

재활분야의 결과측정에서 라쉬분석 활용에 관한 고찰

이재신*, 박소연**

*건양대학교 작업치료학과

**세연아동발달연구소

국문초록

재활분야에서 사용하고 있는 평가도구는 대부분 서열척도인 리커트 척도로 구성되어 있다. 이러한 평가도구는 선택된 항목의 난이도 수준이나 분포에 따라 그 결과가 달라질 수 있으므로, 항목에서 획득한 총점을 환자의 기능수준으로 인식하거나 환자간의 능력을 비교하는 것은 제한적이다. 최근에는 서열척도를 등간척도로 변환시켜주고, 객관적으로 평가항목을 분석할 수 있는 문항반응이론의 활용이 대두되고 있다. 문항반응이론에 기초한 라쉬분석은 기존의 서열척도로 구성된 평가항목의 점수를 등간척도로 변환시킨 로짓점수로 항목의 난이도를 알 수 있고, 변환식에서 산출된 환자의 능력점수는 다른 환자와 객관적인 비교가 가능하다. 라쉬분석은 평가도구의 항목선정과 척도분석이 가능하여 평가도구 개발에 활용되고, 환자의 능력추정과 검사동등화를 실시할 수 있어 컴퓨터 개별화 검사(Computerized Adaptive Testing), 문제은행의 항목 제작 등에 사용되고 있다. 본문에서는 라쉬모형에 관한 이론적 배경과 함께 최근 재활분야에서 라쉬분석이 활용된 연구결과를 요약하였다. 현재 재활분야에서 사용하는 기존의 평가도구에 대한 신뢰도와 타당도를 검증하는 연구는 물론, 새로운 평가도구 개발에 라쉬분석이 활용된 연구가 대부분이었다. 평가항목의 단일구성(unidimensionality)정도를 알 수 있는 적합도 검증과 문항의 난이도 분석결과는 환자의 능력에 맞는 과제 및 항목을 선정할 수 있어 중재계획 및 치료에도 도움이 될 수 있다.

주제어 : 라쉬분석, 재활, 평가도구

I. 서론

일반적으로 널리 활용되고 있는 측정방법은 수리적이고 통계적인 면을 강조하는 양적 접근법으로, 문항의 수준이나 개인의 능력보다는 검사 및 집단을 중시하는 전통적인 측정이론과 방법에 의한 고전검사이론(Classical Measurement Theory)을 적용하여 사용하고 있다. 그러나 이러한 양적 접근법은 평가도구에서 획득한 점수에 의해 개인이나 집단별로 다른 사람과 비교하여 얼마나 능력이 많고 적은지를, 또는 시간이 지남에 따라 변하는 향상정도를 통계학적으로 유의한 진전정도로 해석하여

임상적인 의미로 기술한다(Velozo & Peterson 2001). 이것은 검사에 포함되는 문항의 난이도 수준과 분포에 따라 능력비교가 다르다는 것을 의미한다. 이처럼 고전검사이론의 문제점은 문항의 산출대상에 따라 다르고, 대상자의 능력은 비교집단에 따라 다르며, 점수로 능력을 추정할 때 오차를 동일하다고 보는 것과, 신뢰도를 모든 집단에 동일하게 보는 것이다. 임상에서 평가도구별 항목점수의 합으로 중요한 임상적 결정을 하지만 임상중재 가이드에는 제한된 정보를 가지고 있다. 왜냐하면, 총점은 동일하더라도 서로 다른 구성요소를 포함하므로 단일 구성을 보여주는 것이 부족하고, 그 의미가 불명확하기

교신저자 : 이재신(jaeshin@konyang.ac.kr)

|| 접수일: 2006.4.15 || 게재승인일: 2006.5.15

때문이다. 또한 대부분이 서열척도이므로 순서비교는 되지만 대상자간 차이는 비교할 수 없는 제한점이 있다 (Chang & Chen, 1995). 이러한 고전검사이론의 단점을 보완할 수 있는 것으로 대두된 것이 현대적 검사이론으로 불리는 문항반응이론(Item Response Theory) 측정모델이다. 본문에서는 최근 널리 사용되는 문항반응이론의 이론적 배경과 함께, 재활분야에서 라쉬(Rasch) 분석을 활용한 연구를 제시함으로써 작업치료 영역에서의 적용 가능성을 알아보고자한다.

II. 본 론

1. 고전검사이론의 제한점

검사이론이란 평가도구의 질을 평가하기 위한 것으로, 신뢰도와 타당도를 검증하기 위한 방법과 절차에 대한 이론이다. 일반적으로 널리 활용되고 있는 측정방법은 수리적이고 통계적인 면을 강조하는 전통적인 측정이론에 의한 양적 접근법의 고전검사이론을 사용하고 있다. 고전검사이론은 평가도구에서 획득한 점수를 개인이나 집단별로 비교하여 얼마나 능력이 많고 적은지, 또는 시간이 지남에 따라 변하는 향상 정도가 통계학적으로 유의한 차이를 보인 결과에 따라 임상적인 의의가 있다고 해석한다. 고전검사이론은 수리적으로 직관적인 정확성 때문에 익숙하고 통계에 대한 이해가 쉽다는 장점을 가지고 있다(Downing, 2003).

