

Development of Korean Depression Screening Assessment: Phase II Preliminary Validation Study

Seowon Yoon¹ Bitna Lee² Kiho Park¹ Sooyun Jung¹ Shin-Hyang Kim¹ Yeonsoo Park³
Seung-Yeon Baik³ Cho Long Kim³ Won-Hye Lee⁴ Younyoung Choi^{2†} Seung-Hwan Lee^{3,5†} Kee-Hong Choi^{1†}

¹Department of Psychology, Korea University, Seoul; ²Department of Adolescent Counseling, Hanyang Cyber University, Seoul; ³Clinical Emotion and Cognition Research Laboratory, Inje University, Goyang; ⁴Department of Clinical Psychology, National Center for Mental Health, Seoul; ⁵Department of Psychiatry, Ilsan Paik Hospital, College of Medicine, Inje University, Goyang, Korea

Recent mental health policies emphasize both primary and secondary prevention of mental illness. In this regard, a screening assessment tool with easy accessibility, high reliability, and validity is crucial for identifying individuals at risk or in need of optimal treatment in various health care settings or national research projects. Even though there are several screening assessment tools available for depressive disorders in Korea, few have been reported to have proper diagnostic sensitivity and specificity for Korean people. None are available for both online and offline assessment, scoring, and interpretation without the payment of an international loyalty fee. The current study aimed to report on our Phase II (second year) validation study of Korean Screening Assessment for Depressive Disorders (K-DEP) including the scale development process, validity evidence, and reliability indices. The K-DEP, a brief screening assessment tool for depressive disorders, is designed to cover major components of the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), diagnostic criteria, the Research Domain Criteria (RDoC), and other screening tools for depressive disorders that can be easily used in community mental health settings, psychiatric hospitals and clinics, or various research projects. The results support the construct validity of the K-DEP with optimal evidence based on its contents, internal structure, relation with other variables, response process, and consequences. We also reported Cronbach's α , sensitivity, specificity and standard error of measurement using Item Response Theory for the reliability index. The implications of the current study and the research project of the third year are also discussed.

Keywords: depressive disorders, screening assessment, K-DEP, validity, reliability, validation, DSM-5

우울장애는 신체적 고통과 동시에 나타나는 경우가 30-50%가량에 달하는, 정신질환 중에서는 가장 높은 유병률을 가진 질환이다 (Kroenke et al., 2011). 주요우울장애의 전 세계 유병률은 4.4%로 이는 2005년과 비교했을 때 18.4% 증가한 수치이다. 이를 인구 수로

환산하면 약 3억 2천 2백만 명에 달하는 사람들이 주요우울장애를 경험하고 있다(World Health Organization [WHO], 2018). 국내의 경우 2015년 한 해 동안 우울장애를 경험한 사람은 61만 명으로 추산된다(Ministry of Health and Welfare & Samsung Medical Center, 2017). 2016년 보건복지부에서 실시한 정신질환 실태역학조사에 따르면, 주요우울장애 평생 유병률은 전체 5.0%(64세 이하: 5.1%)로 나타났다. 이처럼 주요우울장애는 여러 정신 장애 중에서도 가장 흔하게 발생하는 질병이다.

질병 부담도의 측면에서 보았을 때도, 주요우울장애는 개인과 국가의 삶에 매우 큰 영향을 미친다(American Psychiatric Association [APA], 2013). 우울장애는 일반적인 의료 장면에서 가장 흔히 나타나는 정신질환이면서도 초래하는 비용도 상당히 크다(Arroll, Khin, & Kerse, 2003). 2008년에 출간된 WHO의 보고서에 따르면,

[†]Correspondence to Kee-Hong Choi, Department of Psychology, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, Korea; E-mail: kchoi1@korea.ac.kr

[†]Correspondence to Younyoung Choi, Department of Adolescent Counseling, Hanyang Cyber University, 220 Wangsimni-ro, Seongdong-gu, Seoul, Korea; E-mail: 1120008@hycu.ac.kr

[†]Correspondence to Seung-Hwan Lee, Department of Psychiatry, Ilsan Paik Hospital, College of Medicine, Inje University, 170 Juhwa-ro, Ilsanseo-gu, Goyang, Korea; E-mail: lshps@hanmail.net

Received Apr 19, 2018; Revised May 17, 2018; Accepted May 18, 2018

This study was supported by a grant of the Korea Mental Health Technology R&D Project, Ministry for Health & Welfare Affairs, Korea (HM15C1169).

