

단축형 심리검사 개발의 측정학적 방법과 타당화: 한국형 역기능우울척도를 중심으로*

서 동 기 이 순 목 김 종 남[†] 최 승 원
한림대학교 성균관대학교 서울여자대학교 덕성여자대학교

채 정 민 정 선 호 조 성 경 김 명 기 Ebesutani Chad
서울사이버대학교 경희대학교 한림대학교 덕성여자대학교

본 연구에서는 집단 연구용으로 개발된 한국형 역기능우울(dysfunctional depression)척도 20문항을 이용하여 단축형 검사 5문항을 선별하고 타당화하였다. 한국형 역기능우울척도의 단축형은 연구용 20문항 중 일부 문항(5문항)을 선별하여 제작된 것으로, 연구용 척도와 단축형 척도 모두 자기보고식 척도이다. 역기능우울 척도는 역기능우울 상태를 측정하기 위한 것으로 일상우울의 상태가 심화되었거나 경우에 따라 일상우울에서는 없는 증상들을 측정하는 도구이다. 연구용 척도는 역기능우울에 대한 학술연구에서 집단수준에서의 추론을 위해 개발되었고, 단축형 척도는 주로 임상 및 상담 현장에서 개인의 역기능 우울 수준에 대한 신속한 진단을 위해 개발되었다. 즉 연구용 척도는 역기능우울에 대한 이론구축 및 관계구조 검증이 목적이 되고, 단축형 척도는 현장에서 개인의 역기능우울 상태를 신속하게 점검하는 것이 목적이다. 본 연구에서 개발한 단축형 척도의 신뢰도와 타당도는 척도의 내적 구조 분석과 기준점을 이용한 분류정확성과 분류일관성 지수에 의해 검정이 되었다.

주요어 : 역기능우울척도, 단축형, 검사 개발, 타당화, 문항반응이론

* 이 논문은 2014년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2014S1A5A2A03064999). 본 연구의 일부는 2017년 한국임상심리학회 춘계학술대회 심포지움에서 구술발표되었음.

[†] 교신저자: 김종남, 서울여자대학교 교육심리학과, (01799) 서울특별시 노원구 화랑로 621
Tel: 02-970-5569, E-mail: kimjn@swu.ac.kr

심리측정학(psychometric)은 인간 행동에 관한 연구로 검사도구(instrument)를 통해 인간의 능력이나 정서, 성격, 태도, 행동 등을 측정하는 학문이다. 이러한 척도(검사)는 개인의 잠재적인 점수를 부여하기 위해 개발되지만 전통적으로 검사의 개발과 타당도 평가는 검사도구 자체의 이론적 구조(구성개념 타당도)의 검증에만 관심을 가져왔다. 즉 대부분의 검사도구의 개발과 타당도에 관한 연구는 검사도구의 이론적 구조에 대한 검증과 다른 준거와의 관계를 보는 준거관련 타당도를 통해 검사점수의 타당성을 간접적으로 보였다. 하지만 검사의 사용현장에서 검사의 목적은 개인에 대한 평가와 의사결정(임상/상담 장면 검사들, 각종 자격시험, 입사시험, 입학시험)이므로 추정된 개인의 점수가 타당한지에 대한 검증은 직접적으로 확인을 하여야 한다. 특히 임상/상담 장면에서는 개인의 상태를 평가하고 결과에 따라 다음 단계의 도움을 제공하는 것이 중요하므로 심리검사로 인해 산출된 점수는 타당하고 정확해야 한다.

심리검사가 개발단계 혹은 연구용이 목적인 경우에는 측정하고자 하는 구성개념과 이론구축 및 관계 구조에 대한 검증을 우선시해야 한다. 즉, 검사개발의 예비단계에서는 측정하고자 하는 구성개념의 내적 구조 및 관련 이론변수들과의 관계를 파악하여 타당화하는 것이 연구의 주된 관심사항이다. 따라서 기존의 검사개발과 타당화에 관한 연구들(김홍석, 최이순, 장요강, 2013; 박정훈, 김율리, 황순택, 2016; 엄진섭, 박지은, 전하정, 정이내, 손진훈, 2017; 이주연, 유성은, 2017)은 척도 내 요인분석과 다른 척도와의 상관을 통해 타당화를 실시하였다. 예를 들어, 본 연구에서 이용한 한국형 역기능우울척도(김종남 외, 2018)는 역기

능우울이라는 새로운 구성개념을 이론화하기 위해 요인분석과 관련 검사와의 상관 분석을 주로 이용하였다. 더 나아가 역기능우울을 일상우울(이순목 외, 2018)과의 단일 차원 상에서 문항의 속성수준(심각도 또는 난이도)을 비교하여 차별화하였다. 개념적으로, 역기능우울은 유전적 취약성과 어린 시절 심리적 외상 경험으로 인하여 강한 우울 성향을 지니게 되고 스트레스 사건에 노출됨으로써(경우에 따라 뚜렷한 스트레스 사건 없이도), 반추와 역기능적 태도와 신념, 비관적 사고와 우울 유발적 귀인 등을 나타냄으로써 심각한 우울 증상과 부적 결과를 초래하는 것으로 정의하였다(김종남, 이순목, 채정민 외, 2015; 이순목 외, 2017). 반면에, 일상우울은 울적함 등 비교적 경미한 수준의 우울을 의미하는 일상 수준의 우울이라고 정의 하였다(이순목 외, 2017; 이순목 외, 2018; 최승원 외, 2018). 이러한 일상우울과 역기능우울의 개념에 대한 논의가 아직 진행 중에 있지만 측정학적으로 우울 수준이란 단일차원 내에서 낮은 수준은 일상우울, 높은 수준을 역기능우울로 구분할 수 있다. 이에 따라 준거집단의 우울속성의 수준이 다르기 때문에 일상우울과 역기능우울을 측정하는 검사도 문항의 속성수준이 낮은 문항들과 높은 문항들로 다르게 개발해야 한다(이순목 외, 2017; 김종남 외, 2018). 결과적으로, 검사를 구성할 때도 준거집단의 속성수준(평균 검사점수)과 검사의 속성수준(평균 검사 심각도 또는 난이도)을 비슷하게 구성하는 것이 검사의 효율성과 정확성 측면에서 중요하다(Algina & Crocker, 1986).

하지만 이러한 연구용 척도는 피험자의 데이터를 신속하게 평가하고 시행하기에는 많은 문항을 포함하고 있어 현장에서 사용하기에

어려움이 있다. 특히 검사 사용의 현장에서 응답이나 내담자를 반복해서 측정하고자 할 때, 연구용 척도는 많은 부담이 된다. 따라서 많은 검사들이 개발된 이후 현장에서 간단하게 사용하기 위해 단축형 척도로 재개발되고 있다. 이러한 단축형 척도는 시간과 비용을 줄일 수 있다는 장점이 있기 때문에 최근 검사장면에서 많이 개발되고 있으며 외국의 경우 수천 개의 논문이 단축형 검사개발에 관한 내용으로 출판되고 있다(예, Butler et al., 2007; Treanor & Donnelly, 2015; Rammsted & John, 2007 등). 대표적으로 Rammsted와 John(2007)는 Big Five 성격검사(John, Donahue, & Kentle, 1991)를 10개의 문항으로 단축한 검사를 개발하였고 GAD-7(Generalized Anxiety Disorder 7-items; Spitzer et al., 2006)은 단축형 검사로 불안증세의 심각성을 신속하게 평가하기 위해 7개의 문항으로 개발되었다. 특히, MMPI(Minnesota Multiphasic Personality Inventory; Hathaway & McKinley, 1943)는 경험적으로 제작되었기 때문에 많은 문항을 포함하고 있어, 심리측정학적으로 비슷한 정보를 제공하면서 보다 간명한 척도를 개발할 수 있는지에 관한 연구가 가장 많이 진행되고 있는 척도 중의 하나이다(예, 김중술, 1999; 정송, 이순묵, 이영선, 이민수, 2011; Kincannon, 1968; McLachlan, 1974 등).

단축형 척도 개발의 관행과 문제점

최근에는 국내에서도 기존에 개발된 검사를 현장에서 신속하게 사용하기 위해 단축형 검사를 개발하거나 단축형 검사를 번안하는 연구가 많이 진행되고 있다. 문항 수가 많아지면 인지기능이 떨어져 피검사자가 집중하여

응답하기 어려울 수 있기 때문에 특히 고령자나 정신질환을 가진 피험자를 위해서 단축형 검사가 개발되고 있다. 엄진섭 외(2017)는 76개의 문항으로 된 분노반응검사(Evans & Stangelang, 1971)를 10개의 문항으로 축소하여 노인용 단축형 분노검사를 개발하였다. 엄진섭 외(2017)는 개발한 단축형 분노검사의 신뢰도와 타당도를 평가하기 위해 내적 일관성 지수(Cronbach $\alpha = .79$)와 구성개념 타당도를 검증하였다. 구성개념 타당도의 두 측면 중 하나인 수렴타당도를 평가하기 위해 상태-특성 분노 표현척도 점수와의 상관과 다른 측면인 변별타당도를 평가하기 위해 적대감 점수와의 상관을 분석하였다. 김홍석 외(2013)는 52문항의 사회적 문제 해결 검사(SPSI-R, D'Zurilla, Nezu, & Olivares, 2002)에 기반하여 25문항으로 구성된 단축형 척도를 개발하였다. 그들은 단축형 척도의 타당화를 위해 확인적 요인분석을 사용하였고 신뢰도로서 내적 일관성 지수(Cronbach $\alpha = .65$)를 보고하였다. 박정훈 외(2016)는 성격신념질문지(PBQ; Beck & Beck, 1991)의 한국어판과 성격신념질문지의 단축형(PBQ-SF; Butler, 2007)을 토대로 문항들을 추출하여 한국어판 단축형 성격신념질문지를 개발하고 타당화하였다. 이주연 외(2017)는 다차원적 경험회피 척도(MEAQ; Gámez et al., 2011)를 타당화하고 요인분석을 통해 각 6개 요인별로 4문항씩을 선별하여 24문항의 한국판 다차원적 경험회피 척도 단축형(K-MEAQ-24)을 개발하였다. 강연욱, 김향희, 나덕렬(2000)는 주의집중능력이 제한되어 있고 오랜 시간 동안 검사를 계속하기가 어려운 치매환자들에게 한국판 보스톤 이름대기 검사(K-BNT) 60문항을 모두 실시하는 것이 어렵다고 보고 짧은 시간에 검사가 가능한 15문항으로 구성된 단

측형을 제작하였다.