그러나 고전검사이론은 검사 항목이나 검사대상자의 개인적인 능력보다는 검사 및 집단을 중시하여 성별이나 지역, 계층, 문화, 인종별 표본대상에 따라 결과에 차이가 있으며, 환자의 능력도 비교집단에 따라 달라진다. 또한 신뢰도를 모든 집단에 동일하게 보고, 점수의 총점으로 능력을 추정할 때 동일한 오차를 보인다고 간주한다. 이러한 총점은 검사에 포함되는 항목의 난이도 수준과 분포에 따라 다르게 나타날 수 있으며, 총점은 동일하더라도 능력 정도가 같다고 할 수 없으므로 수행 능력 정도를 기술하는 것이 제한적일 수밖에 없고, 도구간의 연계도 부족하다(Velozo & Peterson, 2001). 또한 임상에서는 평가도구로 측정한 총점으로 환자의 상태를 판정하거나 치료적으로 중요한 결정을 하지만 임상중재가이드로는 정보가 부족하다. 왜냐하면, 평가도구를 구성하고 있

는 항목들이 단일구성의 요소를 만족하고 있는지에 대한 정보가 부족하며 명확하지 않게 표현되어 있기 때문이다 (Chang & Chan, 1995; Downing, 2003).

2. 문항반응이론

고전검사이론의 제한점을 보완할 수 있는 것으로 대두된 것이 현대적 검사이론으로 불리는 문항반응이론이다. 고전검사이론이 표본집단 전체와 검사의 총점으로 항목특성과 환자능력을 추정하는 것에 비해 문항반응이론은 각각의 평가항목과 그 항목에 대한 반응을 근거로 항목의 특성과 환자능력에 관한 정보를 추정한다 (Piquero, MacIntosh, & Hickman, 2001).

문항반응이론은 환자의 잠재적 특성을 각 평가항목에 보인 반응으로 예측할 수 있다고 가정하여, 환자의 평가항목에 대한 반응과 잠재적 특성 사이의 관계를 확률에 의하여 모형화한 것이다. 항목에 대한 반응은 이분척도나 다분척도로 이루어진 서열척도의 평가도구와 동일한 척도로 관찰된다. 측정한 항목에 보이는 반응을 관찰하여 직접 측정하지 않은 항목에 대한 잠재적 특성을 확률적으로 분석하여 추정하고, 추정결과에 따라 측정하고자 하는 특성을 환자간에 비교할 수 있다(지은림과 채선희, 2000; Alphen, Halfens, Hasman, & Imbos, 1994).

문항반응이론을 적용하기 위해서는 국소독립성(local independence)과 단일 구성(unidimensionality) 요소를 충족시켜야 한다. 국소독립성이란 환자가 어떤 평가항목에 보인 반응이 다른 항목에 대한 반응에 영향을 주지 않는 상호 독립적이라는 의미이며, 단일 구성은 측정하고자 하는 고유한 특성만을 측정한다는 것이다. 문항반응이론을 적용하면 서열척도로 측정한 항목의 결과를 등간척도의 형태로 변환할 수 있다. 문항반응이론의 커다란 장점 중의 하나는, 추정치의 불변성으로 측정의 객관성을 가장 잘 만족시켜 준다는 것이다(Cella & Chang, 2000; Downing, 2003).

문항반응이론에 기초한 라쉬모형은 평가도구의 항목이나 척도의 제작 및 분석, 환자 능력 추정, 다른 평가도구로 평가한 환자의 능력이 비교 가능한 검사동등화(test equating) 등에 사용된다. 그리고 적은 수의 문항으로도 타당한 점수 획득이 가능한 컴퓨터 개별화 검사(computerized adaptive testing), 문제은행의 항목 제작,

운동경기의 결과판정이나 경기단의 성공과 실패 확률 예측, 의학적 임상진단, 물리치료 평가도구 개발에의 적용 등 다양한 인간과학분야에서 활용되고 있다(Bond & Fox, 2001). 또한 세계적으로 의학교육분야에서도 널리 사용된다. 예를 들면, 미국의 의사면허시험을 관리하는 National Board of Medical Examiner(NBME)에서는 시험에 관한 검증과 개발자에게 기본적인 심리측정 정보를 제공하는 데 이용하며, 아울러 간호학을 비롯한 전문의료분야의 면허시험에도 활용된다(Downing, 2003). 라쉬모형은 재활치료 과정 동안 임상적인 변화를 알아보기 위하여 반복측정하는 평가도구를 분석하는 데에도 활용되고 있다(Chang & Chan, 1995).

3. 라쉬모형

문항반응이론을 근거로 한 라쉬모형은 여러 모형 중에서도 그 이론이 간결하고 자료분석이 편리하여 교육이나 의학의 측정 및 평가분야에 널리 보급되어 있다. 1980년대 후반부터 재활치료 과정에서 사용하는 여러 평가도구에 적용되어 기존 평가도구의 단점을 보완하고 환자의 변화를 측정하는 데 유용하게 사용되고 있다(Nilsson, Sunnerhagen, & Grimby, 2005). 라쉬모형은 환자가 지니고 있는 능력이나 항목의 난이도에 따라 반응이 다양하게 나타나는 것을 실제 측정에서는 완벽하게 예측할 수 없기 때문에 로지스틱 함수를 이용하여 측정치산출로 추정한다(지은림과 채선희, 2000).

