

한국판 치매 평가 검사의 타당도와 신뢰도 연구*

최진영

성신여자대학교

나덕렬

성균관 대학교 의과대학 삼성의료원 신경과

박선희

박은희

성신여자대학교

본 연구는 Mattis(1988)의 Dementia Rating Scale(DRS)의 한국판인 치매 평가 검사(Korean-Dementia Rating Scale: K-DRS, 최진영, 1998)의 타당도와 신뢰도를 검토하기 위하여 행해졌다. 143명의 정상 노인이 서울-경인 지역에서 표집되었으며 20명의 알츠하이머성 치매(dementia of the Alzheimer's Type: DAT) 환자가 신경과 외래에서 표집되었다. 타당도의 여러 차원을 살펴보기 위하여 세 국면의 타당도를 검토하였다. 첫 번째로 고려된 것은 영어권에서 개발된 검사인 DRS를 국어로 번역하는 과정에서 원 과제가 검사하고자 한 인지 과정을 K-DRS가 다루도록 내용 타당도를 유지시키는 일이었다. 두 번째 단계에서는 경험적으로 타당도를 검증하는 작업을 시도하여 구성 타당도와 기준 타당도를 살펴보았다. 구성 타당도는 K-DRS의 총점과 MMSEK의 총점간의 상관을 구하여 살펴봤는데 이들간의 상관계수는 .82로 유의미하게 높았다. 진단을 기준으로 한 기준 타당도는 전체 집단을 대상으로 할 경우 상관 계수가 .66이었고 연령과 학력으로 구분하여 산출한 경우 고령 저학력 노인 집단을 제외하고 모두 상승하여 높은 상관을 보여주었다. 마지막으로 검사내의 소검사들이 치매의 변별에 기여하는 정도를 알아보았다. 기억, 관리 기능 소검사가 유의미하게 기여하였으며 다른 소검사들은 유의미한 수준에 미치지 못했다. 요약하자면, 여러 면에서 한국판 치매 평가 검사는 치매를 진단하는 검사 도구로서의 타당성이 검증되었다. 20명의 정상 노인들과 11명의 DAT 환자들을 대상으로 2주 간격으로 실시된 검사-재검사 신뢰도는 .96이었고, 3명의 채점자간의 신뢰도는 .99로 매우 높게 나타났다.

노인성 치매 인구의 증가(통계청, 1997)와 함께 치매에 대한 사회적 관심이 높아지고 국가적인 차원의

대책도 추진되고 있다(보건복지부, 1996). 그러나, 치매 환자를 간편하고 정확하게 진단할 수 있는 도구

* 이 논문은 1996년도 한국학술진흥재단의 공모과제 연구비에 의하여 연구되었음.

가 현재 국내에 부재하여 치매에 관한 정책이나 연구의 효율성이 상당히 떨어지고 있다. 치매 환자의 의학 연구, 임상 및 기초 연구가 활발히 이루어지기 위해서는 측정이나 지수화가 가능한 도구가 필요하다. 치매를 진단하고 경과를 측정할 수 있는 도구의 개발은 국내 의료계는 물론 국가적으로도 필요한 과제이다. 이러한 필요성에 부응하여 한국판 치매 평가 검사(Korean-Dementia Rating Scale: K-DRS; 최진영, 1998)가 국내 노인 인구를 위하여 개발되었다.

Dementia Rating Scale(DRS; Mattis, 1988)은 치매, 특히 노인성 퇴행성 치매 혹은 알츠하이머성 치매(Dementia of the Alzheimer's Type: DAT)를 감별하기 위하여 미국에서 제작된 도구이다. DAT는 대뇌 피질의 신경 세포들이 퇴행하므로써 일어나는 인지행동상의 기능 저하를 주 특징으로 하는 질환으로 노화에 따라 발병률이 증가한다(Cummings & Benson, 1992). DAT의 진단 도구는 이러한 뇌와 그 기능상의 변화를 측정할 수 있어야 하는데 직접적인 신경 병리적인 퇴행의 관찰을 통해서 환자를 진단하는 경우는 매우 드물다. DAT 환자의 대뇌 피질 변화의 지표인 신경섬유 뭉치들(neurofibrillary tangles)과 신경반(neuritic plaque) 및 기타 뇌의 변화는 사후 부검(autopsy)이나 생존시 생검(biopsy)을 통한 뇌 조직 검사가 이루어져야 확인된다. 물론 뇌의 단층 촬영(Computerized Tomography: CT)이나 자기 공명 촬영(Magnetic Resonance Imaging: MRI) 등의 신경조영 촬영 기술들(neuroimaging technologies)은 뇌의 구조적인 변화를 측정하는데 유용하게 사용되고 있다. 그러나 DAT 환자의 뇌 변화는 초기에는 세포 수준에서 이루어지며, 중기로 들어서야만 피질의 구(sulci) 혹은 열(fissures)이 깊어지는 것을 특징으로 하는 뇌 구조의 변화가 관찰된다. 그러나 중기에 비로소 관찰되는 이러한 뇌 구조의 변화도 매우 비특이적이므로 DAT 진단에 결정적인 도움을 주지는 못한다.

이러한 신경조영 촬영 기법들이 뇌의 구조 변화를 감지하기 위하여 사용되는데 반해 최근에 사용되는 양전자 방출 단층 촬영(Positron Emission Tomography: PET)과 단일 광전자 방출 촬영(Single Photon

Emission Computed Tomography: SPECT)은 뇌의 기능상의 변화를 측정하는 목적으로 사용되고 있다. 뇌의 기능상의 변화가 구조 변화에 선행하기 때문에 이러한 첨단 뇌영상 기술은 DAT 뇌의 초기 변화를 감지할 수도 있다. 그러나, 이 검사에서 관찰되는 초기 뇌 변화는 정상 노인 및 다른 신경 질환 환자와의 구분이 어려워 DAT의 질병특유(pathognomonic)의 특징으로 간주되지는 않는다. 또한 PET나 SPECT 상에서 관찰된 뇌의 국소적인 기능 저하가 그와 연관된 인지 기능의 변화로 명확하고 직접적으로 이어지지 않기 때문에 DAT 진단을 위하여 사용하는 데에는 한계가 있다(Cummings & Benson, 1992).