주요우울장애는 모든 신체 및 정신질환들 가운데 3위에 해당하는 부담도를 보였고, 이 지수를 소득 기준 중, 상위 국가로 한정하였을 때는 전체 1위에 해당하였다(WHO, 2008).

주요우울삽화는 특히 첫 삽화 기간 동안 초기에 치료하는 것이 예후에 중요한 영향을 주었으며, 발병과 치료 사이의 기간이 짧을 수록 예후에 긍정적인 영향을 주었다(Ghio, Gotelli, Marcenaro, Amore, & Natta, 2014). 또한 우울증을 치료하지 않고 방치할 경우 해마 회백질의 부피가 감소하는 등 신경학적 문제가 발생할 수 있다고 보고되었다(Sheline, Gado, & Kraemer, 2003). 이처럼 우울증의 치료에 있어서 조기 선별 및 개입은 점차 중요한 이슈로 주목을 받고 있으며, 미국 예방 서비스 대책 위원회(U.S. Preventive Services Task Force)에서는 일차의료기관 등에서도 우울증 선별을 통해 조기에 정확한 진단과, 효과적인 치료를 받을 수 있도록 권장하고 있다(Sharp & Lipsky, 2002). 다시 말해, 일차 의료기관을 방문하는 환자들 중 우울장애를 호소하는 환자들을 선별해내어 초기에 적절한 치료를 받도록 하는 것이 중요하며, 이를 위해서는 일차의료 장면에서 간편하면서도 높은 정확도로 환자를 구분해낼 수 있고 비전문가도 사용이 가능한 선별도구가 필요하다(Russell, 1994).

이처럼 주요우울장애 및 관련 질환의 제1차 및 2차 예방을 위해 임상장면에서의 정확한 조기진단의 중요성이 증가하였으며 더불어 일차적으로 환자를 판별하여 적절한 치료 기회를 제공할 수 있는 우울증 선별 도구의 활용이 매우 중요해지고 있다. 미국에서는 각 연령에 따라 다른 선별도구를 권고하고 있는데, 성인 및 노인의 우울증의 경우에는 Beck Depression Inventory (BDI), Beck Depression Inventory-II (BDI-II), Center for Epidemiologic Studies Depression Scale (CES-D), Geriatric Depression Scale (GDS) 등의 다양한 선별 도구를 권고하고 있다(Sharp & Lipsky, 2002).

Jung 등(2017)이 현재까지의 우울장애 선별검사들을 리뷰한 결과, The Nine-Item Patient Health Questionnaire(PHQ-9)의 경우 우울증 선별도구로서 신뢰도와 타당도를 검증한 연구가 있었지만 검사 대상자의 규모가 주요우울장애 74명, 대조군 75명으로 상대적으로 작았다(An, Seo, Lim, Shin, & Kim, 2013). 한국의 임상장면에서 흔히 사용되는 또 다른 선별도구로는 BDI-II와 CES-D가 있다. BDI-II를 사용함에 있어서는 국내 대규모 연구 R&D나 정기적이고 대규모의 임상활용을 위해서는 해외에 저작권료를 부담해야 하는 비용부담이 있으며, 사회적, 직업적, 일상적 기능의 손상수준을 묻는 문항이 없었다(Jung et al., 2017). BDI-II의 경우 민감도와 특이도가 모두 .9 이상으로 보고된 바 있으나 임상군의 샘플 사이즈가 31명에 불과했으며(Sung et al., 2008), 이중맹검으로 시행되지 않았다는 문제점이 있었다(Jung et al., 2017). 또한, BDI-II 매뉴얼에

서 제시하는 mild와 moderate의 기준 점수인 14점과 20점으로는 효과적으로 우울장애, 주요우울장애를 구분해내지 못하는 문제점이 있었으며, IRT 분석 결과 수면이나 성욕과 같은 신체증상과 관련된 몇몇 문항에서 낮은 변별도 지수와 문항들의 반응곡선이 적절치 않게 나타난 점이 지적되었다(Park, Yoon, Lee, & Choi, 2018). 이는 우울척도 특히 신체증상 문항에 반응하는 한국인의 고유한 특성에 기인하는 것일 수 있다.