라쉬모형은 환자가 평가항목에 성공할 확률을 수식으로 나타낸 것으로, 환자 반응(vi)은 환자 능력(v)과 항목의 난이도(I) 중 어느 것이 얼마나 더 높은가에 의해 결정되는데 이 차이(v-I)에 따라 성공할 확률이 어떻게 변할 것인가를 나타낸 것이다(Alphen et al., 1994; Antonucci, Aprile, & Paolucci, 2002; Nilsson et al., 2005).

<라쉬모형>

$$P(v_i) = \exp(v-I) / \{1 + \exp(v-I)\}$$

라쉬모형은 항목의 난이도에 따라 환자능력별로 다른 반응을 보인다고 가정하므로, 난이도만을 고려한 1모수 로지스틱 모형과 동일하다. 1모수 모형에서 정답을 맞출

확률은 문항특성곡선에 의해 계산되고, 각 문항에서의 난이도상 기울기는 동일한 S자형의 곡선형태로 표시된다. 이것은 불가능한 능력수준과 검사반응간의 확률관계를 규명할 수 있다. 라쉬모형에서 환자 능력과 항목의 난이도 수준이 동일할 때 성공할 확률은 0.5가 된다. 환자의 능력이 항목의 난이도보다 높으면 0.5보다 높아지고, 능력이 낮을수록 0.5보다 낮아 0에 가까워진다. 다음 수식은 환자와 항목의 수준이 동일한 경우 성공할 확률을 보여주는 것이다(지은림과 채선희, 2000; MacKnight & Rockwood, 2000).

$$v-I = 0-0 = 0$$

$$\exp(v-I) = \exp 0 = 1$$

$$P(v_i) = \exp(v-I) / \{1 + \exp(v-I)\} = 1 / (1+1) = 0.5$$

라쉬모형을 평가도구에 적용하면 평가환자와 평가항목에 대한 적합도 검증이 가능하여 평가항목에 대한 부적합한 대상자와 부적합한 항목을 제거하고 적합한 대상자에서 관찰된 항목의 반응을 재분석할 수 있다. 재분석에서는 환자 능력과 항목 난이도 차이를 산출하여, 환자가 성공할 확률을 추정한 로짓(logit) 점수로 변환한다(MacKnight & Rockwood, 2000; Tennant, McKenna, & Hagell, 2004).

4. 모형 적합도(Goodness-of-fit)

라쉬모형은 관찰된 항목의 반응이 정답을 맞힐 확률에 대한 이론적인 반응유형과 얼마나 다른지를 평가할 수 있다. 이것은 적합도 지수라는 잔차분석으로 가능하다. 즉 모형적합도를 나타내는 통계치(fit statistic)인 평균자승잔차(mean square residual : MNSQ)는 라쉬모형에 적용가능성을 나타내는 지수로서, 관찰된 점수와 모형에 의해 예측되는 점수차를 말하는 것이다(Nilsson et al., 2005).

적합도 지수는 단일 구성을 평가하는지에 대한 타당도를 평가하는 지표로서, 내적합 지수(information-weighted fit statistics: infit)와 외적합 지수(outlier sensitive fit: outfit)로 나타낸다. 내적합 지수는 환자의 능력수준과 가까운 항목에 영향을 주는 반응에서 예기치 못한 행동에 더 민감한 지수이고, 외적합 지수는 환자의

능력수준과 동떨어진 항목에서 예기치 못한 행동에 더 민감한 지수라 할 수 있다. 이러한 적합도 지수를 이용하여 모형에 부적합한 항목이나 환자를 선별할 수 있다 (Handa & Massof, 2004; McHorney, Haley, & Ware, 1997; Nilsson et al., 2005; Wright & Linacre, 1992).

환자나 항목의 모형이 기대로부터 벗어난 정도를 나타내는 관찰치와 예측치는 1에 가까울수록 적합도가 높은 것으로 해석된다. 일반적으로 2.0 이상일 때 부적합 항목이나 부적합 환자로 판정을 내리게 된다. 2.0 이상이라는 것은 통계적 가설 검정 원리에 의하여 관찰될 확률이 0.05 이하라는 의미이다. 즉 측정 결과 환자나 항목의 적합도 통계치가 2.0 이상으로 부적합 판정을 할 때는 적합한 자료가 아니라는 의미이다. 적합도 판정은 일차원적 자료의 종류나 연구자에 따라 판정기준이 다르게 적용되고 있다(Bond & Fox, 2001).

항목이 부적합으로 판정되는 이유는 부적절한 용어를 사용하였거나 애매하게 표현되어 이해하기 어렵거나 측정하고자 하는 특성 이외에 다른 영향을 받는 내용을 포함하고 있기 때문이다. 부적합한 환자로 판정되는 원인에는 검사자의 추측이나 잘못된 판단, 환자의 불안상태 또는 경험 부족에 의한 문제, 특정분야의 다른 문제를 가지고 있는 경우 등이 있다. 이러한 진단결과는 반응형태분석으로 문제점을 발견하고 해결해주는 데 도움이 된다(지은림과 채선희, 2000; Wolfe, Michaud, & Pincus, 2004).

5. 난이도

라쉬모형을 적용하면, 리커트 척도로 평가한 서열척도인 원점수를 등간척도로 변환시켜 항목의 난이도나 환자의 능력을 일직선상에서 나열하여 환자들이 지닌 특성이나 능력을 규명할 수 있다. 즉 서열척도가 등간척도로 변환된 로짓점수를 통해서 단순히 항목들의 난이도 순서뿐만 아니라 항목들 간의 간격을 알 수 있기 때문에, 항목 간의 간격에 따라 난이도의 차이를 비교해 볼 수 있다 (Duncan, Bode, Lai, & Perera, 2003; Fisher, 2003).