이와 같이 DAT 진단에는 한 가지의 결정적인 검사나 생물학적 지표(biological marker)가 없는 실정이다. 결국 DAT 진단을 위해서는 신경심리 검사, 신경학적 검사, 혈액 검사, 뇌 촬영 등이 동원되며 이 중에서 진단의 첫 단계인 치매의 유무 및 정도를 평가하는 데에는 자세한 신경심리학적인 평가가 필수적이다. DAT의 진단을 위해서는 보통 인지 행동적인 기능 감퇴가 주된 진단 기준이 되기 때문이다. 치매 진단시 가장 많이 사용되는 기준은 National Institute of Neurology and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA; McKhann et al., 1984)과 Diagnostic and Statistical Manual of the Mental Disorders IV (DSM-IV; American Psychiatric Association, 1994)이 내용은 두 개의 기준인데 서로간에 다소의 차이는 있지만 두 기준 모두 기억력을 포함한 두 가지 이상의 인지 기능 감퇴가 확증되어야 한다고 규정하고 있다(진단 기준에 관한 자세한 내용은 Cummings & Benson, 1992, 및 최진영과 이소애, 1997, 을 참조하기 바람). DRS는 DAT 환자들의 진단에 필수적인 대표적인 신경심리 검사 중 하나이며(Lezak, 1995) 진단 검사 도구로서의 타당성이 입증되었다(Chase et al., 1984; Coblenz et al., 1973; Nadler, Richardson, Malloy, Marran, & Hostetler Brinson, 1993; Teri, Borson, Kiyak, & Yamagishi, 1989; Vitaliano et al., 1984).

DRS의 타당도는 여러 차원에서 연구되어 왔는데 제일 먼저 고찰된 것은 구성 타당도였다. DRS의 목적이 치매 환자, 특히 DAT 환자의 인지 기능 감퇴를 측정하기 위해 개발되었으므로 인지 기능을 측정하는 다른 심리측정 도구들과의 상관 정도가 검토되었다. DRS 총점이 다른 인지 검사들과 높은 상관을 보이므로써 DRS는 인지 기능을 측정하는 도구로서의 타당성을 지지 받았다(Chase et al., 1984; Coblenz et al., 1973). Coblenz 등(1973)은 Wechsler Adult Intelligence Scale(WAIS; Wechsler, 1955)과 Wechsler Memory Scale(WMS; Wechsler, 1945)의 언어 기억 검사(짝-연합 학습 검사; paired associate learning subtest)와 DRS 총점간의 상관 관계를 검토하였다. 그 결과 WAIS와 DRS의 총점간의 상관은 뇌 손상 환자를 포함한 집단에서 .75였고, 이 중 WMS의 언어 기억 점수의 평균에서 표준편차를 감한 점수보다 더 낮은 점수를 얻은 뇌 손상 환자 집단만을 대상으로 할 경우 상관이 .86으로 증가하였다.

대뇌 기능을 물리적으로 측정한 지표와 DRS 간의 상관 관계를 통해서도 검사의 뇌 기능을 측정하는 도구로서의 타당도가 지지되었다. Chase 등(1984)은 DRS 총점과 대뇌 피질의 포도당 신진대사 지수간의 상관을 고찰하여 .59의 유의미한 상관을 얻었다. 그는 또한 WAIS와 WMS 등의 심리측정 검사와도 비교적 높은 상관을 얻어 Coblenz 등(1973)의 결과를 반복 검증하였다. 전두엽의 뇌혈류량과 DRS 총점간의 상관을 구한 연구(Gardner, Oliver, Fisher, & Empting, 1981)에서도 .85의 높은 상관이 관찰되었다.

Kaszniak(1986)은 DRS 총점이 정상 집단과 DAT 집단을 유의미하게 변별함을 보고하여 진단을 기준으로 한 기준 타당도를 확립하였다. 또 Vitaliano 등(1984)은 경도에서 경중도 정도의(mild to moderately severe) DAT 노인들과 정상 노인 집단간의 DRS 수행을 비교하였다. 경도, 경중도의 환자 집단과 정상 집단 모두 간에 DRS 전체 수행에서 유의미한 차이를 관찰하였다. 경도와 경중도 환자 집단간에는 5개의 소검사들 모두에서 유의미한 수행 차이를 보고하였고 정상인과 경도의 DAT 환자간에는 기억, 관리

기능, 구성 소검사에서 유의미한 차이가 관찰되었다. 이 외에도 이 연구는 DRS 총점과 환자의 일상 생활상의 기능 수준간의 상관을 살펴보았다. 이는 최근 더 강조되는 생태학적 타당성을 검토하는 과정으로 검사실에서의 수행 수치와 일상 생활에서 환자가 얼마나 정상적으로 살 수 있는지 여부와 상관을 살펴보는 작업이다. Vitaliano와 그의 동료들은 기본적인 기능인 식사와 개인 위생 등을 측정하는 일상 생활 척도(Activities of Daily Living: ADL)와 좀 더 복잡한 인지 과정을 요구하는 취미 생활과 읽기 등의 기능을 측정하는 독립성 척도들과 DRS 총점간에 각기 .76과 .56의 상관 계수를 얻어 생태학적인 타당성을 입증하기도 하였다.