기존의 여러 연구들(예, 김홍석 외, 2013; 박정훈 외, 2016; 엄집섭 외, 2017)에서 제작한 단축형 검사는 개인점수에 대한 정보를 제공하는 것이 목적임에도 요인분석을 통한 내적 구조 검정과 원검사와의 상관분석을 통한 수렴타당도만을 보고하였고 개별 문항에 대한 분석은 소홀하였다. 연구용으로 개발된 비교적 충분한 문항 수를 가진 심리검사의 단축형 척도를 개발하기 위해서는 구조적으로 연구용 척도와 비슷한 구조를 보일 뿐만 아니라 피험자의 속성수준을 정확하게 측정하기 위해 피험자의 속성수준을 잘 변별하고 비슷한 수준의 난이도 문항들을 선별하여 검사를 구성해야 한다. 그러나 기존의 연구들은 단축형 검사의 문항을 선별하기 위한 통계적인 방법을 전혀 설명하지 않거나 요인분석을 통해 요인계수가 높은 문항만을 선별하여 구성하였다. 요인계수는 하나의 구성개념 관점에서 상하의 속성 수준에 있는 집단 구분이 되는 경우 달라질 수도 있다. 즉 속성수준이 중간인 집단을 준거로 하여 요인계수가 정확하고 속성수준이 낮거나 높은 사람에게서는 정확하지 않을 수 있다. 따라서 기존의 단축형 검사 개발은 검사의 요인구조에 중심을 둔 문항선별인데 요인계수는 문항이 가지는 변별도의 정보를 제공하지만 각 문항이 요인의 척도 상에서 어느 정도의 수준 또는 위치에 있는지의 정보는 포함하지 않는다. 요인분석의 주된 목적은 문항들이 엮어내는 개념구조의 파악에 있고 그 개념의 척도 상에서 문항이 어디에 위치하는지는 알려주지 않는다. 즉, 많은 문항들이 그 개념의 척도 상에 유사한 위치에 몰려있어서 일부만 남기고 나머지는 제외해도 될 경우에 제외하기 위한 필요한 정보를 제공하지 않는

다. 그러나 일반적으로 사용되는 문항반응이론에서는 주어진 개념구조에서 문항 자체의 기능을 파악하는데 중점을 두기 때문에 변별도는 물론 개별 문항들이 주어진 개념의 척도 상에서 어디에 위치하는 지에 대한 정보(속성 수준 또는 문항 난이도로 부르며, 질환이 측정대상일 경우 심각도라고 한다)인 속성 수준을 제공한다. 척도제작의 실제에서, 요인분석을 통해 어떤 요인을 측정하는데 변별도가 동일하게 높은 문항들이 확보되고 나면, 문항반응이론에 의한 분석은 척도 상에서 각 문항들이 어느 위치에 몰려 있는 지를 파악가능하게 해주며, 그들 중에서 그 위치에서 정보가 가장 높은 것 몇 개만을 선정하는 것이 가능해진다. 또한 해당 요인의 척도 상, 전 범위에 걸쳐서 문항들이 고르게 위치하도록 문항 수효를 조정하는데도 큰 도움이 된다. 따라서 이러한 피험자의 속성수준과 측정시간의 비선형적인 관계를 가정한 문항반응이론(item response theory)을 이용한 문항선별 방법이 검사개발에 많이 적용되고 있다(예, Fraley, Waller, Brennan, 2000). 검사 중심인 요인분석과 문항 중심인 문항반응이론은 수학적으로 어느 한 쪽의 추정치가 다른 쪽의 추정치와 1차식의 관계를 가지는 의미에서 선형적으로 동일한 방법(Takane, 1987)이다. 그러나 요인분석은 요인계수를 주로 이용하여 검사 구조의 타당성을 높이는 것이 주요 목적이고 문항반응이론은 문항이 가지는 속성수준(심각도/역치)을 이용하여 피험자의 속성수준을 측정하는 것이 목적이다. 단축형 검사개발의 목적과 문항반응이론의 사용목적이 일치하기 때문에 국내에서도 검사의 문항선별을 위해 문항반응이론을 적용하고 있다(정송 외, 2011; 조용래, 서동기, 2017).

정송 외(2011)는 완전-정보 문항기반 요인분석(full-information item factor analysis)을 이용하여 상호 관계 및 검사의 요인구조를 파악하고, 더 나아가 문항반응이론적 접근법을 이용하여 단축형 검사를 제작하는 방법을 제안하였다. 그러나 정송 등(2011)의 연구에서도 추정방법으로 보상적 다차원 문항반응이론(compensatory multidimensional item response theory)적 접근을 이용한 TESTFACT(Bock et al., 2003)를 사용하였으나 실제로는 요인계수만을 이용하여 단축형을 구성하였다. 이와 달리 강연욱, 김향희, 나덕렬(2000) 등은 원척도의 문항 속성수준(심각도)을 기준으로 하여 4개의 단축형 검사들이 동등한 검사의 속성수준을 가지도록 단축형 검사를 개발하였다. 최근에는 문항반응이론을 이용한 컴퓨터 개별적응 검사(CAT; computerized adaptive testing)가 단축형 검사의 대안으로 개발되고 있는데, CAT의 근본적인 알고리즘 또한 피험자의 속성수준에 가장 가까운 문항 속성수준을 보이는 문항을 할당하여 가장 효율적인 검사를 구성하는 것이다(Weiss, 1985). 특히 CAT은 준거집단 개개인의 속성수준(conditional trait level)을 측정하는 문항을 제시하기 때문에 문항의 많은 노출(item exposure)을 피할 수 있고 개인마다 맞춤형으로 동등한 검사를 제공한다(Stocking & Lweis, 1998). 주로 자격증 검사나 능력검사에서 문항수를 줄이기 위해 CAT이 적용되었으나 최근에는 심리검사에서도 CAT의 적용이 늘고 있다(예, Walter, Becker, Bjorner, Fliege, Klapp, & Rose, 2007).

단축형 척도에서 표준과 기준점수

개인의 점수를 부여하고 진단하기 위한 검

사의 개발에서는 검사구조에 대한 타당한 증거를 제시하는 물론 개인점수의 분포에 근거하여 표준표도 제시해야 한다. 표준표에는 표준이 보다 상세하게 제시되고 원점수에 대한 변환을 통해 상대비교 정보가 포함된 지수(예: 표준점수, T점수, 학년점수, 연령점수 등)들이 제공된다. 만약 검사의 목적이 현장에서 실질적으로 개인의 점수를 측정하는 것이라면 표준표는 반드시 제시되어야 한다(김종남 외, 2017). 그런데 심리검사에서 표준표의 제시는 상대평가 목적에서 필수이지만 절대평가 목적이 강한 검사에서는 표준이나 표준표를 제시할 필요가 없다. 상대평가가 목적인 규준참조 검사(norm-referenced)검사는 표본에서 피험자간의 비교로 피험자의 상대적 위치에 중점을 두는 반면 절대평가 가 목적인 준거(영역)참조(criterion-referenced or domain-referenced)검사는 표본에 관계없이 피험자가 어떠한 영역에 포함되느냐 되지 않느냐의 내용에 중점을 둔다(이순목, 김청택 외, 2001; 이순목, 반재천 외, 2007). 따라서 임상이나 상담 장면에서 사용하는 검사는 개인별 상대비교 보다는 문제나 질병의 잣대 위에 개인을 위치시키는 절대평가의 목적이 강하기 때문에 영역참조검사라고 할 수 있다. 따라서 국내의 임상이나 상담 장면에서도 영역참조적 방식으로 검사를 개발해 오고 있고 이러한 영역참조검사에서는 표준이 필요 없다(예, 김교현, 권선중, 김세진, 이순목, 2011; 김종남, 이홍표, 이순목, 2011; 이순목, 김종남, 2009; 이순목, 김교현 외, 2011; 이형초, 최윤경 외, 2007). 본 연구에서 단축형으로 개발하고자 하는 역기능우울척도 또한 피험자들의 개인차를 보는 것이 아니라 역기능우울 영역에 피험자가 포함되는지를 평가하기 위한 이분적 모형이다. 따라서 역기능우

울을 선별하는 것이 주요한 목적인 역기능 우울검사는 절대평가이기 때문에 규준은 의미가 없다.

특히 단축형 검사의 경우는 문항의 수가 작기 때문에 원점수의 분포가 좁아서 상대평가에 필요한 규준을 작성하기조차 어렵다. 따라서 절대평가가 목적인 단축형 검사는 규준(norm)이 아닌 기준점수(cut score)를 이용하여 피험자를 선별하는 것이 바람직하다. 절대평가에서는 평가대상을 등급화하거나 범주화한다. 따라서 건강/질병, 합격/불합격, 상/중/하, 관리대상/전문수준의 도움 필요/치료기관 수용 등과 같이 측정 영역 내 절대척도 상의 범주에 개인을 분류 또는 위치시키는 것이 목적이기 때문에 상대평가를 위한 정보보다 범주를 구분하는 기준 점수에 미달인지 해당되는지의 정보만이 필요하다(김종남 외, 2015).

대부분의 경우 단축형 척도는 산업/조직/임상/상담 현장에서 개인의 상태를 절대평가하기 위한 척도이기 때문에 기준점수에 의해서 개인이 어느 범주에 속하는지를 판단하여야 한다. 그러나 기존의 연구들(예, 김홍석 외, 2013; 박정훈 외, 2016; 엄집섭 외, 2017)은 절대평가를 목적으로 하는 검사들임에도 불구하고 기준점수가 제시되지 않는 경우가 많았다. 특히 단축형 검사에서도 기준점수를 제공하는 경우가 거의 없을 뿐더러 기준점을 결정하는 방법들이 자세하게 소개되어 있지 않고 있다. 따라서 본 연구에서는 한국형 역기능 우울검사(김종남 외, 2018)를 이용하여 단축형 검사를 개발하는 과정을 보이고 단축형 검사의 기준점수를 결정하고 검정하는 방법을 소개하였다.

연구 목적

한국형 역기능우울척도(김종남 외, 2018)는 집단에 대한 추론으로, 우울과 관련된 여러 개념을 측정하는 척도들을 연구대상인 모집단에 실시하여 우울과 연관된 개념들 간의 관계에 대한 이론 구축 및 가설검정에 사용될 수 있게 개발되었다. 본 연구에서는 역기능 우울의 이론구조를 밝힌 연구용 척도(20문항) 중 일부 문항을 선별하여 문항반응이론을 이용하여 검사정보(test information)의 큰 손실이 없는 범위에서 신속하게 개인점수를 산출할 수 있는 단축형 척도를 구성하고자 하였다. 연구용 척도는 역기능우울에 대한 학술연구에서 수집된 표본에 대한 모집단에서의 현상에 대하여 추론할 목적으로 개발되었고, 단축형 척도는 주로 임상 및 상담 현장에서 개인의 현재 역기능우울 수준에 대한 신속한 파악을 목적으로 개발되었다고 볼 수 있다.

본 연구는 문항반응이론의 변별모수를 요인 분석의 요인계수와 같이 사용하여 문항을 선정하는데 사용할 수 있음을 보였고 더 나아가 문항반응이론의 심각도(난이도) 모수를 이용하여 측정하고자 하는 준거 집단을 정확하게 측정하는 단축형 검사개발의 방법을 소개하고자 하였다. 특히 전문가 집단의 평가와 ROC 곡선을 이용하여 기준점을 선정하는 방법을 소개하였고 분류정확성과 분류일관성을 통해 개발된 단축형 검사의 타당성을 검정할 수 있음을 보였다.

결과적으로 본 연구는 한국형 역기능우울척도를 이용하여 1) 검사정보의 손실이 적도록 문항반응이론을 활용하여 피험자의 역기능우울을 측정하는 문항으로 구성된 단축형 척도 개발방법을 소개하였고, 2) 단축형 척도를 타

당화하고 준거(영역)참조검사의 논리에 준해서 기준점 선정 방법을 설명하고, 3) 기준점 선정의 타당화 방법인 분류일관성과 분류정확성을 소개하였다.

방 법

연구대상

본 연구에서는 김종남 외(2018)의 연구에서 사용한 자료를 활용하여 단축형 척도를 구성하였다. 역기능우울 검사를 위한 연구용 척도는 서울소재 대학병원 정신건강의학과에서 우울장애로 진단받고 치료 중인 환자 214명과 상담센터에서 우울문제를 주호소로 상담 중인 내담자 186명에게 실시하였다. 표 1은 수집된 자료의 인구통계학적 특성을 보여준다. 총 365개의 자료를 수집하였고, 그 중 연령이 대상 범주를 벗어난 3명의 자료와 연령에 응답하지 않은 2개의 자료를 완전 제거(list-wise deletion)하였다. 결국 360개의 자료로 분석을 실시하였으며 각 변수의 부분 결측은 다른 값으로 대체(imputation)하는 것보다는 모든 응답 자료를 사용해서 모형에 기반하여 추정하는(model-based estimation) 방식(Little & Rubin, 2002)으로 진행하였다. 표본의 크기가 충분하지 않다고 보고 확인적 요인분석을 위해서는 Mplus(Muthén & Muthén, 2011)를 이용하여 평균-분산을 조정된 가중최소제곱추정(WLSMV) 방식을 이용하였다.

