkin 등(1999)도 교육 수준이 높을수록 신경심리검사의 재평가에서 수행이 더 많이 향상되는 경향을 보고한 바 있다. 또한 성별은 EVLT의 장기 지연회상의 연습효과에서만 유의미한 영향을 미치는 것으로 나타났는데, 남성이 여성보다 연습효과가 더 작은 경향을 보였다. 본 연구 참여자가 여성에 지나치게 편중되어 있고, 남성 참여자의 수가 제한적이어서 이러한 결과를 일반화할 수는 없다. 일반적으로 언어적 검사에서 여성의 수행이 더 좋은 것으로 알려져 있지만, 연습효과에 대한 성별 효과를 직접적으로 보고한 선행 연구는 찾아보기 어렵다. Ratcliff, Birzescu와 Ganguli(2003)는 2년 간격으로 10년에 걸쳐 노인들의 인지기능을 추적한 종단 연구에서 교육 수준과 성별에 따른 다른 추이를 보고한 바 있다. 이들은 인구통계학적 변수와 신경심리검사 수행 변화 간의 관계를 살펴보았는데, 교육을 더 많이 받은 사람수록 언어 유창성 검사를 비롯한 몇몇 신경심리검사에서 수행이 덜 저하되고, 남성보다 여성이 언어적 기억검사를 포함한 일부 검사들에서 수행이 덜 감퇴하는 경향이 있음을 보고했다. 본 연구의 결과를 고려할 때, Ratcliff 등(2003)의 연구 결과는 교육 수준이 높을수록, 남성보다 여성이 해당 검사들에서 연습 효과가 더 커서 인지노화에 따른 인지기능 감퇴를 상쇄시켰기 때문일 수 있다.

검사-재검사 간격은 주요 일화 기억검사의 연습효과에는 영향을 미치지 않았지만, DST 바로 따라하기와 CDT에서는 유의미한 영향을 미쳤다. 두 검사 모두에서 검사 간격이 길수록 연습 효과가 적게 나타났는데, 검사-재검사

간격이 길어질수록 사전 검사 경험으로 인한 연습효과보다는 노화에 따른 인지기능 감퇴가 더 큰 영향을 미치기 때문으로 해석할 수 있다. 하지만 이러한 양상이 주요 일화 기억검사와 K-BNT, SRFT의 복사 시행에서는 관찰되지 않는 것으로 보아, 적어도 본 연구에서와 같이 검사-재검사 간격이 50개월 이내인 경우 신경심리검사의 연습효과에 미치는 영향이 제한적인 것으로 보인다. WAIS와 WMS에서도 검사-재검사 간격은 반복측정에 따른 수행변화를 유의미하게 설명하지 못했다(Basso et al., 2002; Lineweaver & Chelune et al., 2003; McSweeney et al., 1993).

마지막으로 EMS의 각 검사에서 검사-재검사에 따른 변화에 대한 기준을 RCIP와 SRB 접근법에 의해 계산하여 제시하였다. 장기 회상 지수와 장기 재인지수에서 위의 두 방법을 통해 수립한 각각의 기준으로 상승자와 감퇴자를 분류하고 이를 비교해본 결과, 일치율이 매우 낮게 나타났다. SRB와 RCIP로 각각 분류된 상승집단, 감퇴집단, 무변화 집단 간에 기저 시점의 일반 인지기능, 연령, 교육수준, 검사-재검사 간격의 차이는 유의미하지 않았다. 하지만 장기 기억지수의 기저수행에서는 차이를 보였다. SRB로 분류된 집단들은 장기 기억지수들의 기저 수행에 유의미한 차이를 보이지 않았지만, RCIP로 분류된 집단들의 경우 상승집단의 기저 수행이 다른 두 집단의 기저 수행에 비해 유의미하게 낮았다. 즉, RCIP는 기저 수행 수준이 낮은 사람에 편중되게 상승자를 분류하는 경향을 뚜렷하게 보였다. 이는 SRB가 기저 수행 수준에 따라 차별적으로 나타나는 연습효과를 고려하는 데 비해, RCIP에서는 기저 수행을 고려하지 않고 고정된 연습효과를 가정하기 때문으로 시사된다. 본 연구