또한 외국에서 개발된 우울 선별 도구의 경우 문화적 차이로 인해 이를 고려하지 않을 시 절단점이 달라지거나 단순 적용하기 어려울 수 있다는 문제점을 가지고 있다(Yoon, Lim, & Han, 2012). 앞서 소개한 척도들은 모두 외국에서 개발된 것으로, Marsella와 Yamada(2007)는 다른 문화권에서 개발된 척도를 사용하여 우울장애와 우울을 연구하기 위해서는 문화적 차이를 고려하여야 한다고 주장했다. 대표적으로, CES-D의 경우 동양 문화권에서 긍정정서를 강하게 억제하는 경향으로 인하여 긍정정서 부족과 관련된 문항의 점수가 우울요인과 관련된 문항의 점수보다 훨씬 높게 나온다는 보고가 있으며 이로 인해 우울증의 유무와 관계없이 점수가 높게 나올 수 있다(Demirchyan, Petrosyan, & Thompson, 2011). 따라서, 한국 문화에서 경험되는 우울장애의 요인과 특성은 외국과 다를 수 있으며, 외국에서 개발되고 타당화된 외국의 척도를 한국의 문화에 맞게 타당화하지 않은 채로 번역하여 사용하는 것은 한국 고유의 우울장애 특성을 포함하는 데 있어 적절하지 않을 수 있다(Kim et al., 2017).

즉, 아직까지는 여러 척도들이 단순 번역되어 타당화 과정 없이 활용되고 있거나, 진단기준의 내용을 모두 포괄하지 못하거나, 한국인의 문항에 대한 반응 특성을 고려하여 선별한 문항을 포함하지 못하였다. 또한 우울장애의 선별에 있어 진단적 민감도와 특이성이 높으며, 신뢰할만한 타당도의 근거를 충분히 제시하는 도구가 개발되어 있지 않다.

이에 본 연구에서는 내부 연구진(임상심리 전문가 3명, 정신과 전문 1명, 측정심리 전문가 1명)과 외부 정신건강 전문가들(정신과 전문 4명, 임상심리 전문가 2명, 측정심리 전문가 2명)의 체계적인 자문 내용을 통합하여, 일차 의료기관 등에서 간편하면서도 정확하게 우울장애 환자들을 선별하는 데 활용할 수 있는 한국형 우울장애 선별도구 문항을 개발하였다. 우선, 문항의 선별에 있어서는 다른 척도와의 관계, 내용 타당도 근거, 내적 요인구조, 문항특성 등을 고려하였고, 최종적으로 DSM-5, RDoC의 진단기준을 포괄하며 신뢰도와 민감도 및 특이도가 높고 타당도 근거가 확보된 문항들이 선별되었다. 다음으로 새로 개발하는 한국형 우울장애 선별 도구의 요인구조를 탐색하고, 기존의 다른 우울장애 선별도구와의

관계를 살펴봄으로써 타당도의 근거를 제시하였다. 또한 다른 도구들의 민감도와 특이도를 제시하고 비교하여 한국형 우울장애 선별도구의 실용적 효용성을 함께 논의하였다.