대상자와 항목간의 관계를 나타낸 그림(items and persons map)의 중심선 왼쪽에는 환자의 능력에 대한 분포를 표시하고, 오른쪽에는 항목의 난이도에 대한 분포를 표시한다. 환자의 능력은 가장 능력이 낮은 사람이

맨 아래에 위치하고 가장 능력이 좋은 사람은 최상위에 위치한다. 그리고 가장 쉬운 항목은 맨 아래에 위치하며 가장 어려운 항목은 최상위에 위치하여 환자의 분포와 항목의 로짓점수를 구분하여 알 수 있다(그림 1. 예시). +와 -로 표시되는 로짓점수로 환자의 능력이나 항목의 난이도를 알 수 있으며, 0은 평균점수이다(Chang & Chan, 1995; Downing, 2003).

6. 라쉬모형의 장점

라쉬분석이 수리적으로 매우 복잡한 모형을 거쳐야 하는 단점에도 불구하고 지속적으로 사용되고 있는 이유는 다음과 같은 장점을 가지고 있기 때문이다. 첫째, 환자의 개인별 잠재적 능력의 수준을 정확하게 분석할 수 있다. 둘째, 잠재능력을 분석하여 수행 가능한 평가항목을 추정할 수 있다. 셋째, 추정된 항목특성은 환자집단의 특성에 영향을 받지 않는다. 즉 항목을 분석하면 능력이 높은 환자집단에서 평가가 실시되었든, 능력이 낮은 집단에서 실시되었든 추정된 항목의 특성은 변하지 않는다

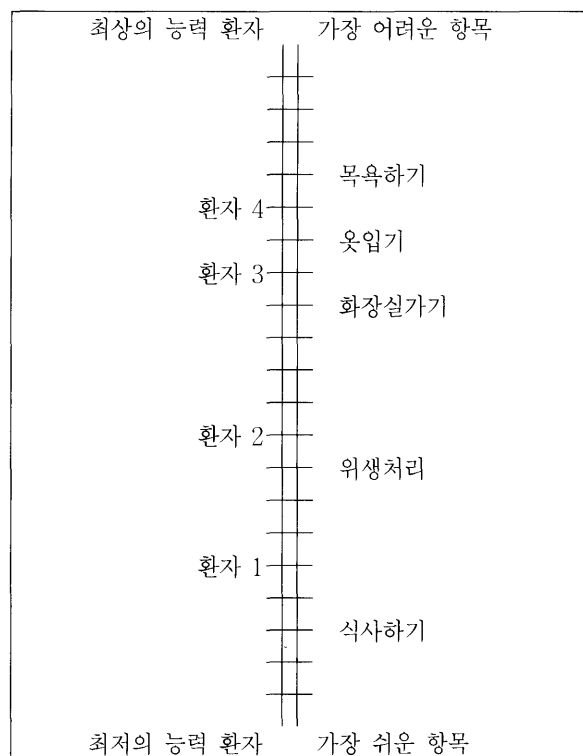


그림 1. 일직선상의 환자별 능력과 일상생활 난이도
(Fisher, 1993)

는 것이다. 넷째, 평가항목의 난이도에 상관없이 환자능력 추정이 가능하여, 환자의 응답에 대한 각각의 신뢰도를 추정할 수 있다. 다섯째, 다수의 영역을 적절하게 이용할 수 있도록 평가영역의 선택이 가능하다. 마지막으로 다른 척도모형과 직접적인 비교가 가능하다. 이러한

객관성으로 최근 라쉬모형의 활용이 증가하고 있다 (Piquero et al., 2001).

표 1. 재활결과측정도구 분석에 활용된 주요연구

저자, 연도	대상자	평가도구	결과
Hsueh et al., 2004	뇌졸중환자와 보호자와의 면담 (N=245)	- Barthel Index - FAI(Frenchay Activities Index)	바텔지표의 항목은 일반노인에게는 너무 쉬운 항목들로 구성되어 있어 부적합한 도구이지만 뇌졸중 환자에게는 단일구성 여부에 적합하다고 지적
Kucukdeveci et al. 2005	16세 이상의 일반인(N=406)	Turkish MMSE	터키에서 MMSE(간이정신상태검사)타당도검사의 사전연구에서 라쉬분석이 부적합(11개중 4 항목만 적합판정)
McHorney et al. 1997	만성환자 (N=3,445)	MOS SF-36 Physical Functioning Scale	리커트척도와 라쉬척도모형으로 신체기능비교, 모든 라쉬모형척도가 개별점수의 임상적 해석에 중요한 관련성
Duncan et al. 2003	급성뇌졸중 환자 (N=696)	Stroke Impact Scale	3개월 치료결과분석, 타당도, 신뢰도, 항목의 난이도
Kornetti et al., 2004	균형조절에 문제가 있는 환자 (N=100)	Berg Balance Scale	항목의 난이도, 척도조정
Veloza et al., 1995	입원재활환자(CVA, SCI, TBI 등) (N=3,056)	Level of Rehabilitation Scale-III	입원과 퇴원결정의 질적인 기능상태평가도구로 검증
Baker, Granger, & Roger, 1997	외래척수손상 재활환자 (N=94)	MRFA MOTrust SF-36, Therapist Ratings	MRFA의 근골격 관련항목과 MOT SF-36의 라쉬와 원점수로 내용타당도, 구성타당도, 기준타당도검사. 진전정도에 대한 치료사의 척도비교
Fisher et al., 1992	미국에서 3년 이하 거주한 타이완 성인 (N=20)	AMPS	AMPS의 처리기술항목의 문화적 차이에 대한 신뢰도, 타당도연구에서 제한 없이 나왔으나 타이완활동을 미국인에게 적용할 후속연구제언
MacKnight & Rockwood, 2000	입원환자 :허약한 노인 (N=204)	Hierarchical Assessment of Balance and Mobility	단일 구성요소분석, HABAM 개정
Malec, 2004	뇌손상 (N=134)	Mayo-Portland Participation Index	심리결과측정의 질적인 면 만족
Tsuji et al., 1999	정상발달아동(6개월~7세) (N=225)	WeeFIM	ADL항목의 적합도검증, 연령그룹별 난이도비교
Duran & Fisher, 1996	신체지체 및 정신지체장애 (N=3,500)	AMPS	운동능력은 남녀간의 차이가 없으며, 처리기술은 여자가 남자보다 약간 우세