의 참여자는 모두 연구 기간 내에 신경정신과적으로 유의미한 변화가 없었던 정상 노인들이다. 그러므로 연구 기간 동안 기저 수행 수준이 낮았던 사람들 중의 더 많은 사람들이 실제로 기억 기능이 유의미하게 향상되었다고 주장할 만한 근거가 없음에도 불구하고 이들에 편향되게 상승자를 분류하는 것은 타당하다고 볼 수 없다. K-DRS의 검사-재검사 규준 연구에서도 RCIP는 기저 검사에서 낮은 수행을 보였던 사람들에게 편중되게 상승자를 분류하는 경향이 있었다(김호영 등, 2010). 따라서 RCIP보다는 SRB가 변화유의성을 판단하는 데 보다 적합한 방법으로 사료된다. Linweaver와 Chelunc(2003)도 변화유의성을 판단하는 여러 방법들 중 기저수행수준이나 인구통계학적 변수들이 검사수행 변화에 미치는 영향을 고려할 수 있는 SRB를 권장하였다.

본 연구의 제한점과 후속 연구에서 추가적으로 검토되어야 할 점들은 다음과 같다. 우선 이 연구의 참여자들은 서울 및 수도권 지역 거주 노인들로 인구통계학 분포가 국내 전체 노인들의 분포와 차이를 보인다. 2005년 현재 60세 이상 국내 노인 인구에서 남성의 비율은 약 42%를 차지하며, 무학자는 전체의 26%, 교육연한이 12 이상인 사람들은 9%로 추정된다(통계청, 2007). 본 연구의 표본은 국내 전체 노인집단에 비해 남성의 수와 무학자의 비율이 적은 데 비해, 고졸 이상의 고교육자의 비율은 지나치게 많고, 전체 표본의 크기도 충분하지 않았다. 추후 모집단을 대표할 수 있는 보다 큰 표본에서 본 연구에서 보고한 인구통계학적 변수들의 영향을 재검토해야 한다. 둘째, 본 연구에서는 검사-재검사 간격이 반복측정에 따른 수행변화에 미치는 영향이 제한적이었는데, 특히 주요 일화 기억검사

들의 수행변화에는 영향을 미치지 않았다. 본 연구에서는 검사-재검사 간격을 특정 간격으로 고정하지 않고 종단연구 자료 중 첫 번째 검사와 두 번째 검사 사이에서 나타나는 변화를 탐색하여 검사-재검사 간격을 통제한 다른 연구들에 비해 검사-재검사 간격의 범위가 매우 크다. 그렇지만 7-15개월 간격에 전체 표본의 약 70%가 포함되어 있고, 34개월 이상의 간격으로 검사를 실시한 사람들은 전체 표본의 15%에도 못 미치는 등 자료가 편중되어 있어 검사-재검사 간격의 영향을 평가하는 데 제한점이 있었다. 기억은 인지 노화에 취약한 영역이어서 노인기 기억 기능의 종단적 평가 결과를 정확히 해석하기 위해서는 검사들 사이의 시간 경과가 수행 변화에 미치는 영향을 정확히 파악하고 있어야 한다. 추후 연구에서 검사-재검사 간격을 폭넓고 다양하게 하여 검사 간격이 기억 검사의 연습 효과에 미치는 영향을 포괄적이고 체계적으로 탐색할 필요가 있다. 셋째, 이 연구에서는 변화 유의성을 판단하는 방법으로 SRB와 RCIP를 이용하여 수행 변화가 정상 범위에서 크게 벗어나는 자들을 편향되지 않게 탐지할 수 있는지를 비교분석하였다. 하지만 정상 집단에서 극단적인 상승 혹은 감퇴를 보인 사람들을 편향되지 않게 변별해낸다는 것만으로는 임상적인 유용성을 보장할 수 없다. SRB와 RCIP와 같이 변화의 유의성을 탐지하는 방법들이 실제 임상 장면에서 유용함을 검증하기 위해서 장기적인 종단연구를 통해 각 방법에 의해 계산된 변화 유의지수가 추후 기억장애의 임상적 발현을 예측하는 데 얼마나 도움이 되는지 확인해야 한다. 혹은 SRB나 RCIP가 뇌졸중이나 뇌 손상과 같이 임상적으로 중요한 사건에 따른 기능 변화를 잘 탐지할 수 있는지를 살펴보아야 한다.