치매 환자의 일상 생활 기능을 예측하는 정도를 살펴본 또 다른 연구(Nadler et al., 1993)에 의하면 DRS 총점은 처방된 약의 올바른 복용, 본인의 안전을 지킬 수 있는지 여부, 위생 관리와 금전 관리 등을 수행하는 능력을 유의미하게 예측하였다(r 의 범위는 .52~.64). 이와 비슷한 연구를 한 다른 연구에서도(Teri et al., 1989) 도구적인 일상 생활 활동(ADL)과 자신을 돌보는 일상 생활 활동으로 측정된 일상 생활에서의 기능 수준과 DRS 총점 및 소검사들 간에 정적인 상관을 보고하고 있다.

DRS의 신뢰도에 대한 연구들은 반분 신뢰도 .90, 1주 후 재검사 신뢰도 .97 등으로 높은 신뢰도가 있음을 증명하였다(Coblenz et al., 1973; Gardner et al., 1981).

국내 치매 환자들의 진단을 위하여 개발된 K-DRS의 사용에 앞서 그 타당도와 신뢰도를 살펴보는 작업은 중요하고 필수적인 절차이다. 본 연구에서는 DAT의 진단 도구로서의 K-DRS의 타당성을 검토하기 위하여 세 가지 접근을 취하였다. 첫 번째 접근은 내용 타당도를 고찰하여 선형적으로 검사가 측정하고자 하는 인지 기능들을 타당하게 측정하도록 검사를 보완하는 작업이다. 두 번째 접근은 경험적으로 구성 타당도와 기준 타당도를 검증하는 작업이었다. 현재 국내에서 치매 진단에 사용되어지는 검사 중 개념적으로 유사하게 인지 기능을 측정하는 Mini Mental

State Exam-Korean (MMSEK; 권용철과 박종한, 1989)과 K-DRS간의 상관 관계를 살펴보므로써 구성 타당도를 측정하고자 하였다. 기준 타당도는 진단 기준을 준거로 하여 K-DRS 총점이 DAT 진단을 변별할 수 있는 정도를 측정하므로써 확인하고자 하였다. 또한 신경심리 검사 수행에 중요한 영향을 미치는 인구통계학적인 요인들을 고려하여 세부적인 기준을 마련할 때 진단 변별력이 증가하는지를 알아보고자 하였다. K-DRS의 성별, 연령, 학력 요인을 살펴본 연구에 의하면 학력과 연령이 수행에 유의미한 영향을 미쳤다(최진영과 이소애, 1997). 이를 근거로 정상 집단을 연령과 학력을 기준으로 4개의 집단으로 나누었을 때 기준 타당도가 증가될 것으로 예측되었다. 세 번째 접근은 K-DRS의 소검사들이 치매 환자를 감별하는데 각자 어느 정도 기여하는지를 검토하여 K-DRS 소검사의 타당도를 고찰하는 작업이었다.

신뢰도 검증을 위해서는 20명의 정상 피험자와 11명의 DAT 환자들을 대상으로 하여 2주 간격으로 실시한 검사-재검사 신뢰도와 3명의 채점자간 신뢰도를 살펴보았다.

방 법

연구 피험자

143명의 건강한 장노년층 정상인들과 20명의 DAT 환자들이 연구에 참여하였다. 정상 집단은 서울-경기

지역의 종교, 복지 기관, 직장, 가정 방문 및 광고 등을 통해 모집되었으며, 다음의 기준에 부합한 사람들이었다. 첫째, MMSEK(권용철과 박종한, 1989)가 22점 이상인 자; 둘째, 신경과 질환의 과거력이 없는 자; 셋째, 고혈압, 심장병, 당뇨 및 두부 외상의 경험이 없는 자; 넷째, 정신과 질환의 과거력이 없는 자; 다섯째, 나이가 만 55세에서 84세 사이인 자. 정상 피험자에 대한 인구통계학적 정보는 표 1에 제시되었다. 4개의 기준의 기초가 된 하위 표집군들에 대한 정보가 세부적으로 명시되어 있다. 20명의 DAT 환자(남:녀=3:17)는 평균 연령이 66.6 ± 9.3 세였고, 평균 교육 연한이 7.3 ± 3.6 년이었다. 삼성 서울 병원 신경과에서 치료를 받고 있는 외래 환자들로서 다음의 기준에 부합하였다. 첫째, NINCDS-ADRDA (McKhann et al., 1984)의 probable DAT 기준에 합치하는 자; 둘째, 경도(mild)에서 경중도(moderately severe) 정도의 DAT 환자로 심한 중증의 환자가 아닌 자; 셋째, 치매 이외의 신경 병력이 없는 자; 넷째, 두부 외상의 경험이 없는 자.

신뢰도 검증을 위하여 재검사를 실시한 피험자들은 연령과 학력이 다양하게 포함되도록 표집되었다.

도 구

K-DRS는 DRS(Mattis, 1988)를 한국어로 전환한 치매 평가 도구로 5개의 소검사로 구성되어 있어 다양한 인지 기능을 측정하도록 고안되었다. 5개의 소검사는 주의(Attention), 관리 기능(Initiation/Persevera-

표 1. 정상 피험자의 인구 통계학적 자료

		사례수	연 령	교 육	성비(남/여)	MMSEK
연 령	학 령					
55~64	0~6	35	59.4 ± 3.0	4.0 ± 2.7	.75	25.9 ± 2.2
	7~	37	59.0 ± 3.2	12.6 ± 3.1	1.85	26.8 ± 1.7
65~84	0~6	36	72.9 ± 5.1	2.8 ± 2.7	.64	25.3 ± 2.2
	7~	35	72.6 ± 4.8	11.4 ± 2.4	.94	26.3 ± 1.9
계		143	66.0 ± 8.0	7.7 ± 5.1	.96	26.1 ± 2.1

tion), 구성(Construction), 개념화(Conceptualization), 기억(Memory)이 있고, 이 소검사들의 점수를 종합한 144점이 만점으로 되어 있다(과제 및 자극에 대한 좀 더 자세한 설명은 최진영, 1998, 과 최진영과 이소애, 1997, 을 참조하기 바람). DRS를 전환하는 과정에서 K-DRS의 내용 타당도를 충족시키고자 다음과 같이 수정, 보완하였다.