표 1. 재활결과측정도구 분석에 활용된 주요연구 -계속-

저자, 연도	대상자	평가도구	결과
Granger et al., 1998	임상가(간호, 작업/물리/언어/오락치료, 복지, 의사) (N=6,645)	FIM Mastery Test	평균능력; 77logit, 분리신뢰도;17, 교육의 초점; 2~6 채점의 정확한 구분
Doctor et al., 2003	의료재활전문가(작업/물리/언어/심리치료, 간호) (N=50)	FIM	FIM척도의 부정확성과 임상적 과잉판단을 줄이는 것이 기능평가의 정확성을 높이는 것임
Chen et al., 2005	입원아동환자 (N=814)	WeeFIM	항목의 적합도 검증, 그룹별(3세 이하, 3~7세, 7세 이상) 난이도 제시
Chang & Chan, 1995	뇌졸중 (N=99)	FIM(입원1주,퇴원, 6주후)	항목난이도, 환자능력, 중등도의 측정으로 건강관리 연구에 최적
Bernspang & Fisher, 1995	우반구뇌졸중(n=71)/좌반구 뇌졸중(n=76)/정상인(n=83)	AMPS	두뇌졸중환자군은 운동기술은 차이가 있으나 IADL 능력차이없으며, 정상인보다 수행능력이 낮다.
Coster et al., 2004	재활치료환자 (입원,외래,가정) (N=477)	58Post-Acute Care FIM, MDS-PAC PF-10, SF-8	적합도검정: 56항목:적합, 6항목: 부적합 항목의 난이도 결과제시
Kjellberg et al., 2003	작업치료사(n=18), 정신장애, 학습장애 및 신경손상환자(n=67)	ACIS-S	ACIS-S(ACIS 스웨덴판)의 내적타당도, 구성타당도, 반응타당도검사결과로 양질의 측정도구입증, 의사소통과 상호관계능력의 항목적합도와 난이도 검증
White & Velozo, 2002	요통환자 (N=942)	Oswestry CS	새로운 척도, 항목의 적합도, 환자의 능력별 난이도 제시

MOS SF-36: Medical Outcome 36-item Short Form Health Survey

MRFA: Medical Rehabilitation Follow Along

MDS-PAC: Minimum Data Set (postacute setting)

FIM: Functional Independence Measure

WeeFIM: Functional Independence Measure for Children

AMPS: Assessment Motor and Process Skill

ACIS: Assessment of Communication and Interaction Skills

Oswestry (Low Back Pain Disability Questionnaire) Classification Scheme

7. 라쉬모형이 활용된 연구

라쉬모형은 1960년 덴마크의 수학자 Georg Rasch에 의해 발표되었고, 미국 시카고 대학 교육학과의 Benjamin D. Wright 교수에 의하여 실용화되고 이론적 발전을 거듭하였다(이충휘, 2004; 지은림과 채선희, 2000).

라쉬모형을 활용한 연구는 메트라인(Medline)과 국가과학기술전자도서관(National Digital Science Library: NDSL)에서 검색하였으며, 최근 재활결과를 측정하는 평가도구의 검증에 라쉬분석을 활용한 연구의 대상자와 타

당도에 관한 결과는 표 1과 같다. 국가별 문화와 언어의 차이를 비교한 평가도구의 타당도에 관한 연구는 주로 유럽에서 연구가 시행되었다(Tennant et al., 2004; Prieto, Novick, Sacristan, Edgell, & Alonso, 2003; Kjellberg, Haglund, Forsyth, & Kielhofner, 2003; Kucukdeveci, Kutlay, Elhan, & Tennant, 2005). 재활분야의 결과측정을 위한 새로운 평가도구의 개발을 위하여 라쉬분석을 활용한 주요연구는 표 2와 같다. 심리측정(psychometrics properties)의 평가도구에 대한 타당도 연구가 대부분으로 기존의 평가도구의 항목을 선택하여

대상자와 항목의 적합도 통계치로 부적합 항목 또는 부적합 환자로 판정을 내린다. 부적합 환자로 판정된 환자의 측정치를 제외하고 적합한 환자의 측정치로 재분석하여 적합한 항목과 항목의 난이도를 알 수 있다.