연구 방법

자료수집절차

본 연구는 한국형 우울장애 선별도구 개발에 대한 3년 연구로 2년 차(2016년 9월부터 2017년 8월)에 수집된 자료로 진행되었으며, 일반인 523명과 우울장애 환자군 90명으로 총 613명이 참여하였다. 피험자 선정 과정에서는 본 연구가 일반인 대상으로 우울선별도구를 개발하는 것을 목적으로 하기 때문에 무선 선정을 원칙으로 특별한 배제 기준을 두지는 않았지만 만 18세 미만의 피험자는 배제하였고($n=1$), 면담 거부($n=11$) 혹은 설문 미완성자($n=26$), 타국적($n=1$, 일본인) 피험자도 배제하였다. 피험자는 연구에 참가할 때 연구에 대한 설명을 들은 후 연구 참여 동의서를 작성하였고, 이후 설문조사와 면담 평가를 실시하였다. 이 과정에서 피험자가 설문지에 응답한 내용과 진단이 서로 오염되지 않도록 하기 위해 설문지 안내와 면담은 서로 다른 연구자가 독립적으로 진행하였다. 본 연구는 관련기관의 IRB 승인을 받아 진행하였다.

활용된 검사의 종류

본 K-DEP의 개발 및 타당화를 진행하기 위해서 기존에 활용되고 있던 우울 평가 도구인 Beck 우울검사2판(BDI-II), 역학연구센터 우울척도(CES-D), PHQ-9를 활용하였고, 진단 평가를 위해서는 구조화된 면담 도구인 Mini International Neuropsychiatric Interview-Plus Version 5.0(M.I.N.I. 5.0) 검사를 함께 실행하였다.

한국형 우울장애 선별도구(K-DEP)

한국형 우울장애 선별도구는 우울장애를 선별하기 위해 본 연구에서 개발된 척도이다. 한국형 우울장애 선별도구를 개발하기 위해 문헌은행이 구축되었으며, 높은 신뢰도와 타당도를 갖춘 10개의 문항을 최종 선별하였다. 문항을 선별하는 데 있어서 일차적으로는 기존의 우울장애 평가도구들과의 상관관계와 문항 난이도 등이 고려되었다. 추가적으로 전문가들의 이론적 논의를 거쳤으며 문항정보곡선에서 높은 정보값을 제공하고 민감도, 특이도가 높은 문항들이 선별되었다. 최종 문항들은 국립국어원에 의뢰하여 맞춤법 표기와 문장의 구성에 대한 감수를 완료하였다.

Beck 우울검사2판(BDI-II)

Beck, Steer와 Brown(2014)이 기존의 BDI를 개정한 2판을 활용하였다. BDI-II는 총 21문항으로 구성되어 있으며 4점 리커트(0-3점) 척도를 사용하고 있다(Beck et al., 2014). 국내에서도 여러 연구들에 의해(Kim, Lee, & Lee, 2007; Lim et al., 2011; Sung et al., 2008; Yu, Lee, & Lee, 2011) BDI-II의 표준화가 진행되었다. Lim 등(2011)의 연구에서 주요우울장애에 대하여 18점을 절단점(cut off score)으로 사용하였을 때, 85.0%의 민감도와 88.1%의 특이도를 보여주었다.

역학연구센터 우울척도(CES-D)

Radloff(1977)가 역학연구에서 일반인의 우울증을 간편하게 측정하기 위해 기존 타당화가 진행된 우울증 평가도구인 BDI, SDS, MMPI-D에서 문항을 선별하여 개발하고, 국내에서도 여러 연구(Cho & Kim, 1993, 1998; Chon & Rhee, 1992)에 의해서 번안 및 표준화된 CES-D를 활용하였다. 총 20문항으로 구성되어 있으며, 지난 7일간의 정서 및 신체증상, 대인관계 관련 증상을 4점 리커트 척도로 측정한다. 국내연구에서는 주요우울장애 환자군을 분류하기 위한 절단점으로 25점을 적용할 때 민감도 91.3%, 특이도 78.8%로 보고되었다(Cho & Kim, 1993).