K-DRS의 주의 소검사에서 변경된 내용은 다음과 같다. 두 종류의 간섭된 셈 과제(AD, AE)에서 “A”를 찾아서 그 개수를 세는 것을 “2”을 찾는 것으로 수정하였다. “언어적인 재인-제시(AH)”에는 5개의 단어가 글로 제시되는데 원 검사 자극들 중에서 “evening”, “plant”, “machine”을 각각 “새벽”, “장미”, “비행기”로 바꾸었다(연세대 사전 편찬 위원회, 인쇄증, 이관용, 1991). 이는 직역인 “저녁”, “식물”, “기계” 등의 어휘 사용 빈도나 친숙성이 원 검사와 차이가 있어 수정한 것이었다. 예를 들면, 국어에서의 “저녁”은 영어의 “evening”에 없는 식사 의미의 영어 어휘 사용 빈도가 “evening”보다 높고 “식물”은 영어의 “plant”에 비해 노인들에게 친숙성이 떨어진다고 판단되었다. 관리 기능의 원 검사에서 수정된 내용은 “자음 보속성” 과제(G)의 하나인 “bee-key-gee”를 반복하여 외우는 대신 “카타파”를 하도록 한 점이다.

개념화 소검사에서는 “상이성(Y)”과 “언어적 회상-문장 작성(AB)” 과제가 원 검사와 다른 자극을 사용하였다. 세 개의 어휘 중 나머지 두 개와 다른 하나를 고르는 과제인 “상이성”의 첫째와 셋째 문항의 자극들이 표 2에서 제시된 것과 같이 변경되었다.

DRS에서는 문항 1의 자극인 “cat”과 “car”가 모두 비슷한 음소로 시작하기 때문에 모두 비슷한 소리

“ca”로 시작한다는 구체적인 오반응을 하여 “dog”이 다르다는 오답을 하는 것이 가능하다. 이에 반해 직역된 자극 “고양이”와 “자동차” 사이에는 공통된 음소가 없어 자극을 바꾸어 공통된 음소가 있게끔 수정하여 제시하였다. 즉, 공통 음소 “자”가 “자동차”와 “자두”에 있어 “자두”와 개념적으로 공통된 “복숭아”와 함께 제시할 경우 “복숭아”가 소리가 달라서 다르다는 오반응이 가능하도록 하였다. 즉, 공통 개념이 동물이 아닌 과일 군으로 수정되었다.

“언어적 회상-문장 작성” 과제의 경우 원 검사는 “man”과 “car”라는 두 개의 단어를 이용하여 문장을 작성하게 하는데 K-DRS에서는 이 단어들을 “아주머니”와 “버스”로 바꾸었다. “man”의 직역인 “사람이나 남자”보다 구어에서 많이 사용되고 친숙성이 높은 “아주머니”와 “자동차(car)”보다는 노인들에게 친숙한 “버스”로 바꾸었다.

K-DRS의 기억 소검사에서 수정된 것은 “지남력” 과제(AC)와 “언어 재인” 과제(AI)였다. DRS의 지남력 과제는 현재 본인이 살고 있는 곳의 대통령, 주지사, 시장 등을 묻는 문항들로 구성되어 있으나 아직 지방자치제의 역사가 짧은 국내 사회에서 모든 노인들이 이러한 것들을 알고 있다고 전제하기가 곤란하여 K-DRS에서는 시간, 계절 및 검사 장소에 관한 질문으로 수정하였다. 언어 재인 과제는 앞서 제시된 “언어적인 재인-제시”에서 읽은 단어들을 재인하는 과제로 앞에서 자극이 수정된 데 따라 재인 과제의 보기 또한 다소 수정되었다. 표 3에서 볼 수 있듯이 직역의 경우 국어에서 어휘 빈도가 매우 낮은 “인치”를 “폭”으로 바꾸고 “evening(저녁)”과 혼동을 주고자 한 “night(밤)”을 “새벽”과 혼동될 수 있게 “아침”으로 바꾸고, 국어의 “머리”보다는 어휘 빈도가 낮은

표 2. 상이성 과제(Y)의 수정된 자극

문항번호	DRS 자극	직역된 자극	K - DRS 자극
1.	dog - cat - car	개 - 고양이 - 자동차	복숭아 - 자두 - 자동차
3.	fish - car - train	생선 - 자동차 - 기차	생선 - 자동차 - 버스

굵은 글씨가 수정된 자극입니다.

표 3. 언어적 재인 과제(AI)의 보기

DRS	K - DRS
evening - head	새벽 - 다리
inch - plant	폭 - 장미
land - open	하늘 - 열다
machine - night	비행기 - 아침
fire - milk	불 - 우유

굵은 글씨가 수정된 자극입니다.

“head” 등을 원 검사와 비슷한 단어로 수정하였다.

요약하면, 한국판 치매 평가 검사(K-DRS)는 언어와 문화가 다른 지역에서 개발된 심리 검사를 번역하는데서 발생하는 문제점을 최대한 보완하여 국내 치매 환자를 진단하는데 타당도를 유지시킬 수 있도록 K-DRS의 자극과 지시들을 재 표준화하였다. 언어적인 자극들은 DRS의 자극들을 번역한 후 국어와 한국 문화에 맞도록 수정되었고 어휘 사용 빈도나 함축되는 의미들을 고려하여 원 문항이 측정하고자 한 바를 최대한 충실히 측정할 수 있게 하여 내용 타당도를 높이고자 하였다.