한국형 역기능우울 연구용 척도

한국형 역기능우울척도(김종남 외, 2018)의

20개 문항을 집단 연구용으로 가정하고 이 척도에 기반하여, 최소한의 문항으로 그러나 높은 효율성을 확보하는 단축형 척도를 개발하고자 하였다. 한국형 역기능우울 척도는 김종남 외(2018)가 우울에 대한 조기 개입이 필요한 준임상 수준을 넘어서는 우울을 측정하기 위해 개발하였다. 리커트 5점 척도(0점 '전혀 아니다', 1점 '별로 아니다', 2점 '약간 그렇다', 3점 '많이 그렇다', 4점 '매우 그렇다')로 점수가 높으면 우울점수가 높다는 의미로 해석하였다. 한국형 역기능 우울척도의 내적 일관성 신뢰도는 .96으로 충분히 높은 것으로 나타났다.

임상/상담 현장에서 내담자들이 전문가를 한번이 아니라 반복적으로 방문하게 되는 경우 처음 방문시의 검사 점수와 추후 방문시의 검사 점수에 변화가 있거나 아니면 초기 점수가 그대로 유지되거나 할 것이다. 내담자의 상황에 주목할 만한 변화가 없다면 검사 점수는 유지가 되겠지만 자연 상태에서 무한정 그렇지는 않고 일정기간 동안에만 유지가 될 것이다. 그 기간이 어느 정도인지가 바로 검사의 안정성에 대한 정보이다. 이 정보를 수집하기 위해서 재검사를 실시한다. 예로서, 자연상태의 정상인 인구 집단에서 0.8 이상의 재검사 신뢰도가 유지되는 기간이 30일이라면, 그 동안에는 특별한 상황 변화가 없는 한 내담자의 검사 점수는 유사할 것으로 간주할 수 있다. 그러나 일단 30일이 지나면 검사의 안정성 범위를 벗어나므로 그 시점에서 다시 검사를 실시하여 기록을 해두어야 할 것이다. 안정성 범위를 벗어나는 동안에 발생하는 자연 상태에서의 변화를 반영한 점수가 필요하기 때문이다. 따라서 검사-재검사 신뢰도는 재검사 기간과 신뢰도 수준을 같이 보고해야 한

표 1. 수집된 자료의 인구통계학적 특성

문항	구분	빈도	퍼센트	전체(결측)
성별	남자	112	31.1	360(0)
	여자	248	68.9	
나이	19세	11	3.1	360(0)
	20대	120	33.3	
	30대	79	21.9	
	40대	77	21.4	
	50대	58	16.1	
	60대(60~64)	15	4.2	
결혼상태	미혼	180	50.0	360(4)
	기혼	137	38.1	
	이혼	31	8.6	
	별거	4	1.1	
	사별	4	1.1	
거주지역	서울/경기	319	88.6	360(1)
	충청	25	6.9	
	강원	1	0.3	
	호남	1	0.3	
	영남	13	3.6	
직업	학생	55	15.3	360(25)
	공무원	18	5.0	
	회사원	70	19.4	
	교육자	17	4.7	
	자영업	26	7.2	
	전업주부	81	22.5	
	무직	59	16.4	
	은퇴	9	2.5	
최종학력	전문직	51	14.0	360(2)
	무학	1	0.3	
	초등학교	8	2.2	
	중학교	18	5.0	
	고등학교	136	37.8	
	대학교	156	43.3	
가족소득	대학원 이상	39	10.8	360(18)
	100만원 미만	39	10.8	
	100~199만원	42	11.7	
	200~299만원	68	18.9	
	300~399만원	58	16.1	
	400~499만원	48	13.3	
	500~999만원	60	16.7	
1000만원 이상	27	7.5		
거주형태	월세 및 전세	172	47.8	360(18)
	자가 소유	170	47.2	
SES	보통이상 상	10	2.8	360(17)
	보통이상 하	15	4.2	
	보통 상	95	26.4	
	보통 하	99	27.5	
	보통이하 상	65	18.1	
	보통이하 하	59	16.4	
우울 진단 여부	예	131	36.4	360(0)
	아니오	229	63.6	

다. 본 연구에서 본검사와 검사개발을 위해서는 전체 360명 자료를 사용해서 분석하였고 이후 임상전문가의 인터뷰를 진행할 66명을 이용하여 검사 재검사를 실시하였다. 현실적으로 360명을 모두 전문가와 인터뷰하기는 어려워 66명에 대해서만 전문가 인터뷰와 검사-재검사를 실시하였다. 본 검사 시행 후 재검사는 평균적으로 32일($SD=13$ 일) 이후에 실시되었다. 재검사에 참여한 66명의 검사-재검사 신뢰도는 0.84로 높은 안정성을 보였다. 이는 연구용으로 사용하기에는 충분히 높은 신뢰도임을 보여준다. 김종남 외(2018)는 역기능우울척도 20문항이 연구용 척도로 사용하기에 타당함을 논문에서 자세히 설명하였다.

본 연구에서는 우울 수준을 3개로 구분하였다. 즉 일상우울(normal depression), 준임상우울(subclinical depression)과 임상우울(clinical depression)로 구분하였다. 또한 본 연구에서는 준임상우울과 임상우울을 하나로 통합하여 역기능우울이라고 하였다(김종남 외, 2018). 기준점 설정의 타당성 검정을 위해 우울수준의 개념적 구분을 숙지한 임상심리전문가 3인이 재검사에 응답한 66명의 내담자들을 면담하여 임상심리전문가의 판단에 의해 우울수준을 분류하였다.

면담을 위해 임상심리 전공 교수 2명이 간이 국제 신경정신학적 면담(유상우 외, 2006)과 DSM-5 진단기준(American Psychiatric Association, 2013), 역기능우울과 일상우울의 특성집합과 문항 내용(김종남, 이순목, 채정민 외, 2015)을 고려하여 면담도구를 구성하였다(부록 1). 일상우울과 준임상우울, 임상우울의 수준을 구분하기 위해서 총 12문항이 채택되었다. 3명의 임상심리전문가가 재검사 및 면담에 동의한 대상자 66명과 평균 20분 내외로

개별면담을 진행하였다. 피면담자는 병원에서 우울장애로 진단받고 치료 중인 환자나 우울을 주호소 문제로 상담센터에서 상담 중인 내담자였으며, 면담자는 개발된 면담도구를 바탕으로 기본적으로 동일한 질문을 하였고, 전문가의 판단에 의해 우울 수준을 판정하였다. 면담을 실시한 결과 일상우울(수준1)에 13명, 준임상우울(수준2)에 25명, 임상우울(수준3)에 21명이 분포되었으며, 면담 당시 우울하지 않음에 해당되는 사람은 2명이었다. 또한 면담 당시 양극성 장애로 판단되는 2명과 수준을 특정하기 곤란한 3명이 있었다.

다분문항반응이론(Polytomous Item Response Theory)

한국형 역기능우울검사는 일상우울과 준임상우울 그리고 준임상우울과 임상우울 집단을 정확하게 선별하는 것이 목적이다. 따라서 일상우울과 준임상우울 그리고 준임상우울과 임상우울 집단의 경계선에 있는 문항으로 검사를 구성하면 보다 정확한 검사(분류정확성이 높은 검사)를 구성할 수 있다(Embreston & Reise, 2000). 문항반응이론은 사람의 속성수준과 문항의 속성수준이 같은 척도 위에 있다고 가정하기 때문에 내용은 같으나 문항의 구성이 다른 두 검사를 시행한 피험자의 점수를 직접 비교할 수 있다. 일반적으로 정답이 있는 인지검사에서는 이분문항반응이론(dichotomous item response theory)을 이용하여 해당 문항에 정답일 확률만을 문항특성곡선(item characteristic curve)으로 표현한다. 이에 반해 다분문항반응이론(polytomous item response theory)은 이분문항반응을 여러 개의 범주를 가진 평정 척도의 경우로 확장한 것이다. 본 연

구는 5점 척도를 이용한 역기능 우울 척도를 이용하였기 때문에 다분문항반응이론을 이용하여 단축형 척도의 문항을 선별하였다. 일반적으로 임상검사에서 다분문항반응이론을 구현하는 모형으로 등급반응모형(graded response model; Samejima, 1969)이 적용되어 왔는데 본 연구에서는 문항 변별모수와 심각도 모수를 모두 추정하는 일반화부분점수모형(generalized partial credit model)을 적용하였다. 일반화부분점수모형은 평정 척도(예: 2점 척도, 3점 척도, 5점 척도 등) 문항의 각 범주(예: 3점 척도면 3개 범주)에서 적확하게 측정되는 속성수준을 제시하기 때문에 척도개발과 타당화 연구에 적합하다고 할 수 있다(Muraki, 1992). 등급반응모형은 범주간의 서열을 가정하는데 평정 척도에서 문항의 응답범주가 서열이 아닌 경우도 있다. 그러므로 범주간의 서열을 가정하지 않는 일반화부분점수모형이 응답자가 서로 다른 각 범주를 의미 있게 구분할 수 있는지 더 정확하게 검정할 수 있다. 일반화부분점수모형은 아래와 같이 표현된다(Muraki, 1992). 아래 공식은 θ_j 를 가진 사람이 $m_i = 5$ 개의 범주를 가진 i^{th} 문항에서 범주 k 를 선택할 확률을 나타낸다.

$$P_{ik}(x_i = k | \theta_j, a_i, b_i, d_{i1}, \dots, d_{i5}) = \frac{\exp \left[\sum_{k=1}^g 1.7a_i(\theta_j - b_i + d_{ik}) \right]}{\sum_{g=0}^{m_i-1} \exp \left[\sum_{k=0}^g 1.7a_i(\theta_j - b_i + d_{ik}) \right]} \quad (1)$$

공식(1)에서 θ_j 는 피험자 j 의 속성수준, a_i 는 변별모수이고, k 는 각 문항의 가능한 범주이고, m_i 은 각 문항의 최대 범주수로 5가 되

고, g 는 선택한 범주, b_i 는 문항 i 의 속성수준(난이도)이고 d_{ik} 는 문항 i 의 k 범주모수이다. 공식(1)의 모수들을 추정하기 위해 첫 번째 범주의 모수는 $0(d_{i,0} = 0)$ 으로 고정하였고 모든 범주모수의 합 또한 0으로 고정하였다 ($\sum_{k=1}^{m_i-1} d_{ik} = 0$).

일반화부분점수모형을 이용하여 문항을 분석하면 5개의 범주를 가진 각 문항은 1개의 변별모수와 고정된 범주모수 0과 4개의 범주모수를 가지게 된다. Muraki(1992)는 $b_i - d_{ik}$ 를 범주 $k-1$ 에서 범주 k 로 넘어가는 역치(threshold) 또는 경계모수(boundary parameter)라고 하였다. 모든 범주모수의 합은 0이기 때문에 실질적으로 5개의 범주인 경우 4개의 경계모수를 추정하게 된다. 따라서 각 문항의 속성수준은 4개의 경계모수의 평균으로 계산되기 때문에 추정할 필요가 없고 4개의 경계모수와 1개의 변별모수만을 추정하면 된다(Muraki, 1992). 그림 1은 일반화부분점수모형에서 변별모수가 다른 두 문항의 범주반응함수를 비교하기 위해 보였다.