우울증 선별도구(PHQ-9)

Kroenke, Spitzer와 Williams(2001)가 일차 의료 장면에서의 간편한 우울증 선별 및 심각도 측정을 위해 개발하고, 국내에서는 An 등(2013), Choi 등(2007), Lim 등(2009)이 각각 국내 타당화 연구를 진행한 PHQ-9를 활용하였다. PHQ-9은 이분구성으로 되어있는데, DSM-IV의 주요 우울장애 진단을 위한 9가지 증상에 대응하는 9개 문항과 이로 인한 일상생활 어려움의 심각도를 측정하는 1개 문항을 포함한다. 국내연구에서 주요우울장애에 대한 절단점은 9점을 사용할 때 민감도 88.5%, 특이도 94.7%로 보고되었다(An et al., 2013).

M.I.N.I. 5.0 검사

본 연구에서 연구 참가자의 정신병리학적특성을 진단하기 위해 Lecrubier 등(1997)이 개발하고 국내에서는 Yoo 등(2006)이 한국어판으로 표준화한 구조화된 면담도구인 간이 국제 신경 정신 인터뷰를 사용하였다. 검사절차로는 연구자가 본 연구의 목적 등을 간단히 설명하고 연구 참가 동의를 받은 뒤 적혀 있는 질문들을 읽고 대답을 받아 적는 절차로 평가가 이루어졌으며, 질문에 적혀 있는 질문 외에도 추가적인 정보(예: 기능 손상 정도) 등을 파악하

기 위해 임상가의 판단에 따라 부가적인 질문을 시행하였다. 진단 면담이 끝난 뒤 임상가는 면담 정보를 토대로 피험자들의 진단명을 정하였고, 이 과정에서 본 연구의 설문지 정보가 활용되지 않도록 하였으며 모든 진단 정보는 임상 심리 전문가의 슈퍼비전을 통해 확정되었다.

연구 과정 및 절차

한국형 우울장애 선별도구의 개발은 척도의 개발 과정 자체가 타당도의 근거를 수집하는 과정임을 강조하는 최근 학계의 동향을 반영하였다. 본 연구에서는 Jung 등(2017)이 소개한 과정을 적용하여, 모든 과정에서 근거 중심의 평가도구 설계체계의 중요성을 반영할 수 있는 근거 중심의 설계 체계(ECD)에 기초하여 검사개발과정을 설계 및 진행하였다(Jung et al., 2017). 전문가 집단을 활용하여 실시한 영역분석으로 다양한 이론에 의해서 제시된 우울의 조작적 정의와, DSM-V와 RDoC의 우울장애 진단기준에 기초하여 기존의 검사들을 분석하고, 하위구성체의 요소들을 통합하여 정리하였다. 또한 임상심리전문가, 임상 및 상담심리 전공 박사 및 석사과정생, 정신과 전문의에 의하여 Focus Group Interview(FGI)를 실시하였다. FGI를 통하여 기존의 평가도구들에서 상대적으로 비중을 두지 않았던 사회, 직업, 그 외 삶의 중요한 영역에서 나타나는 우울을 드러내는 행동을 수집하여 문항에 반영하였다. 마지막으로 임상심리전문가, 정신과 전문의, 심리측정 전문가로 구성된 전문가 자문단이 우울의 조작적 정의, 문항들이 진단기준에 부합하는지, 개별 문항들이 측정하기에 적합한지 등을 검토하였고, 우울장애 선별도구의 의미, 지향점, 문항의 앵커링, 수량, 채점방식의 적합성을 평가받았다. 분석에 관해서는 Mplus(Muthén & Muthén, 1988-2015)를 이용한 탐색적 및 확인적 요인분석을 실시하고, IRTPRO (Cai, Thissen, & Du Toit, 2011)를 이용하여 문항반응이론을 활용한 각 문항의 특성을 추정하였다. 또한 문제은행(Item Bank)을 구축함에 있어서 최종문항들을 도출하기 전 주어진 피험자로 최대한 많은 파일럿문항을 고려하기 위해서 링킹(Linking/Equating)방법으로 자료를 수집하고 분석하였다(de Ayala, 2009; Kolen & Brennan, 2014). 구축된 문제은행을 통해서 최종문항의 도출은 Test Assembly를 위한 Matlab을 활용하여 최적의 문항을 도출하였다.