절 차

K-DRS의 본 검사에 앞서 간단한 신경심리학적 면접과 노인용 MMSEK(권용철과 박종한, 1989)로 이루어진 사전 스크리닝 과정이 모든 피험자들에게 실시되었다. 이 절차는 피험자의 기본 인적 자료와 더불어 피험자가 각 집단의 포함 기준(inclusionary criteria)에 부합하는지 여부를 검토하기 위한 것이었다. K-DRS 시행은 표준적인 절차에 따라 이루어졌으며(자세한 실시 절차는 최진영, 1998, 을 참조하기 바람), 노안이나 청력 약화 같은 비인지적인 기능의 장애로 검사 수행이 영향을 받지 않게 하기 위하여 면접시 이를 확인하고 필요하다면 돋보기를 사용할 수 있도록 제공하거나 지시문을 크게 읽어 주었고 이러한 조정이 불가능한 경우에는 검사를 실시하지 않았다.

결 과

K-DRS의 타당도를 경험적으로 검토하기 위하여 세 가지 접근 방법을 취하였다. 첫 번째 접근은 K-DRS의 구성 타당도를 살펴보는 것이었다. 구성 타당도란 검사가 측정하고자 한 이론적 개념 혹은 특성을 얼마나 잘 측정하는지를 살펴보는 과정으로 타당화 과정에서 가장 핵심이 되는 것으로 간주되고 있다(Anastasi & Urbana, 1997). 현재 사용되고 있는 전반적인 인지 기능을 측정하는 검사 중 노인 인구에게 많이 실시되고 있고, 치매 진단에 활용되고 있는 MMSEK(권용철과 박종한, 1989)와의 관련성 정도를 검토함으로써 인지 측정 도구로서의 타당도가 검토되었다.

두 번째 접근 방법은 K-DRS의 진단적 유용성을 알아보기 위하여 검사가 정상인과 DAT 환자를 변별 진단할 수 있는 정도를 살펴봄으로써 기준 타당도를 확인하는 작업이었다. K-DRS의 목적이 치매 환자를 진단하고 분류하는 것이므로 다른 검사에 비해 기준 타당도의 검증이 매우 중요하다. 이와 더불어 세분화된 규준의 사용이 기준 타당도를 증가시킬 가능성을 살펴보았다. K-DRS 수행은 연령이 높을수록 또 학력이 낮을수록 감퇴된다는 것을(최진영과 이소애, 1997) 앞서 언급하였다. 그러므로, 학력과 연령으로 세분화된 하위 규준을 사용할 때 검사의 기준 타당도가 증가할 것으로 예측되어 4개의 하위 규준으로 나누었을 때의 기준 타당도의 변화를 알아보았다. 또한 흔히 평균 - 2 표준편차로 경계점(cut off score)을 산출하여 인지 기능의 정상 여부를 나누는 기준점으로 사용하는데 K-DRS에서 이 방법을 사용하는 것의 타당성 또한 살펴보았다. 마지막으로 DAT를 변별하는데 특히 유의미한 소검사가 무엇인지 살펴보았다.

신뢰도는 2주 간격으로 실시된 검사와 재검사 총 점간의 상관과 3명의 채점자간 점수의 상관을 계산하여 살펴보았다.

구성 타당도

한국판 MMSE 혹은 MMSEK(권용철과 박종한, 1989)는 국내에서 표준화되어 사용되어지고 있는 인지 기능 검사 중 치매 진단을 위해 가장 많이 실시되고 있는 검사이다. 물론 치매 환자의 진단을 위해 개발된 도구는 아니지만 전반적인 인지 기능을 측정하는 도구이기에 K-DRS와의 상관관계가 유의미하게 나타날 것으로 예측되었다. 총 163명의 정상 및 DAT 피험자들의 MMSEK 점수와 K-DRS 점수와의 상관관계를 분석한 결과, 상관 계수(Pearson r) .82를 얻었다($p < .01$).

기준 타당도

전체 표집을 대상으로 기준 타당도를 구하였는데 진단(정상 집단 대 DAT 집단)과 K-DRS 총점간의 양의 상관 계수(point biserial correlation: r_{pb})는 .66($p < .01$)이었고, 정상 집단과 DAT 환자 집단간에 총점 평균에서의 차이는 유의미한 것으로 나타났다($F = 126.84$, $p < .0001$). 이는 K-DRS가 두 집단을 구별해 주고 있다는 것을 보여주는 결과였다. 4개의 하위 집단별로 규준을 세분화하였을 때 각 집단 내에서의 진단과 K-DRS 총점간의 상관관계를 산출하였으며 그 결과는 표 4에 제시되어 있다.

고령 저학력 노인 집단을 제외하고는 전체 표집을 대상으로 했을 경우 보다 상관 계수들이 모두 상승

표 4. 4개 집단의 K-DRS 총점과 진단간 상관

연 령	학 력	r_{pb}	N
55 ~ 64	≤ 6	.866	39
	> 6	.765	42
65 ~ 84	≤ 6	.597	46
	> 6	.872	36

하였다.

이상 수행(abnormal performance)의 기준으로 많이 사용되는 평균에서 표준편차의 2배를 감하여 얻은 경계 점수가 정상인과 DAT 환자를 올바르게 변별할 확률을 구한 결과 비교적 높은 수준의 변별력이 관찰되었다. 전체 정상 피험자를 기준으로 구해진 경계 점수는 평균 125.1에서 표준편차 9.5의 2배 값을 감한 106.1이었다. 이를 기준으로 할 경우 정상인을 정상으로 올바르게 판별한 비율이 97.2%였고 DAT 환자를 DAT로 올바르게 판별한 비율이 80%였다. 4개의 하위 집단 각각에 대해 평균과 표준편차를 기준으로 산출한 경계 점수들의 판별력이 표 5에 제시되었듯이 정상인을 정상으로 올바르게 판별할 비율이 고령 저학력 집단에서 증가하고 다른 집단에서는 비슷한 수준을 유지하였다. DAT 환자를 올바르게 판별할 비율은 저령 저학력과 고령 고학력 집단에서는 증가하고 나머지 두 집단에서는 감소하였다.