일반화 부분화점수모형에서는 각 범주함수가 만나는 지점에서 아래로 수직선을 내려서 X축에서 만나는 점이 경계모수(t)가 되고 각 경계모수들의 평균이 그 문항의 속성수준(심각도= b)이 된다. 그림 1은 같은 문항속성수준 0(경계모수=-1.5, -0.5, 0.5, 1.5)을 가지고 있으나 변별모수 1.0과 0.5을 가진 문항의 범주반응함수를 보여주고 있다. 변별모수가 1.0인 범주반응함수를 나타낸 왼쪽 그래프는 0.5의 변별모수에 대한 범주반응함수인 오른쪽 그래프보다 첨도가 크다. 문항반응이론에서 변별모수는 문항정보(item information)로서 그 문항이

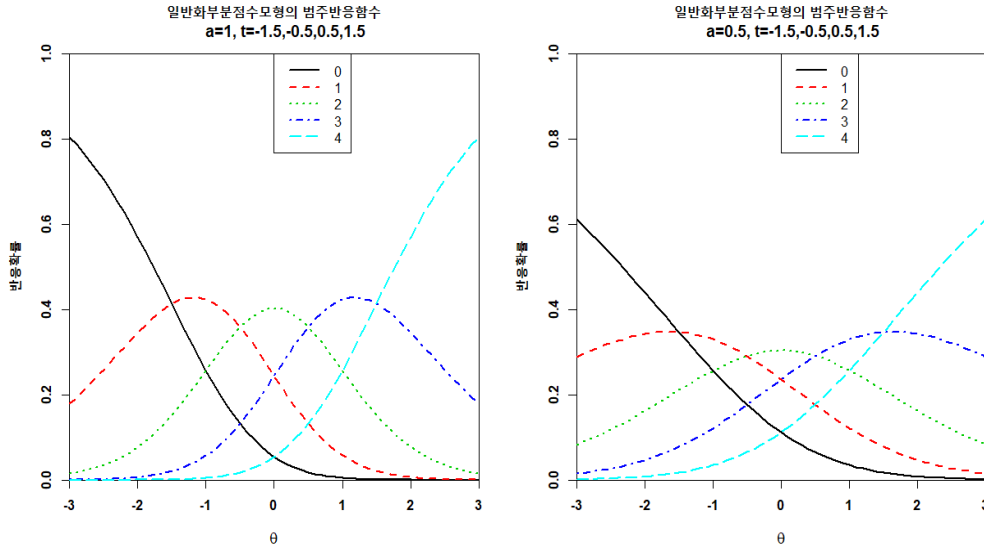


그림 1. 일반화부분점수모형에서 변별모수가 다른 두 문항의 범주반응함수 비교

측정하고자 하는 잠재변수를 정확하게 측정하는 정도라고 할 수 있기 때문에 단축형 검사에서 문항을 선별할 때, 피험자의 속성수준을 잘 측정하는 왼쪽문항을 선택해야 함을 보여준다. 일반화부분점수모형의 문항 모수를 추정하기 위해 R 프로그램(R development core team, 2008)에서 제공하는 'ltm' 패키지(Rizopoulos, 2006)를 이용하여 MML(marginal maximum likelihood) 추정방식을 이용하였고 360명의 피험자 모수를 추정하기 위해 EAP(expected a posterior) 추정방식을 이용하였다.

단축형 검사의 기준점 선정 방법

김종남 외(2018)의 연구에서 개발한 한국형 역기능우울척도 20문항을 가지고 실시한 기준점 미팅에서 단축형 척도의 기준점을 선정하였다. 1차 기준점 회의에서 역기능우울척도에 대해 일상우울, 준임상우울, 임상우울의 총 세 수준을 가정하고, 일상우울과 준임상우울이

나뉘는 지점을 기준점 1, 준임상우울과 임상우울이 나뉘는 지점을 기준점 2로 하였다. 이를 위해 총 12명의 임상심리 전문가들이 수정된 안고프방법(Modified Angoff; Angoff, 1971)에 따른 일련의 과정을 시행하였다.

이후 각 수준의 경계에서 상위 수준에 포함되는 마지막 인물(minimally acceptable person, MAP) 또는 경계선 인물 100명의 문항에 대한 응답이 어떻게 분포될 것인지를 결정하도록 하는 연습시행을 실시하였다. 그리고 단축형 척도에 선정된 문항에 대해 수준 간 경계선 인물 100명의 반응을 추정하도록 하였고 그 결과를 공유하였다. 문항정보(난이도 등), 점수 분포, 패널 간 일치도, 패널 내 일치도 등의 정보를 참조하여, 극단 값(최소/최대 결과를 보인 패널)에 대해서 토론하며 점수 간 격차를 줄였다. 3회기를 진행하여 역기능 우울검사 단축형에 대해서 2개의 기준점수를 도출하였다(김종남 외, 2018)에서와 동일한 절차를 시행함).

통상적으로 기준점 설정 시 하나의 방법으로 결정하는 것보다 타당성을 높이기 위해 두 가지 이상의 방법을 같이 사용할 것을 권장하고 있다. 따라서 본 연구는 기준점을 선정하기 위해 기준점 선정미팅과 함께 ROC 분석을 통해 기준점을 결정하였다. 임상장면에서 단축형 척도는 기준점수를 정하여 응답자들의 우울 수준이 있음/없음 수준으로 분류하는 준거(영역)참조 검사로서의 용도를 가진다. 준거 영역을 구분하는 논리와 함께 각 범주에 분류되기 위한 기준점수를 산출해야 하고, 그 기준점수의 타당성을 검증하기 위해 기준 점수에 의해서 내담자들을 분류한 결과와 그 내담자들을 임상장면 전문가들이 분류한 결과가 얼마나 일치하는지에 대한 분류일관성(classification consistency)을 조사해야 한다. 분류하는 작업은 준거 영역에서 이루어지므로 분류일관성은 곧 타당화의 증거가 된다(김종남, 이순목, 채정민 외, 2015).

표 2는 내담자를 면담한 임상가의 판단과 개발된 척도를 이용하여 진단한 결과의 관계를 표현한 것이다. 민감도(sensitivity)는 임상전문가들이 우울증이 있다고 판단한 사례를 척도점수가 우울증이 있다고 제시할 확률(true positive/(true positive+false negative))이고 특수도

(specificity)는 임상전문가들이 우울증이 없다고 판단한 사례를 척도점수가 우울증이 없다고 제시할 확률(true negative/(true negative+false positive))이다.

척도개발자는 이분적 의사결정일 때 민감도와 특수도를 최대로 하고 오류탐지와 오류통과를 최소화하는 기준점수를 구하고자 ROC 곡선(송상욱, 2009; Spackman, 1989)을 이용할 수 있다. ROC 곡선의 분석을 위해서는 ‘진실’에 대한 준거로서의 전문가 판단(임상적 면접 후)을 구하는 과정이 필요하다. 따라서 임상가의 판단과 검사의 결과를 함께 이용하기 위하여 우울 수준 결정에 참가한 66명의 검사-재검사 표본 자료를 사용하였으며 이 분석에 적절하지 않은 7명을 제외한 나머지 59명을 사용하여 ROC 곡선을 분석하였다. ROC 곡선 분석에서 곡선아래지역(Area Under Curve; AUC)의 기준은 0.5이며 이 기준보다 높은 값이 나와야 의미 있는 검사이며, 1.0일 때 가장 완벽한 검사를 의미한다(Muller, 2005). Muller 외(2005)는 논문에서 AUC 수준에 대해 0.9~1.0은 아주 높음, 0.80~0.90은 높음, 0.70~0.80은 보통, 0.6~0.70은 낮음, 0.50~0.60은 곤란으로 나타내었다. 따라서 AUC가 0.7 이상이 되어야 어느 정도 의미가 있는 검사도구라고 할 수

표 2. 임상심리전문가의 판단과 검사점수에 의한 진단에 따른 4종류의 조건부 상황

정확도(Accuracy)=민감도+특수도		우울검사(척도)	
		정상(Negative)	우울증(Positive)
임상심리 전문가	우울증(Positive)	오류통과 (False Negative: FN) =1-민감도	정확탐지 (True Positive: TP) =민감도(Sensitivity)
	정상(Negative)	정확통과 (True Negative: TN) =특수도(Specificity)	오류탐지 (False Positive: FP) =1-특수도

있다. 기준 점수 교정을 위해 민감도와 특수도를 동시에 높이는 기준점수를 만들고자 R 소프트웨어(R Development Core Team, 2008)에서 제공하는 'RPCR' 패키지(Sing, Sander, Beerenwinkel, & Lengauer, 2005)를 이용하였다. 일반적으로 ROC분석은 하나의 기준점만을 정하기 위해 사용 가능한 방법이므로 본 연구는 먼저 일상우울과 준임상우울을 기준으로 민감도와 특수도를 이용하여 분류정확성을 최대로 하는 기준점을 정하고 그 다음 준임상우울과 임상우울을 기준으로 민감도와 특수도를 이용하여 분류정확성을 최대로 하는 기준점을 정하였다.

민감도와 특수도가 최대가 되는 기준점을 정하고 그 기준점을 이용하여 분류일관성과 분류정확성을 측정하여 기준점의 타당성을 검토하였다. 분류일관성은 결정된 기준점에 근거하여 내담자가 두 번의 무선 반복된 검사에서 동일한 수준에 분류될 확률을 말한다. 즉 결정된 기준점을 가지고 두 이론적 모형에 의해 예측된 관찰 점수의 분포에 기초해서 분류표를 작성하여 동일한 수준으로 분류될 확률을 구하는 것이다. 반면에 분류 정확성은 결정된 기준점에 근거하여 추정된 관찰점수 분포를 통해 얻은 분류와 이론적 모형에 의해 예측된 진점수에 의한 분류를 비교하여 동일한 수준으로 분류될 확률이다(Lee, 2007). 분류일관성과 분류정확성은 외견상 동형검사 신뢰도라고 할 수 있다. 즉 우울의 수준을 측정하기 위하여 방식이 다른 두 개의 척도를 사용한 모습이 된다. 분류일관성은 두 검사간의 피험자 점수의 상대적 위치가 유지되는 정도로 규준지향 신뢰도라고 할 수 있고 분류일치도는 피험자의 상대적 위치를 넘어 두 검사간의 절대적 위치가 일치하는 정도로 준거지향

신뢰도라고 할 수 있다(김지윤, 여성철, 이순목, 2009; MaGraw & Wong, 1996). 그런데 신뢰도라는 개념은 규준참조검사간의 관계에서 일관성, 준거참조검사간의 관계에서 일치도로 사용되는 해당 검사들간의 관계에 사용되는 용어이고, 해당 검사와 다른 준거와의(예, 진리라고 가정하는 검사, 임상전문가의 판단)관계를 보는 검사의 신뢰도(분류정확성)는 타당도라고 할 수 있다(공시타당도 또는 예측타당도). 단축형 검사의 다분문항에 대한 분류일관성과 분류정확성은 MULT-CLASS 소프트웨어(Lee, 2007, 2008)에서 산출하였다.