1980년대 이후 재활연구 분야에서 광범위하게 사용되고 있는 문항반응이론은 최근 평가분야에서 시험개발이나 데이터분석에 꾸준히 도입되고 있다. 이처럼 라쉬분석은 재활분야에서 많이 사용하고 있는 리커트 척도에 적용될 수 있어 기존 평가도구의 타당성 연구와 새로운 평가도구의 개발에 따른 항목의 단일 구성과 항목의 난이도를 확인하는 데 유용하다. 또한 각각의 문항에 대한 개인별 정보를 얻을 수 있어 개인 비교연구에 유용하며, 의학교육에서 해부학 및 임상실습 평가척도(Downing, 2003)에도 활용된다. 국내의 보건의료분야에서는, 국가시

험의 문항을 좀더 향상된 분석을 수행하기 위하여 문항반응이론으로 의사국가시험의 문항난이도와 분별도 분석(한국보건의료이국가시험원, 2006), 간호사국가고시에 문항반응이론의 적용타당성에 관한 연구(정금희와 임미경, 2005)가 있었다. 물리치료에서는 뇌성마비아동을 대상으로 실시한 한국판 대동작 기능 평가도구에 라쉬분석을 활용한 연구(박소연, 2005)와 작업치료에서는 라쉬분석을 이용한 뇌졸중 환자용 일상생활수행 평가도구 개발(이재신, 2006)이 있었다.

표 2. 평가도구 개발에 활용된 주요연구

저자, 연도	대상자	평가도구	결과
Coster et al., 2004	신경계, 근골격계, 복합의료장애 (N=477)	Post-Acute Care, Medical Outcome Study 8-item Short Form Health Survey, FIM, Mimimum Data Set	일상생활의 결과측정에 유용한 응용인지기능평가 항목 개발(Applied Cognition Scale)
Fujiwara et al., 2004	뇌졸중 (N=73)	Stroke Impairment Assessment Set trunk item, FIM, Trunk Control Test	뇌졸중 연구결과에 유용한 신뢰도, 타당도, 반응도의 새로운 TIS(Trunk Impairment Scale) 개발
Velozo & Peterson 2001	노인(평균 76세) (N=106)	University of Illinois at Chicago Fear of Falling Measure	평가항목의 적합성, 난이도를 통하여 환자의 능력에 따른 임상적 중재가이드제공
Hart & Wright, 2002	재활환자 (N=92,343)	Self-report health status, PF-10, Oswestry, Neck Disability Index, Lysholm Knee Questionnaire	신뢰도와 타당도 검증된 Functional Health Status(FHS-36)가 개발되어 역동적이며 컴퓨터로 그룹분석 및 환자개인의 임상적 의사결정에 적용
Tennant et al., 2004a	문제은행	Quality of Life Instrument	삶의 질에 관한 심리평가도구 개발에서 라쉬분석의 개념 및 적용에 대한 소개
Wolfe et al., 2004	NDB (N=16,519)	SF-36, WOMAC, VAS(Visual Analog Scale), Grobal VAS	Short Arthritis Assessment Scale의 타당도와 신뢰할만한 간단한 척도개발로 초기관리에 유용
Wolfe et al., 2004	관절염환자 (N=14,038)	HAQ 20항목에 31질문	HAQ-II 개발

WOMAC: Western Ontario McMaster Osteoarthritis Index

NDB: National Data Bank for Rheumatic Diseases

HAQ: Health Assessment Questionnaire

III. 결 론

문항반응 이론은 문항에 대한 반응과 잠재적 특성 사이에 관계를 수리적 함수로 나타낸 것이다. 즉 보이지 않는 잠재적 특성을 관찰되는 문항의 반응에 의하여 측정하고자하는 특성을 확률적으로 분석하여 추정하고 추정결과에 따라 대상자들을 비교할 수 있다. 문항반응이론에 기초한 라쉬모형은 평가대상자와 평가문항의 적합성을 검정할 수 있으며, 서열척도가 등간척도로 변환된 로짓점수로 환자의 능력이나 항목의 난이도를 알 수 있어 순서와 차이를 직접 비교할 수 있다. 또한 재활과정 중의 임상적 변화를 객관적으로 알아보는데 타당하다. 검정력이 우수하고 객관적인 장점을 가진 라쉬분석의 활용이 국내 재활분야에서는 연구보고가 미약하다.

라쉬분석의 활용이 최근의 연구경향이므로 작업치료에서도 다양한 평가도구에 대한 타당성 연구와 우리문화와 실정에 맞는 새로운 도구의 개발이 가능하며 활성화가 바람직하다. 이러한 연구결과는 작업치료와 관련된 모든 질환의 대상자에게 확대하는 연구가 필요하며, 일상생활수행에 관한 평가도구는 물론 심리측정요소가 가미된 모든 기능의 평가도구의 개발에 효과적이다. 이처럼 라쉬분석을 적용한 연구로 개발된 객관적인 평가도구는 질적으로 우수한 상위수준의 평가도구로서 작업치료 계획과 중재전략에도 효과적으로 활용될 것이다.

참 고 문 헌

박소연. (2005). *한국판 대동작 기능 평가도구의 Rasch 분석*. 서울: 연세대학교 박사학위논문.