표 5. 4개 하위 집단의 경계 점수를 이용한 변별력

연 령	학 력	경계 점수	분 류		비 율(%)		총 계(n)
			정 상		DAT		
			정 상	환 자	정 상	환자	
55~64	≤ 6	103.1	97.1 (34)	2.9 (1)	0 (0)	100 (4)	39
	> 6	119.4	97.3 (36)	2.7 (1)	80 (4)	20 (1)	42
65~84	≤ 6	116.4	100 (36)	0 (0)	50 (5)	50 (5)	46
	> 6		97.1 (34)	2.9 (1)	0 (0)	100 (1)	36
계		106.1	97.2(139)	2.8 (4)	20 (4)	80(16)	163

소검사의 변별력

5개의 소검사들이 DAT를 변별하는 정도를 고찰하기 위하여 일반 선형 모델(General Linear Model)을 통한 분석을 시행하였다. 그 결과는 표 6에 제시되었다. 소검사들 중에서 기억 소검사가 가장 유의미했으며($p < .0001$), 다음으로는 관리 기능 소검사가 유의미한 것으로 나타났다($p < .006$).

표 6. 5개 소검사의 AD와 정상 변별 정도

소 검 사	F	유의도	관찰된 검증력
주 의	3.68	.057	.478
관리기능	7.73	.006	.789
개 념 화	3.62	.059	.472
구 성	.31	.578	.086
기 역	98.47	.000	1.000

General Linear Model

신뢰도

2주 간격으로 실시한 검사-재검사 신뢰도는 .96으로 높은 수준이었으며, 3명의 채점자들간의 신뢰도는 .99로 매우 높게 나타났다($p < .0001$).

논 의

K-DRS의 타당도가 선형적으로 보완되었으며 경험적으로 구축되었고 신뢰도 또한 입증되었다. 전반적인 인지 기능 검사인 MMSEK와의 유의미한 상관을 통하여 K-DRS가 전반적인 지적 기능을 측정한다는 것을 보여주었다. MMSEK와 K-DRS 총점간의 .82라는 상관 계수는 DRS의 구성 타당도 연구(Chase et al., 1984; Coblentz et al., 1973)에서 얻어진 .67~.86의 상관 계수와 비슷하거나 높은 수준으로 K-DRS의 구성 타당도가 지지되었다. 즉, K-DRS의 구성 타당도는 원 검사와 유사한 수준임을 시사한다.

진단과 K-DRS 총점간의 상관을 통해 살펴본 기준

타당도 또한 지지되었다. K-DRS의 진단 기준 타당도도 DRS 연구와 유사한 수준의 변별력을 지니고 있는 것이 증명되었으며 원 DRS에서는 사용하지 않은 세부 규준의 사용으로 그 변별력을 더 향상시킬 수 있었다. 기준 타당도를 상승시키기 위하여 연령과 교육에 따라 규준 집단을 세분화하였을 때 한 집단을 제외한 세 집단에서 기준 타당도가 증가되는 것이 관찰되었다. 이를 통해 DAT 환자의 진단시, 피검자의 연령과 교육 수준에 해당하는 세분화된 규준을 사용함으로써 더 정확한 진단을 얻을 수 있을 것이라는 추론이 가능하였다. K-DRS 총점은 정상과 DAT를 잘 변별하였으나 판별력 결과는 추후 연구에서 좀 더 많은 환자들을 포함시켜 이를 재확인해야 할 필요성을 시사하기도 한다.

소검사별로 나누어 살펴보았을 때 기억과 관리 기능 소검사가 DAT를 탐지하는데 더 유용하다는 결과를 얻었다. 이는 DRS 연구 결과(Monsch et al., 1995)와 일관된 것으로 소검사 중 기억과 관리 기능 검사가 진단적 변별력에 가장 크게 기여한다는 것을 의미한다.

검사-재검사 신뢰도와 채점자간 신뢰도는 매우 높은 것으로 나타났는데, 특히 .99에 달하는 채점자간 신뢰도는 K-DRS의 실시 및 채점 기준을 엄격하고 명료하게 설정한 결과에 따른 것으로 생각된다. 그러나 3명의 검사자가 사전에 검사의 실시와 채점에 대해 집중적으로 함께 훈련받았다는 사실을 간과할 수는 없을 것이다.

본 연구에서 드러난 고령 저학력 노인의 상대적으로 저조한 진단 변별력을 높이기 위해서는 65세 이상 저학력 노인 인구에 대한 자세한 후속 연구가 절실히 요구된다. 이는 국내 65세 이상 노인의 50%가 저학력으로 분류되며 특히 여자 노인 중 90% 정도가 저학력으로 분류된다는 점(통계청, 1997)을 감안하면 이 집단의 진단 정확성을 높이는 것은 임상적으로 매우 중요한 일이다. 부연하자면 65세 이상의 정상 노인들 중에 K-DRS 수행이 매우 저조한 경우가 적지 않게 있어 초기 DAT 환자들의 수행 점수와의 변별력이 다른 집단에 비하여 떨어진 것이 관찰되었는데