결 과

단축형 척도 개발

다분문항반응이론의 일반화부분점수모형을 이용하여 연구용 척도 20개 문항의 문항모수(변별모수와 속성수준)를 추정하였다(표 3). 단축형 척도 구성을 위한 문항선정 첫 번째 기준은 일상우울집단과 준임상우울집단을 나누는 피험자의 속성수준과 가장 가까운 문항속성수준을 보이는 문항을 선정하고 준임상우울집단과 임상우울집단을 나누는 피험자의 속성수준에 가장 가까운 문항 속성수준을 보이는 문항을 선택하였다. 단축형 척도 구성을 위한 문항선정 두 번째 기준은 기준점수를 기준으로 변별모수가 높은 문항을 선정하는 것이다. 문항정보함수(item information function)는 그 문항이 얼마나 측정하고자 하는 속성수준을 정확하게 측정하는지에 대한 정도를 제공하는 지수인데 변별모수는 문항정보함수와 상관이 있으므로 변별모수가 높으면 문항정보함수의

표 3. 20개 문항으로 구성된 역기능우울 연구용 척도의 각 문항의 변별도, 심각도 및 4개의 경계모수

문항번호*	문항	변별도	난이도	경계 모수1	경계 모수2	경계 모수3	경계 모수4
basic4	밤에 잠을 깊게 못 잔다.	0.700	-0.158	-0.883	-0.231	0.876	-0.395
basic7	머리가 아프거나 어지럽다.	1.043	0.198	-0.741	-0.532	0.995	1.069
basic14	팔다리가 무겁다.	1.128	-0.148	-1.267	-0.600	0.465	0.809
basic19	친구들과 만나고 싶지 않다.	0.799	0.364	-0.900	0.183	1.119	1.054
basic24	나도 모르게 눈물이 쏟아진다.	1.198	0.237	-0.512	-0.065	0.655	0.870
basic25	전에 비해 식욕이 떨어졌다.	0.459	0.808	-0.391	0.197	1.668	1.758
basic29	아무런 의욕이 없다.	2.060	0.066	-0.988	-0.327	0.488	1.092
basic31	별일이 없어도 기분이 처진다.	2.203	-0.134	-1.167	-0.763	0.372	1.022
basic32	쉽게 자포자기하는 심정이 된다.	1.539	0.177	-1.217	-0.242	0.712	1.454
basic38	평상시와 달리 기억력이 떨어진다.	0.930	-0.150	-1.359	-0.988	0.456	1.290
basic39	눅을 잃고 멍하니 앉아 있다.	1.327	0.279	-0.850	-0.148	0.823	1.290
basic45	전보다 잘 웃지 않는다.	1.272	-0.024	-1.239	-0.471	0.546	1.069
basic48	요즘 따라 행동이 둔해졌다.	1.572	0.047	-1.198	-0.463	0.606	1.242
basic51	대인관계가 어렵다.	0.862	0.111	-1.220	-0.009	0.600	1.075
basic56	현재 상황에 대해 죄책감을 느낀다.	1.044	0.338	-0.751	0.003	0.635	1.465
basic60	보통 때보다 성욕이 저하된다.	0.491	0.037	-1.324	0.093	1.252	0.127
basic63	우울한 기분에서 회복이 잘 안 된다.	2.837	-0.009	-1.145	-0.352	0.344	1.119
basic64	세상에 내 편은 하나도 없다는 생각이 든다.	1.447	0.397	-0.562	0.172	0.986	0.993
basic65	나 스스로에 대해 비참한 생각 (기분이 든다.	1.542	-0.080	-1.088	-0.607	0.582	0.794
basic70	무슨 일이든 자신이 없다.	1.537	0.100	-1.133	-0.306	0.461	1.377

주. *문항번호는 예비검사에 사용된 최초 문항번호로 김중남 외(2017)의 연구에서 가져왔음. 음영 처리된 열은 최종 선정된 5문항을 표시함.

척도가 높아지므로 문항정보함수를 비교하여 기준점에서 척도가 높은 문항을 선정하면 된다. 그림 2는 연구용 척도 20문항의 추정된 문항 모수를 이용하여 문항정보함수(왼쪽 그림)와 360명의 피험자 속성수준의 분포(오른쪽

그림)를 나타내었다. 문항반응이론은 모수추정을 위해 집단의 속성이 표준정규분포를 따른다는 가정 하에 추정하기 때문에 그림 2의 360명의 피험자 속성수준의 분포는 표준정규분포를 따른다. 본 연구에 사용된 피험자가

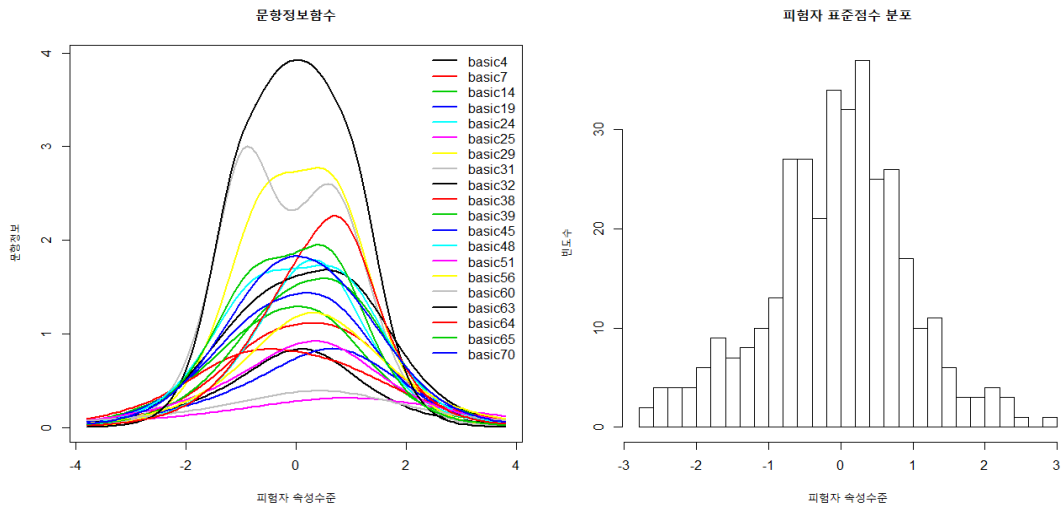


그림 2. 역기능우울 연구용 척도를 이용한 20개 문항의 정보함수와 360명의 속성수준 분포

우울장애로 진단받고 치료 중인 환자와 상담 센터에서 우울문제를 주호소로 치료 중인 내담자가 대부분이기 때문에 문항의 속성수준 또한 높지가 않다. 만약 전체 모집단을 일반인으로 두고 문항속성수준을 추정하였다면 역기능우울 연구에 사용된 문항속성수준은 일상우울 연구에서 사용된 문항속성수준에 비해 상당히 높았을 것이다(이순목, 김중남, 채정민, 2018). 본 연구는 임상집단에 준거를 두고 단축형 척도 5문항을 선별하기 위해 표 3과 그림 2의 정보를 이용하여 기준점에 가까운 부분에서 문항을 선택하였기 때문에 기준점 또한 상대적으로 높지가 않았다. 임상전문가들의 경험적 판단에 의해 일상우울집단과 준임상우울집단의 경계 속성수준은 우울집단에서 상위 50~55% 정도가 된다고 보고 상위 50~55% 피험자의 속성수준 0.0~0.125(표준점수)에 위치하는 문항의 속성수준을 보이고, 변별모수가 1.00 이상인 3개 문항(basic48, basic70, basic29)을 선택하였다. 준임상우울집단과 임상우울집단으로 분류되는 피험자의 속성수준은

상위 60~70% 정도가 된다고 보고 속성수준 0.25~0.52(표준점수) 사이에 있는 문항의 속성수준을 가지고 변별모수가 1.00 이상인 6개(basic7, basic19, basic24, basic39, basic56, basic64) 문항 중 최종 임상전문가들의 경험적 판단을 고려하여 2개의 문항(basic7, basic24)을 선택하였다. 이때 문항내용 상 신체, 정서, 동기, 행동, 인지 등 우울의 다양한 요소가 포함되도록 하였다. 결과적으로 단축형 척도를 5개(basic7, basic24, basic29, basic48, basic70)의 문항들로 구성하였다. 최종 결정된 5문항으로 구성된 역기능우울의 단축형 척도는 부록 2에 제시하였다.

표 3에서 음영 처리된 최종적으로 선정된 단축형 5개 문항에 대한 문항 변별모수와 문항속성 수준을 볼 수 있다. 20개의 문항으로 구성된 연구용 척도를 준거로 하여, 단축형 척도 5개 문항이 가지고 있는 상관은 0.93으로 충분히 높은 공시타당도(concurrent validity)를 보였다.

단축형 척도 5문항의 검사정보가 연구용척

도 20문항과 비슷하여 우울특성을 측정하는데 5문항으로 충분하다고 할 수 있다. 그림 2에서 피험자의 속성수준 범위가 -3에서 3이라고 가정할 때 각 20개의 문항정보 값을 합한 검사정보는 99.22이었다. 반면에 단축형 척도 5개 문항을 이용하였을 경우 검사정보는 28.75를 보였다. 검사정보는 각 문항정보를 더하여 계산하므로 단축형보다 연구용의 검사정보가 높을 수밖에 없다. 그러나 평균 문항정보(검사정보/문항의 수)는 연구용 척도의 경우 4.96이고 단축형 척도의 경우 5.75로 높음을 알 수 있다. 측정학적으로 검사의 효율성(test efficiency)은 검사에 각 문항이 포함됨으로써 증가되는 정보의 양으로 결정이 되는데 가장 효율성이 높은 검사는 가장 적은 수의 문항으로 피험자의 잠재특성을 정확하게(높은 정보를 획득) 측정하는 검사이다(Lunz, Bergstrom, & Wright, 1992). 연구용 척도를 이용하면 문항 수가 많아져 더 정확하게 피험자의 점수를 추정할 수 있으나 대체로 문항정보가 낮은 문항들을 검사에 많이 포함시킴으로서 임상/상담 현장에서의 신속한 선별용으로서는 검사의 효율성이 낮아진다. 반면에 단축형을 사용할 경우 전체 검사정보는 낮으나, 대체적으로 문항정보가 높은 문항들로 검사를 구성하였기 때문에 검사의 효율성은 높아진다. 즉 단축형 검사의 정확성은 연구용 척도에 비해 낮으나 그에 비해 효율성을 많이 높인 검사라고 할 수 있다. 만약 문항의 정보를 고려하지 않고 무선적으로 5문항을 선별할 경우 본 연구에서 선별한 단축형 검사의 검사정보보다 낮게 된다. 예를 들어 임의적으로 5문항(문항 1, 2, 3, 4, 5)를 선별하여 검사정보를 계산하면 문항정보가 높은 문항으로 구성된 단축형 검사의 검사정보(28.75)보다 훨씬 낮은 19.47이 된다.

단축형 척도의 타당화

현장에서 전문가가 내담자의 역기능우울 정도를 선별할 경우 5개 문항으로 구성된 단축형 척도를 사용할 수 있는가에 대한 신뢰도와 타당도 검정을 실시하였다. 연구용 척도에 비해 단축형 척도는 개인 추론용이므로 높은 신뢰도(.80 이상)를 요구한다. 단축형 척도 5개 문항은 연구용 척도의 20개 문항 속에 랜덤하게 포함되어 검사-재검사가 실시되었다. 1차 본검사 집단에서 단축형 척도의 내적 일관성 지수는 .83이었고 재검사 집단에서 단축형 척도의 내적 일관성 지수는 .82이었다. 평균 30일의 간격을 두고 실시된 단축형 검사의 검사-재검사 신뢰도는 .79이었다. 검사-재검사 신뢰도와 각 집단에서 구한 내적일관성 지수를 감안하면 진점수 수준에서도 신뢰도는 만족할만한 수준이다. 본검사 자료(N=360)에서 5개 문항 간 기술통계 결과는 표 4와 같다.