이재신. (2006). *라쉬분석을 이용한 뇌졸중 환자용 일상생활수행 평가도구 개발*. 서울: 연세대학교 박사학위논문.

이충휘. (2004). *연구방법론*. 서울: 계축문화사.

정금희, 임미경. (2005). 간호사 국가시험 항목분석에 항목반응이론의 적용가능성. *보건의료교육평가*, 2(1), 23-29.

지은림, 채선희. (2000). *측정 Rasch 모형의 이론과 실제*. 서울: 교육과학사.

한국보건의료인국가시험원. (2006). *한국보건의료인국가*

시험원 연구보고서 총론집

Alphen, A. V., Halfens, R., Hasman, A., & Imbos, P. (1994). Likert or Rasch? Nothing is more applicable than good theory. *Journal of Advanced Nursing*, 20, 196-201.

Antonucci, G., Aprile, T., & Paolucci, S. (2002). Rasch Analysis of the Rivermead Mobility Index: A Study Using Mobility Measures of First-Stroke Inpatients. *Archives of Physical Medicine and Rehabilitation*, 83, 1442-1449.

Baker, J. G, Granger, F. & Roger. A. (1997). A Brief Outpatient Functional Assessment Measure: Validity Using Rasch Measures. *Archives of Physical Medicine and Rehabilitation*, 76, 8-13.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: fundamental measurement in the human sciences*. New Jersey: Lawrence Erlbaum associates, publishers Mahwah.

Bernspang, B., & Fisher, A. G. (1995). Differences Between Persons With Right or Left Cerebral Vascular Accident on the Assessment of Motor and Process Skills. *Archives of Physical Medicine and Rehabilitation*, 76, 1144-1151.

Cella, D., & Chang, C. H. (2000). A Discussion of Item Theory and Its Applications in Health Status Assessment. *Medical Care*, 38(9), 66-72.

Chang, W. C., & Chan, C. (1995). Rasch Analysis for Outcomes Measures: Some Methodological Considerations. *Archive of Physical Medicine and Rehabilitation*, 76, 934-939.

Chen, C. C., Bode, R. K., Granger, C. V., & Heinemann, A. W. (2005). Psychometric Properties and Developmental Differences in Children's ADL Item Hierarchy. *American Journal of Physical Medicine and Rehabilitation*, 84(9), 671-679.

Coster, W. J., Haley, S. M., Ludlow, L. H., Andres, Bond, T. L. Y., & Ni, P. S. (2004). Refining the Conceptual Basis for Rehabilitation Outcome Measurement Personal Care and Instrumental Activities Domain. *Medical Care*, 42(1), 62-72.

- Coster, W. J., Haley, S. M., Ludlow, L. H., Andres, P. L., & Ni, P. S. (2004). Development of an applied cognition scale to measure rehabilitation outcomes. *Archives of Physical Medicine and Rehabilitation*, 85, 2030-2035.
- Doctor, J. N., Wolfson, A. M., Mcknight, P., & Burns, S. P. (2003). The effect of an accurate FIM ratings on prospective payment: A study of clinician expertise and FIM rating difficulty as contributing to inaccuracy. *Archives of Physical Medicine and Rehabilitation*, 84, 46-50.
- Downing, S. M. (2003). Item response theory: Applications of modern test theory in medical education. *The metric of Medical Education*, 37, 739-745.
- Duncan, P. W., Bode, R. K., Lai, S. M., & Perera, S. (2003). Rasch Analysis of a new stroke-specific Outcome Scale: The stroke impact scale. *Archives of Physical Medicine and Rehabilitation*, 84, 950-963.
- Duran, L. J., & Fisher, A. G. (1996). Male and female performance on the Assessment of Motor and Process Skills. *Archives of Physical Medicine and Rehabilitation*, 77, 1019-1024.
- Fisher, A. G., Liu, Y., Velozo, C. A., & Pan, A. W. (1992). Cross-cultural assessment of process skills. *American Journal of Occupational Therapy*, 46(10), 876-885.
- Fisher, A. G. (1993). The assessment of IADL motor skills: An application of many faceted rasch analysis. *American Journal of Occupational Therapy*, 47(4), 319-329.
- Fisher, A. G. (2003). *Assessment of Motor and Process Skills: Volume 1- development, standardization, and administration manual* (5th ed.). Colorado: Three star press, Inc.
- Fujiwara, T., Liu, M., Tsuji, T., Sonoda, S., Mizuno, K., Akaboshi, K. (2004). Development of a new measure to assess trunk impairment after stroke. *American Journal of Physical Medicine and Rehabilitation*, 83(9), 681-688.
- Granger, C. V., Deutsch, A., & Linn, R. (1998). Rasch analysis of the Functional Independence Measure(FIM) mastery test. *Archives of Physical Medicine and Rehabilitation*, 79, 52-57.
- Handa, V. L., & Massof, R. W. (2004). Measuring the severity of stress urinary incontinence Using the incontinence impact questionnaire. *Neurourology and Urodynamics*, 23, 27-32.
- Hart, D. L., & Wright, B. D. (2002). Development of an index of physical functional health status in rehabilitation. *Archives of Physical Medicine and Rehabilitation*, 83, 655-664.
- Hsueh, I. P., Wang, W. C., Sheu, C. F., & Hsieh, C. L. (2004). Rasch analysis of combining two indices to assess comprehensive ADL function in stroke patients. *Stroke*, 35, 721-726.
- Kjellberg, A., Haglund, L., Forsyth, K., & Kielhofner, G. (2003). The measurement properties of the Swedish version of the assessment of communication and interaction skill. *Scandinavian Journal of Caring Science*, 17, 271-277.
- Kornetti, D. L., Fritz, S. L., Chiu, Y-P, Light, K. E., & Velozo, C. A. (2004). Rating scale analysis of the Berg Balance scale. *Archives of Physical Medicine and Rehabilitation*, 85, 1128-1135.
- Kucukdeveci, A. A., Kutlay, S., Elhan, A. H., & Tennant, A. (2005). Preliminary study to evaluate the validity of the mini-mental state examination in a normal population in Turkey. *International Journal of Rehabilitation Research*, 28, 77-79.
- MacKnight, C., & Rockwood, K. (2000). Rasch analysis of the hierarchical assessment of balance and mobility(HABAM). *Journal of Clinical Epidemiology*, 53, 1242-2347.
- Malec, J. F. (2004). The Mayo-Portland Participation Index: A brief and psychometrically sound measure of brain injury outcome. *Archives of Physical Medicine and Rehabilitation*, 85, 1989-1996.
- McHorney, C. A., Haley, S. M., & Ware, J. E. (1997).