이의 원인을 다각도로 탐색해 볼 필요가 있다고 생각된다. 첫째로, 아직 발견되지 않은 DAT 환자가 정상인으로 포함되었을 가능성이다. 이에 대한 여부는 종단적인 연구를 통하여 확인해야 할 것이다. 본 연구의 정상인 기준이 엄격했기 때문에 이 가능성은 크지 않다고 생각되나 이를 확인할 필요는 있다고 생각된다. 둘째로, 인지 기능 발달에 중요한 교육의 영향을 고려해 보아야 한다고 생각된다. 우선 치매에 대한 기본 가정은 이전에 습득된 인지 기능이 일정 시간 이후 지속적으로 감퇴되는 것으로 전제되어 있다 (Cummings & Benson, 1992). 그러나 국내 노인들 특히, 고령 저학력 노인의 경우 정규 교육 경험의 부족으로 고등 인지 능력이 충분히 발달되지 못했을 것이라는 추론이 가능하다. 본 연구에서는 표집 지역이 다른 지역에 비해 비교적 교육 수준이 높고 문화적 자극이 많은 서울-경기 지역으로 제한되었음에도 불구하고 이러한 가능성이 시사되므로 전국 노인을 대상으로 할 경우 정규 교육 부족의 영향은 더욱 심할 것으로 예상된다. 따라서 보다 광범위한 표집 지역을 대상으로 추후 연구가 계속될 필요가 있다. 본 연구 결과 나타난 전체 집단에 비해 고령 저학력 노인 집단에서 상대적으로 낮아진 기준 타당도와 진단 변별력을 보완하기 위해 교육 효과가 가장 크게 나타난 개념화 소검사의 점수를 제외한 조건에서 판별 분석을 실시하여 그 차이를 비교해 보았다. 그 결과 모든 소검사를 포함한 경우 정상 대 DAT 환자 집단의 전체 변별력이 93.9%인데 비해 개념화 소검사를 제외한 경우 전체 변별력이 95.9%로 다소 상승하였다. 더 나아가 연령과 학력으로 구분한 4개 하위 집단 각각에서 개념화 소검사를 제외한 경우의 변별력의 차이를 살펴본 결과 고령 저학력 집단에서만 전체 판별률이 82.6%에서 89.1%로 상승함을 발견할 수 있었다. 이러한 결과는 K-DRS를 사용하여 65세 이상의 고령 저학력 노인의 치매 여부를 감별할 때는 개념화 소검사를 제외하고 사용하는 것이 더 타당할 가능성을 시사한다. 그러나 본 연구의 환자 집단의 표집 수가 충분히 많지 않았으므로 이 가능성을 일반화하기 위해서는 보다 더 많은 환자 집단을 대상으로 한 연구

가 필요할 것이다.

또한 기본적인 교육이 인지 발달에 필수적인 것은 물론 노화에 따른 인지 기능의 감퇴를 둔화시켜주는 역할을 하고 있어 저학력자들의 기저 인지 기능이 저조하였을 뿐만 아니라 노화에 따라 인지 기능의 감퇴가 급속하게 일어났을 가능성이 있다. 본 연구에서 학력을 기준으로 하여 여러 개의 하위 집단으로 나누어 분석해 본 결과 무학의 노인들에게서만 연령 효과가 유의미하게 관찰된 점이 이를 지지한다. 즉, 무학 노인들만이 연령이 증가함에 따라 K-DRS 총점이 유의미하게 감퇴하는 추세를 보였다. 현재 국내 DAT 발병률이 선진국 보다 높게 관찰된 것(우종인 등., 1997; Park, Ko, Park, & Jung, 1994)도 기본 교육 성취의 저조함이 국내 노인들을 치매 위험 인구가 되도록 했을 가능성을 시사하는 것이다.

셋째로, K-DRS와 같은 구미에서 유입된 신경심리학적 검사 도구가 국내 저학력 노인에게는 생소하여 이들에게 특히 불리했을 가능성이다. 정상적인 인지 노화의 핵심적인 특징이 새로운 자극을 처리하는 능력이 감퇴하는 것이기 때문에(Malec, Ivnik, & Smith, 1993) 학교 과제와 유사한 성격을 띠는 신경심리학적 검사들은 학교에서의 학습이 거의 없었거나 매우 적었던 국내 저학력 노인들에게 아주 생소한 자극이 되어 이들의 수행을 저조하게 하였을 가능성이 있다. 이러한 의문점들을 검증하는 후속 연구는 국내 치매 환자는 물론 우리와 사회 경제적인 구조가 유사한 개도국의 치매 진단 및 연구에 중요한 디딤돌이 될 것으로 사려된다.

추후 K-DRS의 타당도가 더 보완되기 위해서는 K-WAIS(염태호, 박영숙, 오경자, 김정규, 이영호, 1992)나 다른 심리 측정 도구들과의 상관을 구해 구성 타당도를 더 심화시키는 작업과 더 크고 다양한 환자들을 표집하여 기준 타당도를 살펴보는 작업이 필요할 것이다. K-DRS의 진단 기능 못지 않게 중요한 경과 측정 도구로서의 타당성을 입증하기 위해 손상 정도가 심한 DAT 환자들을 포함하여 이들을 구별하는 점수와 프로파일을 구해 보는 작업 역시 필요하다. 최근 강조되는 생태학적 타당도를 구축하기 위해 치매

환자의 일상 생활 지표들과 K-DRS 점수간의 관계를 고찰해 보는 연구도 도움이 될 것으로 생각된다.

요약하자면, 본 연구는 K-DRS의 타당도와 신뢰도가 국내 노인 인구에서도 유지되고 있다는 것과 연령과 학력 수준의 효과를 감안한 세분화된 규준의 제시가 좀 더 타당한 치매 도구로써의 역할을 하도록 돕는다는 것을 입증하였다. 보다 다양한 표집군과 다양한 방법으로 추후 K-DRS의 타당도와 신뢰도 연구는 계속될 필요가 있다고 생각된다.