표 4를 보면 평균은 대체로 1점(별로 아니다)과 2점(약간 그렇다) 사이에 있고 표준편차는 1.00이상으로 평균과 표준편차 모두 20문항인 연구용 척도의 자료와 유사하다. 단축형 척도에 사용된 문항들은 왜도 첨도가 절대값 2보다 작으므로 모두 정규분포를 따른다고 볼 수 있다. 다분상관계수는 .50에서 최고 .76 사이에 있다. 연구용 척도(김종남 외, 2018)에서는 2개의 특수요인을 가지는 준쌍요인 모형(bifactor-like model)을 최종모형으로 선정하였으나, 단축형 척도에서는 5개의 문항으로 구성되었기 때문에 1요인구조 외에는 더 많은 요인을 가지기는 어렵다. 특히 쌍요인은 모형은 단축형 점수 산정 시 일반요인만 해석하여 문항총점으로 개인 점수계산이 가능하므로(Reise, Morizot, & Hays, 2007) 1요인 구조로 모형검

표 4. 역기능우울척도 단축형 문항의 기술통계

일련 번호	문항	M	SD	왜도	첨도	상관계수				
						문항1	문항2	문항3	문항4	문항5
1	머리가 아프거나 어지럽다	1.77	1.28	0.17	-.96	1.00	-	-	-	-
2	나도 모르게 눈물이 쏟아진다	1.69	1.39	0.29	-1.18	.61*	1.00	-	-	-
3	아무런 의욕이 없다	1.77	1.19	0.09	-1.05	.59*	.64*	1.00	-	-
4	요즘 따라 행동이 둔해졌다	1.67	1.26	0.06	-.89	.61*	.62*	.73*	1.00	-
5	무슨 일이든 자신이 없다	1.66	1.29	0.07	-1.00	.50*	.62*	.76*	.69*	1.00

주. M과 SD는 5점 척도인 원자료에서 구한 것이고 상관은 범주 변수에 대한 요인분석을 위해 구한 다분상관(polychoric correlation)임. * $p < .05$

정을 하였다. 연구에서 사용된 표본에 대해 연구용 척도도 1요인 구조로 나타나는지 확인적 요인분석을 실시한 결과 전반적 합치도가 무난하게 나왔다($\chi^2_{(190)}=15413.626(p=.00)$, RMSEA=.091(.084~.098), CFI=.967, TLI=.963, WRMR=1.210). 이에 따라 단축형 검사의 요인 구조를 탐색하지 않고 직접 내적 구조의 타당도를 검증하기 위해 1요인의 확인적 요인분석을 한 결과는 표 5와 같다.

표 5에서 5개 문항은 1요인으로 잘 수렴하고 있고 전반적 합치도 역시 무난하였다($\chi^2_{(5)}$

표 5. 단축형 척도의 1요인 모형에 대한 확인적 요인분석 결과(표준화 계수, N=360)

문항 일련번호	요인계수
1	.70*
2	.76*
3	.88*
4	.83*
5	.83*

주. 측정오차간 고정된 상관에 대한 수정지수(MI)가 10을 넘는 경우는 없음. * $p < .05$

=37.618($p=.00$), RMSEA=.135(.096~.176), CFI=.989, TLI=.977, SRMR=.033). 그리고 문항간 잔차 상관이 낮아 수정지수가 10을 넘는 경우가 없었다. 이것으로 역기능우울 단축형 척도는 1요인 구조이고 특별히 문항들 간 방법효과를 공유하는 경우가 없는 척도로 검정이 되었다. 즉, 연구용 척도뿐만 아니라 단축형 척도의 내적 구조도 타당하다고 할 수 있다.

기준점 선정 미팅에서 단축형 척도 5문항을 이용한 일상우울과 준임상우울의 기준점은 11점이었고, 준임상우울과 임상우울의 기준점은 14점이었다.

그림 3은 ROC 분석을 통해 일상우울과 준임상우울을 나누는 방법을 보여준다. 일상우울과 준임상우울의 AUC는 0.82로 높았고(그림 3의 왼쪽) 민감도와 특수도를 최대화하는 기준점은 민감도와 특수도가 만나는 지점인 8점에서 결정되었다(그림 3의 오른쪽).

그림 4는 ROC 분석을 통해 준임상우울과 임상우울을 나누는 절차를 보여준다. 준임상우울과 임상우울의 AUC는 0.95로 검사의 정확도 수준은 아주 높은 수준이었고(그림 4의 왼

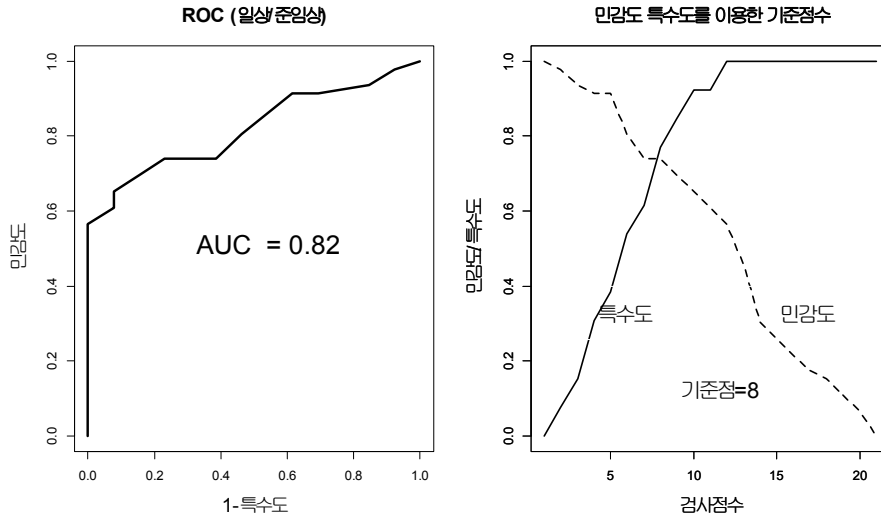


그림 3. 일상우울과 준임상우울의 기준점을 찾기 위한 ROC분석 결과

쪽) 민감도와 특수도를 최대화하는 기준점은 특수도와 민감도가 만나는 점인 14점에서 결정되었다(그림 4의 오른쪽).

기준점 설정에서 기준점 회의를 통해 얻은 결과(기준점수 11점과 14점)와 ROC 곡선 분석을 통해 얻은 결과(기준점수 8점과 14점)는 일

상우울과 준임상우울 간 기준점수에서 달랐다. 준임상우울과 임상우울 간 기준점수는 전문가 합의와 경험적 자료에 의한 점수가 일치하여 그대로 14점으로 하였다. 그러나 일상우울과 준임상우울 간 기준점수에 대해서는 추가 논의가 필요하였다. 즉, 역기능우울척도 단축형

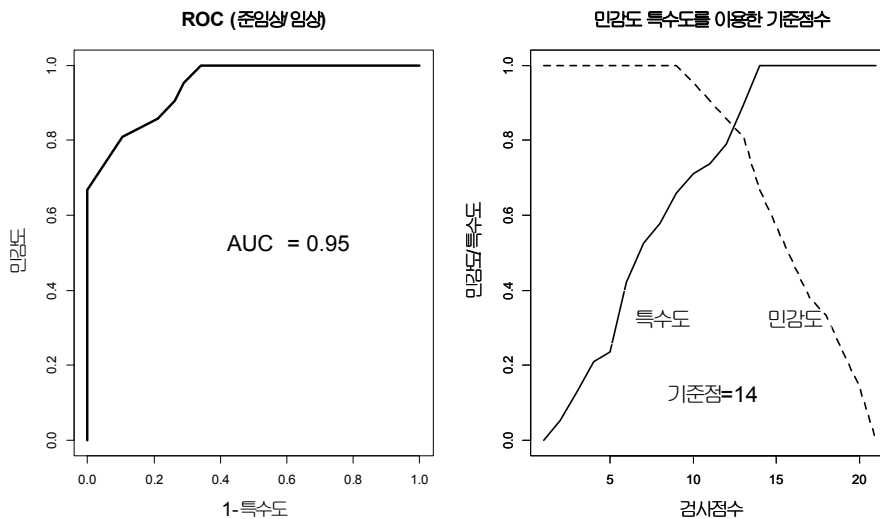


그림 4. 준임상우울과 임상우울의 기준점을 찾기 위한 ROC 분석 결과

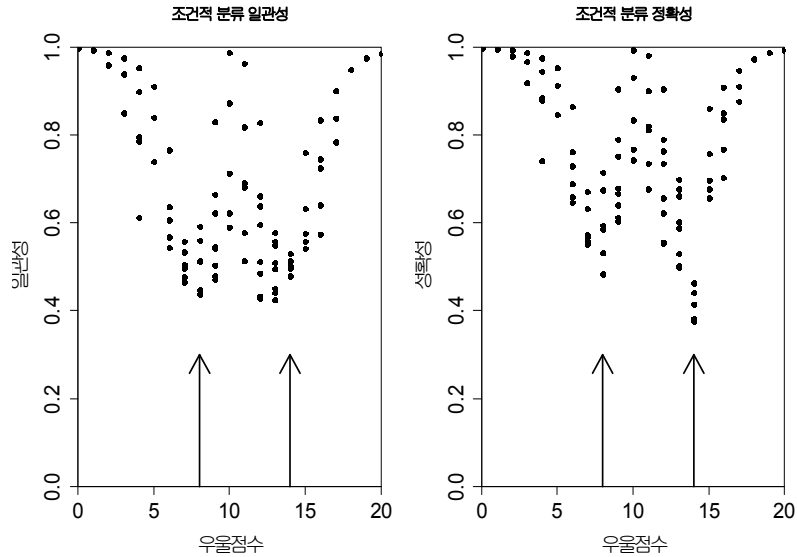


그림 5. 5문항으로 구성된 역기능우울척도 단축형의 분류일관성과 분류정확성

에서는 특수도보다 민감도를 높이는 것이 중요하다. 임상심리전문가들의 의견에 따라 일상우울과 준임상우울의 기준점은 8점, 준임상우울과 임상우울의 기준점은 14점으로 결정하였다. 이는 역기능우울척도 단축형은 우울을 보이지 않는 내담자를 선별하는 것보다 우울을 보이는 내담자를 선별하는 것을 중요하게 생각하는 척도임을 반영한 것이다.

결정된 기준점수 8점과 14점이 신뢰롭고 타당한지 검증하기 위한 지수로 분류일관성과 분류정확성을 살펴보았다. 임상심리 전문가가 판단한 내담자들(66명)의 분류결과와 단축형 척도의 기준점수에 기반하여 분석한 결과 분류일관성 지수는 0.75, 분류정확성 지수는 0.82로 상당히 높은 것으로 나타났다(그림 5).

그림 5는 단축형 척도 5개 문항에 대하여 일상우울과 준임상우울 간 기준점수 8점, 그리고 준임상우울과 임상우울 간 기준점수 14점이 주어진 조건에서 분류일관성과 분류정확

성 지수를 보여준다. 단축형 척도를 이용한 개인점수들이 기준점수 8점과 기준점수 14점 근처에서 가장 낮은 분류일관성과 분류정확성을 보이지만, 기준점수에서 멀어질수록 그 지수들의 값은 높아짐을 알 수 있다. 이러한 분류일관성과 분류정확성의 유형은 타당한 기준점수를 가진 검사의 전형적인 모습이다(Lee, 2008). 20개 문항으로 구성된 연구용 척도와 비교해서는 그 지수들의 값이 낮지만 단축형 5개 문항으로도 일상우울과 준임상우울 그리고 준임상우울과 임상우울을 분류하기에는 충분하다고 판단된다. 따라서 단축형 척도로 5개의 문항을 사용할 경우 기준점 8점과 14점을 이용하여 일상우울과 준임상우울, 준임상우울과 임상우울을 구분하기에 타당하다고 볼 수 있다.