연구대상의 특성

본 연구에는 M.I.N.I. 5.0 진단을 통하여 일반군 523명, 우울장애로 진단받은 환자군 90명으로 구성된 총 613명의 피험자를 대상으로 분석이 실시되었다. 남자는 194(31.6%)명, 여자는 419(68.4%)명으로 여성의 수가 상대적으로 많았으며 평균 연령은 만 36.23세(SD =

Table 1. Demographic Characteristics (N=613)

Category	Classification	Number of persons	Proportion (%)	
Age	19 years or older	23	3.8	
	20-29	275	44.9	
	30-39	90	14.7	
	40-49	120	19.6	
	50-59	58	9.5	
	60-69	24	3.9	
	70 years or older	18	2.9	
	Unanswered	5	.8	
	Mean (SD)	36.23 (14.16)		
Education	Less than 13 years	130	21.2	
	13-16	377	61.5	
	17 years or longer	90	14.7	
	Unanswered	16	2.6	
	Mean (SD)	14.8 (2.71)		
Marital state	Unmarried	224	36.5	
	Married	358	58.4	
	Others	31	5.1	
Gender	Male	194	31.6	
	Female	419	68.4	
Non-clinical group		523	85.31	
Clinical group	Major depressive disorder	40	6.5	
	Persistent depressive disorder	50	8.2	
	Bipolar I disorder	13	2.12	
	Bipolar II disorder	4	.65	
	Severity of suicide	No suicide tendency	505	82.4
		Suicide tendency low	62	1.1
Suicide tendency mid		38	6.2	
Suicide tendency high		8	1.3	
Total		613	100	

14.16)였다. 피험자의 정신병리학적 특성으로는 주요우울장애는 6.5%, 우울장애 14.7%, 불안장애는 16.0%였다(Table 1).

결 과

이전까지 사용되던 대부분의 우울장애 선별도구들이 고전검사이론에 근거하여 제작되었다는 점과 차이를 두어 문항반응이론(IRT)을 도입하여 검사 문항을 선정하였다. 총 383문항의 예비 문항에 대해서 각각의 문항의 문항특성곡선(ICC)과 문항정보곡선(IIC)을 분석하여 각 문항이 우울이라는 잠재변수를 얼마나 잘 측정하고, 각기 다른 우울 수준에서 가장 많은 정보량을 제공하는지를 파악하였다. 이 과정을 거쳐 총 10개의 소수의 최적 문항들을 판

Table 2. 10 Items Selected from Stage 2

Item number	Contents	DSM-5 category
Item 1	(Past 2 weeks) During most of the day, it was depressing.	Depressed mood
Item 2	(Past 2 weeks) I could not live happily.	Diminished interest
Item 3	(Past 2 weeks) I felt anxious or irritable	Psychomotor agitation/retardation
Item 4	(Past 2 weeks) I did not even want to speak or move	Fatigue
Item 5	(Past 2 weeks) I felt my life was worthless.	Feelings of worthlessness/guilt
Item 6	(Past 2 weeks) I had diminished concentration ability or decision-making ability than before.	Diminished ability to think or concentrate/ indecisiveness,
Item 7	(Past 2 weeks) I felt annoyed and wanted to die.	Suicidal ideation
Item 8	(Past 2 weeks) I felt like I could not get out of the trap.	Anxiety
Item 9	(Past 2 weeks) I felt that I had nothing more to look forward to.	Hopelessness
Item 10	(Past 2 weeks) It was difficult to fall asleep or often awake	Sleep disturbance

Table 3. Comparison of the Item Characteristics of the Integrated Group and the Patient Group