마지막으로 이 연구에서는 기저 검사와 이 후 첫 번째 재검사만을 비교하여 자료를 분석하였다. 임상 장면에서는 경과 추적을 위해 두 번 이상의 재평가가 이루어지는 경우도 흔히 있어, 두 번 이상의 반복적인 측정 과정에서 나타나는 수행 변화의 특징에 대한 탐색도 필요하다.

이상의 연구 결과에서 다음과 같은 결론을 내릴 수 있다. 첫째, 기억 검사의 검사-재검사에서 검사 질차나 수행 전략과 관련된 검사 특정적 연습 효과가 문항 내용에 자체에 대한 기억보다 더 큰 영향을 미치는 것으로 시사된다. 둘째, 기억 검사는 지능 수준과 같은 일반 인지능력 수준을 평가하는 검사들에 비해 재검사 시 평균으로의 회귀가 강력하게 나타나므로 점수변화를 해석할 때 기저 수행 수준으로 반드시 고려해야 한다. 셋째, 기억 기능의 검사-재검사 시 점수 변화는 해당검사에서 개인의 기저수행 수준, 일반 인지능력, 인구통계학적 변수들에 따라 변산이 크므로 이를 고려해서 해석해야 한다. 특히 평균적으로는 재검사에 따른 점수 변화가 유의미하지 않았더라도 개인의 특징에 따라 점수 변화의 편차가 클 수 있음을 염두에 두어야 한다. SRB와 같이 재검사에 따른 점수 변화의 개인차를 고려하는 분석방법은 변화의 유의성을 보다 타당하게 해석하는 데 도움이 될 것이다.

참고문헌

김호영, 석정서, 최진영 (2010). 한국판 치매 평가 검사(K-DRS)를 이용한 노인인지기능 평가에서 변화유의성 탐지: 50세 이상 장 노년 검사-재검사 기준 연구. 한국심리학회

지: 임상, 29권, 495-513
최진영 (1998). 한국판 치매 평가 검사: Korean-Dementia Rating Scale. 서울: 학지사.
최진영 (2007). 노인 기억장애 검사. 서울: 학지사.
통계청 (2007). 2005 인구주택총조사보고서: 표본조사결과, 제10권. 고령자. 대전: 통계청.
American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders. 4th ed.* Washington DC: American Psychiatric Association.
Anastasi, A. (1988). *Psychological testing (6th ed.)*. New York: Macmillan Publishing Company.
Basso, M. R., Carona, F. D., Lowery, N., & Axelrod, B. N. (2002). Practice effects on the WAIS-III across 3- and 6-month intervals. *The Clinical Neuropsychologist*, 16, 57-63.
Beglinger, L. J., Gaydos, B., Tangphao-Daniels, O., Duff, K., Kareken, D. A., Crawford, J., Fastenau, P. S., & Siemers, E. R. (2005). Practice effects and the use of alternate forms in serial neuropsychological testing. *Archives of Clinical Neuropsychology*, 20, 517-529.
Benedict, R. H. B., & Zgaljardic, D. J. (1998). Practice effects during repeated administrations of memory tests with and without alternative forms. *Journal of Clinical and Experimental Neuropsychology*, 20, 339-352.
Bush, R. M., Chelune, G. J., & Suchy, Y. (2006). Using norms in neuropsychological assessment of the elderly. In D. K. Attix & K. A. Welsh-Bohmer (Eds.), *Geriatric neuropsychology: Assessment and intervention* pp. 133-157. New York: Guilford Press
Chelune, G. J. (2002). Assessing reliable