논 의

단축형 검사의 목적은 최소한의 문항으로 신뢰성이 높고 타당한 측정도구를 만드는 것이다. 이러한 목적을 달성하기 위하여 본 연구에서는 다음과 같은 측정학적 특성을 고려하였다. 첫째, 검사는 검사대상자들을 잘 선별하는 것이 일차적인 목적이므로, 검사대상자들 간의 점수 차이를 최대한 반영한 기준점수 근처에 위치하는 문항들을 선별하였다. 둘째, 비슷한 문항들이 반복적으로 검사에 포함된다면 검사 효율성이 낮아지므로 검사의 효율성을 높이기 위해 문항정보가 높은 문항을 선별하였다. 셋째, 단축형 검사의 기준점수를 결정하기 위해 두 가지 방법을 사용하였으며 분류 정확성과 분류일관성을 높이는 기준점을 선정하였다. 본 연구에서 개발한 단축형 검사는 위의 세 가지 측정학적 특성을 모두 고려하여 개발된 검사라고 볼 수 있다.

본 연구를 통해 선별된 5개 문항으로 구성된 단축형 척도의 신뢰도는 양호한 수준으로 나타났다. 내용관련 타당도도 단축형 척도 문항 선정의 기반이 되는 원척도(20문항)의 제작 과정에서 상당히 자세히 제공되고 있기 때문에 검정되었다고 할 수 있다(김종남 외, 2018). 연구용 원척도(20문항)의 타당화를 위한 분석 자료에서 내적 구조가 검증되었고, 역기능우울 척도로써 연구용 척도 대신 단축형 척도를 사용하여 검사를 진행하여도 문제가 없음을 검정하였다. 뿐만 아니라 단축형 검사의 기준점수1과 기준점수2를 이용한 분류일관성과 분류정확성이 상당히 높게 나왔기 때문에 실제적으로 현장에서 사용하기에 타당하다고 할 수 있다. 결론적으로, 본 연구는 문항반응이론을 활용하여 측정학적 관점에서 단축형 검사

를 개발하였고 단축형 검사를 타당화하였다고 할 수 있다. 따라서 앞으로 기존의 우울척도들과 비교했을 때 본 연구에서 개발된 단축형(현장용) 척도가 재난 상황이나 급박한 상황에서 쉽고 빠르게 사용될 수 있을 것으로 기대한다.

특히 실용성 측면에서 본 연구에서 개발한 단축형 척도의 가치는 높게 평가될 수 있다. 검사의 양호도를 평가하는 기준으로 신뢰도, 타당도, 수용가능성, 그리고 실용성을 언급할 수 있는데(이순목, 1999), 특히 실용성은 경제성과 편의성(conveniency)으로 나누어진다. 경제적 실용성은 시간적, 금전적 측면에서 평가될 수 있다. 편의적 실용성은 실시 지침의 명료함, 사용 절차의 간단함이라고 할 수 있는데, 본 연구에서 개발된 단축형 척도는 임상/상담 현장에서 매우 중요한 시간적 경제성과 사용 절차의 간단함이라는 장점을 가지고 있으며 반복 사용이 용이하다는 강점이 있다. 그러나 단축형 척도의 해석 시 다음과 같은 유의점을 고려해야 한다(김종남, 이순목, 채정민 외, 2015; 이순목, 김종남, 채정민, 최승원, 서동기, 2018).

첫째, 김종남 외(2018)가 발표한 한국형 역기능우울척도가 이제 개발된 단계이고 역기능우울에 대한 개념 또한 논의 중(김종남, 2014; 이순목 외, 2017; 최승원 외, 2018)으로, 역기능우울 단축형 척도의 내적 구조 관련 타당도는 검정되었다고 할 수 있으나 다른 변수와의 관계 속에서 준거관련타당도는 아직 제시되지 못하였다. 이후에도 역기능우울 단축형 척도의 다양한 타당화 연구가 계속 이루어질 필요가 있다.

둘째, 단축형 척도의 ROC 분석에는 59명의 자료가 사용되었는데, 표본 수가 크지 않다는

제한점이 있다. ROC 분석을 위해 우울장애로 진단받은 환자와 우울이 주호소인 상담센터 내담자를 임상심리전문가가 직접 대면하여 자체 개발한 면담 스케줄에 따라 구조화된 면담을 진행하여 우울 수준을 분류하였는데, 이때 구조화된 면담에 참여하는데 동의한 연구 참여자의 수가 많지 않아 더 많은 표본을 구할 수가 없었다. 보다 많은 자료를 확보하였다면 더 높은 수준의 민감도와 특이도, 분류정확성과 분류일관성 결과를 얻을 수 있었을 것으로 예측된다.

셋째, 역기능우울척도 단축형 기준점수의 타당성에 대한 경험적 증거가 필요하며 전문가와 연구자들의 피드백을 통해 계속해서 검증 및 수정, 보완이 이루어질 필요가 있다. 우울연구나 임상 및 상담 현장에서 척도가 사용되는 가운데 역기능우울척도 단축형의 현재 기준점수가 정확한지, 조정이 필요한지를 판단할 수 있을 것으로 기대된다. 그리고 본 척도는 전체 모집단을 임상집단을 준거로 해서 만들었기 때문에 일반인을 대상으로 사용하기 위해서는 앞에서 언급한 문항속성수준과 문항정보함수 그리고 기준점의 재선정 등 측정학적 특성을 고려하여 단축형 검사를 새롭게 구성해야 할 것이다.

넷째, 단축형 검사를 구성할 경우 몇 개의 문항으로 단축해야 할지는 통계적인 방법만으로 결정할 수 없고 측정하고자 내용을 잘 파악할 수 있는 연구자의 판단이 필요하다. 문항반응이론을 통해 연구용 척도와 단축형 척도의 검사정보함수(test information function)를 비교해서 어느 정도 원척도의 검사정보의 손실을 감수하면서 평균 검사정보를 높일지를 결정해야 한다. 즉 연구용 척도의 정확성을 최소로 희생하면서 적은 수의 문항에서 최대

한의 효율성을 담보할 수 있는 방향으로 단축해야 한다. 이는 측정학적 측면뿐만 아니라 현장의 요구를 고려해서 신중하게 결정해야 할 것이다.

마지막으로 역기능우울척도 단축형의 활용과 관련하여 다음과 같은 유의점을 고려해야 한다(김종남 외, 2015; 이순목 외, 2018). 첫째, 역기능우울 탐지를 최대한 높이고자 할 경우에 기준점수를 낮게 함으로써 실제 역기능우울을 보이는 피험자들이 좀 더 많이 역기능우울로 선별(screening)되도록 해야 할 것이다. 그러나 이때 정확통과 비율(특수도)이 감소하고 오류탐지가 증가하는 것을 감수하는 비용이 따른다. 둘째, 정확통과를 높이고 오류탐지의 감소가 목적이라면 기준점수를 높게 함으로써 목적을 달성할 수 있다. 그러나 이때는 정확탐지 비율(민감도)이 감소하고 오류통과가 증가하는 비용을 감수해야 한다. 셋째, 어떤 목적으로, 어떤 비용을 감수할 것인가에 대한 결정은 통계적인 것이 아닌 개념적인 판단을 통해 내릴 수 있다. 사회적으로 어떤 것이 더 추구되느냐에 따라 그에 따른 비용을 감수하는 방향으로의 선별이 이루어져야 할 것이다. 즉, 정확탐지 비율을 증가시키는 이점이 오류탐지 비용보다 크다면 기준점수를 내려야 할 것이고, 역기능우울을 경험하고 있는 사람만을 실수 없이 가려내는 것이 연구의 목적이라면 오류통과는 문제가 되지 않으므로 기준점수를 높여야 할 것이다. 결론적으로 역기능우울 단축형 척도의 사용에 있어서 개발자들이 제공하는 기준점수는 철칙이 아니고 척도사용의 목적에 따라 융통성 있게 적용함으로써 검사의 가치를 살리고 질 높은 사용이 되도록 하는 것이 중요하다(김종남, 이순목, 채정민 외, 2015). 이렇게 검사의 측정학적 측면을 넘

어 기능적인 측면을 타당도의 증거로 간주하는 관점도 있기 때문에(예, Cronbach, 1988), 본 연구에서 개발한 단축형 검사의 기준점은 기능주의적 관점에 근거하여 타당도를 높이기 위해 기준점을 낮게 잡았다고 볼 수 있다.

참고문헌

- 김연옥, 김향희, 나덕렬 (2000). 한국판 보스톤 이름대기 검사(K-BNT)의 병렬 단축형 개발. *대한신경과학지*, 18(2), 144-150.
- 김교현, 권선중, 김세진, 이순목 (2011). 저수준 도박행동 연구를 위한 개념화 및 척도개발. *한국심리학회지: 일반*, 30(2), 599-628.
- 김종남 (2014). 한국형 우울척도 개발을 위한 이론적 모형 설정. 2014년 한국심리학회 연차대회 자료집(구술 발표). 8월 30일 이화여자대학교. 서울: 한국심리학회.
- 김종남, 이순목, 채정민, 최승원, 서동기, 조철현, 최윤경 (2015). 문화반영적 우울모형 개발 및 그에 기반한 선별용 우울척도의 제작과 타당화 보고서: 역기능-부적응 우울 예비문항 개발 보고서. 서울여대 산학협력단.
- 김종남, 이순목, 서동기, 최승원, 채정민 (2017). 한국형 역기능우울 검사의 사용자 교본. 서울여대 산학협력단.
- 김종남, 이순목, 최승원, 채정민, 서동기 (2017). 문화반영적 우울모형 개발 및 그에 기반한 선별용 우울척도의 제작과 타당화: 일상우울 검사 개발 보고서. 서울여대 산학협력단.
- 김종남, 이순목, 최승원, 채정민, 서동기, 최윤경, 조철현 (2017). 한국형 역기능 우울 척도의 문항 개발 및 문항 분석. *한국심리학회: 건강*, 22(4), 849-882.
- 김종남, 이순목, 최승원, 채정민, 서동기, 이현정, 원은수 (2018). 한국형 역기능우울척도의 개발 및 타당화. *스트레스研究*, 26(2), 103-114.
- 김종남, 이홍표, 이순목 (2011). 문제 도박의 조기탐지 및 대처를 위한 가족용 척도의 개발. *한국심리학회지: 일반*, 30(1), 135-168.
- 김지윤, 여성철, 이순목 (2009). 급내 상관과 일반화가능도 이론 간의 이론적 및 경험적 비교. *한국심리학회: 일반*, 28(3), 659-686.
- 김중술 (1999). *다면적 인성검사*. 서울대학교 출판부.
- 김홍석, 최이순, 장요강 (2013). 단축형 사회적 문제해결 검사의 타당화. *Korean Journal of Clinical Psychology*, 32(3), 611-625.
- 박정훈, 김울리, 황순택 (2016). 한국어판 단축형 성격신념질문지의 타당화 연구. *신경정신의학*, 55(2), 103-113.
- 송상욱 (2009). Receiver Operating Characteristic (ROC) Curve를 이용한 민감도와 특이도 측정. *Korean Journal of Family Medicine*. 30(11), 841-842.
- 유상우, 김영신, 노주선, 오강섭, 김찬형, 남궁기, 채정호, 이기철, 전성일, 민경준, 오동재, 주은정, 박현주, 최영희, 김세주 (2006). 한국판 Mini International Neuropsychiatric Interview 타당도 연구. *대한불안학회지*, 2(1), 50-55.
- 엄진섭, 박지은, 전하정, 정이내, 손진훈 (2017). 노인용 단축형 분노반응검사의 개발 및 타당화. *감성과학*, 20(1), 95-104.