- Evaluation of the MOS SF-36 Physical Functioning Scale(PF-10):II. Comparison of relative precision using likert and Rasch scoring methods. *Journal of Clinical Epidemiology*, 50, 451-461.
- Nilsson, A. L., Sunnerhagen, K. S., & Grimby, G. (2005). Scoring alternatives for FIM in neurological disorders applying Rasch analysis. *Archives Neurology Scandinavia*, 111, 264-273.
- Piquero, A. R., MacIntosh, R., & Hickman, M. (2001). Applying Rasch modeling to the validity of a control balance scale. *Journal of Criminal Justice*, 29, 493-505.
- Prieto, L., Novick, D., Sacristan, J. A., Edgell, E. T., & Alonso, J. (2003). A Rasch model analysis to test the cross-cultural validity of the EuroQoL-5D in the schizophrenia outpatient health outcomes study. *Acta Psychiatrica Scandinavica*, 107(416), 24-29.
- Tennant, A., McKenna, S. P., & Hagell, P. (2004). Application of Rasch analysis in the development and application of Quality of Life Instruments. *Value in Health*, 7, S22-S26.
- Tennant, A., Penta, M., Tesio, L., Grimby, G., Thonnard, J. L., Slade, A., et al. (2004). Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model the PRO-ESOR Project. *Medical Care*, 42(1), 37-48.
- Tusji, T., Liu, M., Toikawa, H., Hanayama, K., Sonoda, S., & Chino, N. (1999). ADL Structure for non-disabled Japanese children based on the Functional Independence for Children. *Archives of Physical Medicine and Rehabilitation*, 78, 208-212.
- Veloza, C. A., Magalhaes, L. C., Pan, A. W., & Leiter, P. (1995). Functional scale discrimination at admission and discharge: Rasch analysis of the level of rehabilitation scale-III. *Archives of Physical Medicine and Rehabilitation*, 76, 705-712.
- Veloza, C. A., & Peterson, E. W. (2001). Developing meaningful fear of falling measures for community dwelling elderly. *American Journal of Physical Medicine and Rehabilitation*, 80(9), 662- 673.
- White, S. J., & Veloza, C. A. (2002). The use of Rasch measurement to improve the Oswestry classification scheme. *Archives of Physical Medicine and Rehabilitation*, 83, 822-831.
- Wolfe, F., Michaud, K., Kahler, K., & Omar, M. (2004). The Short Arthritis Assessment Scale: A brief assessment questionnaire for rapid evaluation of arthritis severity in research and clinical practice. *The Journal of Rheumatology*, 31(12), 2472-2479.
- Wolfe, F., Michaud, K., & Pincus, T. (2004). Development and validation of the health assessment questionnaire II. *Arthritis & Rheumatism*, 50(10), 3296-33.
- Wright, B. D., & Linacre, J. M. (1992). Combining and splitting categories. *Rasch Measurement Transactions*, 6(3), 233-235.

Abstract

A Review of the Application Rasch Analysis for Rehabilitation Outcome Measure

Lee, Jae-Shin*, Ph.D., O.T., Park, So-Yeon, Ph.D., P.T.**

*Dept. of Occupational Therapy. Konyang University

**Seiyon Child Development Center

This paper reviews recent advances that have been made in the field of theory, application of Rasch analysis. Most outcomes measures in the health care are based on an ordinal scale rather than an interval or ratio scales on the score. It's limitation is that sample-dependent, in that all statistics are confounded with the particular sample of patients who completed the assessment.

Rasch analysis has been increasingly advocated and used in health care especially rehabilitation, In this analysis, the ordinal data are converted to interval scales expressed as logit. It was used to examine the overall quality of the existing instruments, and develop the new instruments. To determine if the items formed a unidimensional measure. Also, it's model implies a consistent relationship between a person's ability and item difficulty and the probability of giving a certain answer to a test item. The scale is reliable and sensitive to change during hospital rehabilitation; the scale is also a valid measure of the patient status both before and after rehabilitation programs. The Rasch analysis based on IRT provides powerful new tools to occupational therapy in rehabilitation.

Key Words : Evaluation Instrument., Rasch Analysis, Rehabilitation Outcome