- neuropsychological change. In R. Franklin (Ed.), *Prediction in forensic and Neuropsychology: New approaches to psychometrically sound statistical practices*, pp. 115-138. Mahwah, NJ: Lawrence Erlbaum Associates Inc.
- Chelune, G. J., Naugle, R. I., Luders, H., Sedlak, J., & Awad, I. A. (1993). Individual change after epilepsy surgery: Practice effects and base-rate information. *Neuropsychology*, 7, 41-52.
- Chey, J., Na, D. R., Park, S., Park, E., & Lee, S. (1999). Effects of education in dementia assessment: Evidence from standardizing the Korean-Dementia Rating Scale. *The Clinical Neuropsychologist*, 13, 3, 293-302.
- Frerichs, R. J., Tuokko, H. A. (2005). A comparison of methods for measuring cognitive change in older adults. *Archives of Clinical Neuropsychology*, 20, 321-333.
- Lineweaver, T. T. & Chelune, G. J. (2003). Use of the WAIS-III and WMS-III in the context of serial assessments: Interpreting reliable and meaningful change. In D. S. Tulsky, D. H. Saklofske, G. J. Chelune, R. K. Heaton, R. J. Ivnik, R. A. Bornstein, A. Prifitera, & M. F. Ledbetter. (Eds.), *Clinical Interpretation of the WAIS-III and WMS-III*, pp. 301-335. New York: Academic Press.
- Mattis, S. (1988). *Dementia Rating Scale (DRS): Professional Manual*. Odessa, FL: Psychological Assessment Resources.
- McSweeney, A. J., Naugle, R. I., Chelune, G. J., & Luders, H. (1993). "T scores for change": An illustration of a regression approach to depicting change in clinical neuropsychology. *The Clinical Neuropsychologist*, 7, 300-312.
- Ratcliff, G., Dodge, H., Birzescu, M., & Ganguli, M. (2003). Tracking cognitive functioning over ten years of followup of a rural elderly community population. *Applied Neuropsychology*, 10, 76-88
- Temkin, N. R., Heaton, R. K., Grant, I., & Dikmen, S. S. (1999). Detecting significant change in neuropsychological test performance: A comparison of four models. *Journal of the International Neuropsychological Society*, 5, 357-369.
- Winblad, B., Palmer, K., Kivipelto, M., Jelic, V., Fratiglioni, L., Wahlund, L.-O., Nordberg, A., Bäckman, L., Albert, M., Almkvist, O., Arai, H., Basun, H., Blennow, K., de Leon, M., DeCarli, C., Erkinjuntti, T., Giacobini, E., Graff, C., Hardy, J., Jack, C., Jorm, A., Ritchie, K., van Duijn, C., Visser, P., & Petersen, R. C. (2004). Mild cognitive impairment: beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *Journal of Internal Medicine*, 256, 240-246.
- Zehnder, A. E., Bläsi, S., Berres, M., Spiegel, R. & Monsch, A. U. (2007). Lack of Practice Effects on Neuropsychological Tests as Early Cognitive Markers of Alzheimer Disease? *American Journal of Alzheimer's Disease and Other Dementia*, 22, 416-426.

원고접수일자 : 2010. 4. 10.

게재확정일자 : 2010. 7. 12.

Meaningful change in memory scores in repeated administration of the Elderly Memory disorder Scale

Hoyoung Kim

Jeanyung Chey

Department of Psychology, Seoul National University

Memory is a very important cognitive domain in neuropsychological evaluation of aged adults, and longitudinal assessments are often recommended for objective evidence of memory decline. Since memory decline and practice effects are both often observed in test-retest memory evaluation, normative standards to determine meaningful or reliable changes in memory performance are necessary in elderly adults. We explored the factors affecting the test-retest performance of memory tests and established normative rates of change on the Elderly Memory-disorder Scale from baseline to follow-up testing among 111 neurologically normal adults aged 60 and older. The results of stepwise multiple regression analyses revealed that the most powerful predictor of performance change was the initial performance level. Education, age and general intellectual functioning were also significant predictors for some test scores. Generally, a person whose baseline performance was lower improved more on the retest. After controlling for one's initial performance, a person who was better-educated, younger, or more intelligent demonstrated greater improvement in one's performance. Normative retest data are presented for two analytic techniques: the Reliable Change Index adjusted for practice (RCIp) and the Standardized Regression-Based technique (SRB). For long-term memory indices of EMS, we classified improver and decliner by using the 90% prediction intervals of RCIp and SRB. We compared the two techniques in the improved group, the declined group, and the unchanged group, and found that RCIp tended to classify low performers as improvers. These results suggest that the SRB, which factors in performance differences at baseline and demographic variables, is a more useful and unbiased method for detecting changes compared to the RCIp.

Key words : memory tests, practice effect, SRB, RCIp, prediction interval