- 이순목 (1999). 대학생선발용 면접의 기능과 방법에 대한 통합적 고찰. *한국심리학회지: 산업 및 조직*, 12(1), 55-76.
- 이순목, 김종남, 채정민, 최승원, 서동기 (2018). 일상우울척도의 개발 및 타당화. *Korean Journal of Clinical Psychology*, 37(3), 411-430.
- 이순목, 김종남, 최승원, 채정민, 최윤경, 서동기, 권호인 (2017). 우울연구를 위한 새로운 방향: 행동과학적 접근의 도입. *한국심리학회지: 일반*, 36, 293-323.
- 이순목, 김교현, 최용성, 이홍표, 김종남, 김수진, 권선중, 김세진, 김인혜 (2011). 한국형 도박행동 H척도의 개발: 고수준 도박자 대상의 연구용 척도. *한국심리학회지: 일반*, 30(4), 1053-1082.
- 이순목, 김종남 (2009). 도박중독 문제의 본질에 충실한 평가/진단 및 비율 산정. *한국심리학회지: 건강*, 14(1), 1-26.
- 이순목, 김청택, 김명소, 설현수, 유태용, 이도형, 임대열. (2001). 영역참조적 검사제작의 논리와 적용. *한국심리학회지: 일반*, 20(2), 305-325.
- 이순목, 반재천, 이형초, 최윤경, 이순영 (2007). 인터넷 중독 진단방식의 전환: 규준참조 평가에서 영역참조평가로. *한국심리학회지: 임상*, 26(1), 213-238.
- 이형초, 최윤경, 이순목, 반재천, 이순영 (2007). 성인 인터넷 중독 진단척도 개발 및 타당화 연구. *한국심리학회지: 임상*, 26(3), 765-791.
- 이주연, 유성은 (2017). 한국판 다차원적 경험 회피 척도 (K-MEAQ)의 타당화 및 단축형 (K-MEAQ-24) 개발. *인지행동치료*, 17(2), 181-208.
- 정 송, 이순목, 이영선, 이민수 (2011). MMPI 단축형 우울척도 개발-문항기반 요인분석의 사용. *한국심리학회지: 일반*, 30(3), 811-830.
- 조용래, 서동기 (2017). 수용과 행동 질문지 제2판(AAQ-II)에 대한 요인분석과 다분문항 이론의 적용: 지역 사회 노인표본을 대상으로. *Korean Journal of Clinical Psychology*, 36(3), 402-417.
- 최승원, 이순목, 김종남, 최윤경, 서동기, 채정민 (2018). 반응성 우울과 내인성 우울의 현대적 재개념화. *Korean Journal of Clinical Psychology*, 37(2), 211-224.
- Algina, J., & Crocker, L. (1986). *Introduction to Classical and Modern Test Theory*. New York: Holt, Rinedhart and Winston Inc.
- American Psychiatric Association. (2013). *DSM-5*. Arlington: VA. American Psychiatric Association.
- Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike(Ed.), *Educational Measurement* (pp. 508-600). Washington, DC: American Council on Education.
- Beck, A. T., & Beck, J. S. (1991). *The personality belief questionnaire*. Unpublished manuscript. Philadelphia: University of Pennsylvania.
- Bock, R. D., Gibbons, R, Schilling, S. G., Muraki, E., Wilson, D. T., & Wood, R. (2003). *TESTFACT 4.0* [Computer program and manual]. Lincolnwood, IL. Scientific Software International.
- Butler, A. C., Beck, A. T., & Cohen, L. H. (2007). The personality belief questionnaire-short form: development and preliminary

- findings. *Cognitive Therapy and Research*, 31(3), 357-370.
- Cronbach, L. J. (1988). "Five perspectives on validity argument". In H. Waine & H. I. Braun (Eds.) *Test Validity*, NJ: LEA. pp.3-17.
- Rizopoulos, D. (2006). ltm: An R package for Latent Variable Modeling and Item Response Theory Analyses. *Journal of Statistical Software*, 17(5), 1-25. URL <http://www.jstatsoft.org/v17/i05/>
- D'Zurilla, T. J., Nezu, A. M., & Maydeu-Olivares, A. (2001). *Social Problem Solving Inventory-Revised Technical Manual*. New York: Multi-Health System Inc.
- Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah: Lawrence Erlbaum Associates.
- Evans, D. R., & Stangelang, M. (1971). Development of the reaction inventory to measure anger. *Psychological Reports*, 29(2), 412-414.
- Gamez, W., Chmielewski, M., Kotov, R., Ruggero, C., & Watson, D. (2011). Development of a measure of experiential avoidance: The Multidimensional Experiential Avoidance questionnaire: development and initial validation. *Psychological Assessment*, 26, 35-45.
- Hathaway, S. R., & McKinley, J. C. (1943). *The Minnesota multiphasic personality schedule*. University of Minnesota Press, Minneapolis.
- John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory Versions 4a and 5*. Berkeley, CA: Institute of Personality and Social Research.
- Lee, W. (2007). Multinomial and compound multinomial error models for tests with complex item scoring. *Applied Psychological Measurement*, 31(4), 255-274.
- Lee, W. (2008). *MULT-CLASS: A computer program for multinomial and compound-multinomial classification consistency and accuracy* (Version 3.0). Iowa City, IA: Center for Advanced Studies in Measurement and Assessment, The University of Iowa.
- Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York, NY: John Wiley & Sons.
- Lunz, M. E., Bergstrom, B. A., & Wright, B. D. (1992). The effect of review on student ability and test efficiency for computerized adaptive tests. *Applied Psychological Measurement*, 16(1), 33-40.
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30-46.
- Muller, M. P., Tomlinson G, Marrie, T. J., Tang, P., McGeer, A., Low, D. E., Detsky, A. S., & Gold, W. L. (2005). Can routine laboratory tests discriminate between severe acute respiratory syndrome and other causes of community-acquired pneumonia? *Clinical Infectious Diseases*, 40, 1079-1086.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16, 159-176.
- Muthén, L. K., & Muthén, B. O. (2011). *Mplus User's Guide*. Sixth Edition. Los Angeles, CA: Muthén & Muthén.
- Spitzer, R. L, Kroenke K, Williams, J. B., & Löwe, B. (2006). A brief measure for assessing

- generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine*, 166, 1092-1097.
- Stocking, M. L., & Lewis, C. (1998). Controlling item exposure conditional on ability in computerized adaptive testing. *Journal of Educational and Behavioral Statistics*, 23(1), 57-75.
- R Development Core Team (2008). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, 41, 203-212.
- Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, 16(1), 19-31.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores (Psychometric Monograph No. 17). Richmond, VA: Psychometric Society. Retrieved from <http://www.psychometrika.org/journal/online/MN17.pdf>
- Sing T, Sander O, Beerenwinkel N., & Lengauer T. (2005). ROCr: visualizing classifier performance in R. *Bioinformatics* 21(20), 3940-3941.
- Spackman, K. A. (1989). "Signal detection theory: Valuable tools for evaluating inductive learning". Proceedings of the Sixth International Workshop on Machine Learning. San Mateo, CA: Morgan Kaufmann. pp. 160-163.
- Takane, Y., & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, 52, 393-408.
- Walter, O. B., Becker, J., Bjorner, J. B, Fliege, H., Klapp, B. F., & Rose, M. (2007). Development and evaluation of a computer adaptive test for 'Anxiety'(Anxiety-CAT). *Quality of Life Research*, 16, 143-155
- Weiss, D. J. (1985). Adaptive testing by computer. *Journal of Consulting and Clinical Psychology*, 53, 774-789.
- 1차원고접수 : 2018. 10. 14.
수정원고접수 : 2019. 01. 22.
최종게재결정 : 2019. 03. 11.

**Psychometric Methods and Validation
of Short Form for the Psychological Scale:
Based on the Korean dysfunctional depression scale**

Dong gi Seo Hallym University	Soonmook Lee Sungkyunkwan University	Jong-Nam Kim Seoul Women's University	Seungwon Choi Duksung Women's University
Jungmin Chae Seoul Cyber University	Sunho Jung Kyunghee University	SunKyung Cho Hallym University	MyungKi Kim

Chad Ebesutani

Duksung Women's University

The purpose of this study was to develop a short form of the Korean dysfunctional depression scale (KDDS) and introduce methods for validation of the short form. The short form was developed by selecting five items from the original 20 items in the Korean dysfunctional depression scale. While the KDDS was developed to measure dysfunctional depression for the purpose of group research, the short form was developed for the practical use at individual levels. The scale for academic study was developed to infer the dysfunctional depression at the group level. So the short form can be widely used by practitioners in industries, organizations, clinical/counselling settings for quick screening of potential patients suffering from dysfunctional depression. We demonstrated the reliability and validity of the short form by analyzing internal structure of the scale and providing classification accuracy and consistency.

Key words : dysfunctional depression scale, short form, development of scale, validation, item response theory

부 록

1. 면담용 검사 도구

* 가장 최근 우울경험을 기준으로 답변해주시기 바랍니다.

1) 일상우울-준임상우울 구분

질문			
1	한 번 우울해지면 어느 정도 지속되나요?(1주일 이상이면 준임상우울)		
2	우울해지기 직전에 어떤 스트레스 사건이 있었나요?	예	아니오
3	우울하다가도 좋은 일이 생겼을 때에 기분이 전환되는 것을 느낀 적이 있습니까?	예	아니오
4	우울한 문제와 관련하여 전문적인 도움(상담이나 심리치료, 약물치료)이 필요하다고 생각하셨나요?	예	아니오
5	요즘 집중하거나 의사결정을 내리는데 힘이 들었습니까? (요즘 자신의 능력을 잘 발휘하고 있다고 생각하시나요?)	예	아니오
6	요즘 미래에 대해 어떤 생각을 가지고 있습니까?		

2) 준임상우울-임상우울 구분(지난 2주 기준)

질문			
1	지난 2주 동안 지속적으로 우울하거나 기분이 가라앉았던 적이 있습니까?	예	아니오
2	지난 2주 동안 거의 매일 식욕이 줄거나 늘어났습니까? 또는 체중이 늘거나 줄었습니까?	예	아니오
3	지난 2주 동안 거의 매일 밤에 잠자는데 어렵습니까?	예	아니오
4	지난 2주 동안 거의 매일 평소보다 말이나 행동이 느려졌거나 또는 초조해 하거나 안절부절 못하거나 가만히 앉아있을 수 없었습니까?	예	아니오
5	지난 2주 동안 우울 증상으로 당신의 직업, 사회생활, 또는 중요한 영역에 있어 어떤 변화가 나타났습니까?	예	아니오
6	지난 2주 동안 반복적으로 자해를 생각하거나 자살하고 싶은 생각이 들었거나 죽고 싶었습니까?	예	아니오

2. 한국형 역기능우울척도-단축형

<p>다음의 문항은 일상생활에서 경험하는 기분, 생각, 또는 행동들에 대한 것입니다. 문항을 읽고 <u>오늘을 포함하여 지난 1주일 간</u> 자신에게 있었던 경험을 가장 잘 나타내는 선택지에 V표 해주 세요.</p>						
번호	문항	전혀 아니다(0)	별로 아니다(1)	약간 그렇다(2)	많이 그렇다(3)	매우 그렇다(4)
1	머리가 아프거나 어지럽다.					
2	나도 모르게 눈물이 쏟아진다.					
3	아무런 의욕이 없다.					
4	요즘 따라 행동이 둔해졌다.					
5	무슨 일이든 자신이 없다.					

채점: 척도에 대한 응답이 끝난 후, 전혀 아니다=0, 별로 아니다=1, 약간 그렇다=2, 많이 그렇다=3, 매우 그렇다=4를 부여. 문항별 점수를 합계하여 척도 점수 산출. 상세지침은 한국형 역기능우울 검사의 사용자 교본(김종남, 이순목, 서동기, 최승원, 채정민, 2017) 참조