참고문헌

- 권용철, 박종한 (1989). 노인용 한국판 Mini-Mental State Examination (MMSE-K)의 표준화 연구. *신경정신의학*, 28, 125-135.
- 보건복지부(1996). 노인 복지. pp. 277-286. 보건사회 백서. 서울: 저자.
- 연세대 사전 편찬 위원회. 연세 말 뭉치 I. 인쇄중.
- 염태호, 박영숙, 오경자, 김정규, 이영호 (1992). 한국형 성인 지능 검사: Korean- Wechsler Adult Intelligence Scale. 서울: 한국가이던스.
- 우종인, 이정희, 유근영, 홍진표, 김창엽, 김용익, & 이강욱 (1997). 한국의 한 농촌 지역에 거주하는 노인에서의 치매 유병율. *신경정신의학*, 36, 93-101.
- 이관용 (1991). 우리말 범주 규준 조사. *한국심리학회지: 실험 및 인지*, 3, 131-160.
- 최진영, 이소애 (1997). 한국판 치매 평가 검사의 규준 연구. *한국심리학회지: 임상*, 16, 2, 423-433.
- 최진영 (1998). 한국판 치매 평가 검사: Korean-Dementia Rating Scale. 서울: 학지사.
- 통계청 (1997). 그래프로 보는 통계 [전자자료]. 서울: 통계청 [출판자와 배포자].
- American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders (4th ed.)*. Washington, D. C.: Author.
- Anastasi, A. & Urbina, S. (1997). *Validity: Basic concepts. Psychological Testing (7th ed.)*. pp. 113-139. New Jersey: Prentice Hall.
- Chase, T. N., Foster, N. L., Fedio, P., Brooks, R., Mansi, L., & Chiro, G. D. (1984). Regional cortical dysfunction in Alzheimer's disease as determined by positron emission tomography. *Annals of Neurology*, 15(Suppl.), S170-S174.
- Coblentz, J. M., Mattis, S., Zingesser, L. H., Kasoff, S. S., Wisniewski, H. M., & Katzman, R. (1973). Presenile dementia: Clinical aspects and evaluation of cerebrospinal fluid dynamics. *Archives of Neurology*, 29, 299-308.
- Cummings, J. L., & Benson, D. F. (1992). *Dementia: A Clinical approach*. Boston: Butterworth-Heinemann.
- Gardner, R., Oliver, S., Fisher, L., & Empting, L. (1981). Mattis Dementia Rating Scale: Internal reliability study using a diffusely impaired population. *Journal of Clinical Neuropsychology*, 3(3), 271-275.
- Kaszniak, A. W. (1986). The neuropsychology of dementia. In I. Grant & K. M. Adams (Eds.), *Neuropsychological assessment of neuropsychiatric disorders*. New York: Oxford University Press.
- Lezak, M. D. (1995). *Neuropsychological assessment (3rd ed.)*. New York: Oxford University Press.
- Malec, J. F., Ivnik, R. J., & Smith, G. E. (1993). Neuropsychology and normal aging: The clinician's perspective. In R. W. Parks, R. F. Zec, and R. S. Wilson (Eds.), *Neuropsychology of Alzheimer's disease and other dementias*. New York: Oxford University Press.
- Mattis, S. (1988). *Dementia Rating Scale (DRS)*. Odessa, FL: Psychological Assessment Resources.
- 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 Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, 34, 939-944.
- Monsch, A. U., Bondi, M. W., Salmon, D. P., Butters, N., Thal, L. J., Hansen, L. A., Wiederholt, W. C., Cahn, D. A., & Klauber, M. R. (1995). Clinical validity of the Mattis Dementia Rating Scale in detecting Dementia of Alzheimer Type: A double cross-validation and application to a community-dwelling sample. *Archives of Neurology*, 52, 899-904.
- Nadler, J. D., Richardson, E. D., Malloy, P. F., Marran, M. E., & Hostetler Brinson, M. E. (1993). The ability of the Dementia Rating Scale to predict everyday functioning. *Archives of Clinical Neuropsychology*, 8, 449-460.
- Park, J. H., Ko, H. J., Park, Y. N., & Jung, C. H. (1994). Dementing among the elderly in a rural Korean Community. *British Journal of Psychiatry*, 164, 796-801.
- Teri, L., Borson, S., Kiyak, A. & Yamagishi, M. (1989). Behavioral disturbance, cognitive dysfunction, and functional skill: Prevalence and relationship in Alzheimer's disease. *Journal of the American Geriatrics Society*, 37, 109-116.
- Vitaliano, P. P., Breen, A. R., Russo, J., Albert, M., Vitiello, M. V., & Prinz, P. N. (1984). The clinical utility of the Dementia Rating Scale for assessing Alzheimer patients. *Journal of Chronic Disorders*, 37, 743-753.
- Wechsler (1955). Wechsler Adult Intelligence Scale. New York: The Psychological Corporation.
- Wechsler (1945). A standardized memory scale for clinical use. *Journal of psychology*, 19, 87-95.

The Validity and Reliability of the Korean Dementia Rating Scale

Jeanyung Chey

Sungshin Women's University

Duk-Ryul Na

Sung Kyun Kwan Medical School
Samsung Medical Center

Sun-Hee Park

Eun-Hee Park

Sungshin Women's University

The validity and reliability of the Korean-Dementia Rating Scale (K-DRS) have been examined. 143 normal elderly and 20 patients diagnosed with the dementia of the Alzheimer's type(DAT) participated in the study. Three approaches were pursued to examine the different aspects of validity. First approach focused on the content validity of the K-DRS. Efforts were made to ensure that the translated tasks would be tapping the same cognitive processes that were purported to be examined in the original Dementia Rating Scale (Mattis, 1988). Secondly, we employed an empirical approach to establish the construct and criterion validity of the test. Correlation between the total scores of the K-DRS and the MMSEK(Kwon & Park, 1989) was sought for the construct validity, which yielded a correlation coefficient of .82. Criterion validity with diagnostic states as the criterion was examined. The correlation between the diagnoses and the K-DRS total scores was .66 when all participants were included in the group. The correlations increased when the normative group was divided by age and years of education, except for one subgroup. The third approach involved examining how well each subtest discriminated the DAT, which showed that Memory and Initiation contributed significantly. It was concluded that K-DRS is a valid dementia assessment tool in terms of its content, construct and criterion validities. Test-retest reliability of two weeks was examined by testing 20 normal elderly adults and 11 DAT patients, yielding .96. Inter-rater reliability involving three raters was .99.