Item	Integrated group (n = 613)		Depressive disorder Patient group (n = 90)		Correlation between total scores	Min value	Max value
	M	SD	M	SD			
Item 1	.54	.96	1.72	1.37	.98**	0	4
Item 2	.87	.98	2.04	1.19	.90**	0	4
Item 3	.62	1.04	1.91	1.46	.98**	0	4
Item 4	.51	.93	1.70	1.35	.98**	0	4
Item 5	.44	.88	1.54	1.31	.98**	0	4
Item 6	.54	.86	1.58	1.17	.96**	0	4
Item 7	.31	.84	1.33	1.37	.93**	0	4
Item 8	.47	.92	1.61	1.33	.97**	0	4
Item 9	.43	.89	1.54	1.36	.96**	0	4
Item 10	.37	.23	.78	.14	.97**	0	1

**p < .001.

별해내어 간편하면서도 우울수준에 대한 많은 정보를 제공할 수 있도록 하였다. 구체적인 선별기준으로는 (1) 고전검사이론에 따른 문항 난이도, 문항과 총점의 상관, BDI-II, PHQ-9, CES-D와의 상관을 통해서 계산된 값을 참고하여 적합하다고 판단된 문항, (2) 이론전문가들의 이론적 논의를 통하여 적합하다고 판단된 문항, (3) 문항반응이론을 통하여 추정된 문항정보곡선(Item Information Curve)을 통해서 본 연구의 검사 목적에 부합하는 우울수준에서 높은 정보값을 제공하고 낮은 측정 오차값을 갖는 가장 최적화된 문항들을 조합, (4) 민감도와 특이도가 높은 문항 조합으로 선별하였다(Table 2).

문항의 특성

Table 2는 분석한 문항들의 내용과 문항이 포함하는 진단의 영역들이며, Table 3은 최종적으로 선별된 문항들의 특성을 나타낸다. 일반인과 우울장애에 환자군을 모두 포함한 통합군의 응답평균은

.31-.87, 환자군의 응답평균은 .78-2.04였다. 문항 총점상관은 .90 이상으로 높은 상관값을 보여주었다. Table 4는 문항반응이론에 따른 문항 난이도는 문항경계값이 겹치지 않고 어느 정도의 간격을 유지하며 문항변별도는 1.35에서 5.54로 모두 양호했다.

요인분석

한국형 우울장애 선별도구(K-DEP)의 내적구조의 타당성을 검증하기 위하여 본 연구에서는 탐색적 요인분석과 확인적 요인분석을 실시하였다. 각 분석을 위해서 수집된 자료를 무선으로 나누어 탐색적 요인분석에서는 307명(일반군: 254명, 환자군: 53명), 확인적 요인분석에서는 306명(일반군: 269명, 환자군: 37명)의 자료를 사용하였으며, 모든 분석은 Mplus(Muthén & Muthén, 1988-2015)를 활용하여 실시하였다. 탐색적 요인분석에서는 1요인 모형이 지지되었으며 확인적 요인분석 결과로도 1요인 모형의 합치도가 양호한 수준으로 관찰되었다(Table 5). Figure 1은 확인적 요인분석 결과

Table 4. Item Difficulty and Item Discrimination Based on Item Response Theory

Item	<i>a</i>	<i>SE</i>	<i>b1</i>	<i>SE</i>	<i>b2</i>	<i>SE</i>	<i>b3</i>	<i>SE</i>	<i>b4</i>	<i>SE</i>
Item 1	5.54	.71	.16	.05	.84	.07	1.34	.10	1.68	.11
Item 2	3.47	.44	-.26	.08	.59	.09	1.34	.13	1.85	.18
Item 3	3.51	.50	.11	.08	.76	.10	1.19	.14	1.73	.21
Item 4	4.18	.61	.23	.07	.95	.12	1.40	.17	1.93	.23
Item 5	4.62	.58	.37	.06	1.03	.09	1.48	.11	2.10	.17
Item 6	3.48	.46	.24	.07	.72	.13	1.87	.10	2.31	.24
Item 7	4.47	.82	.73	.09	1.07	.11	1.50	.16	2.79	.94
Item 8	3.90	.51	.57	.09	1.16	.11	1.57	.14	1.90	.32
Item 9	3.72	.46	.39	.07	1.03	.10	1.46	.12	1.96	.17
Item 10	1.35	.28	-1.55	.23	1.15	.22				

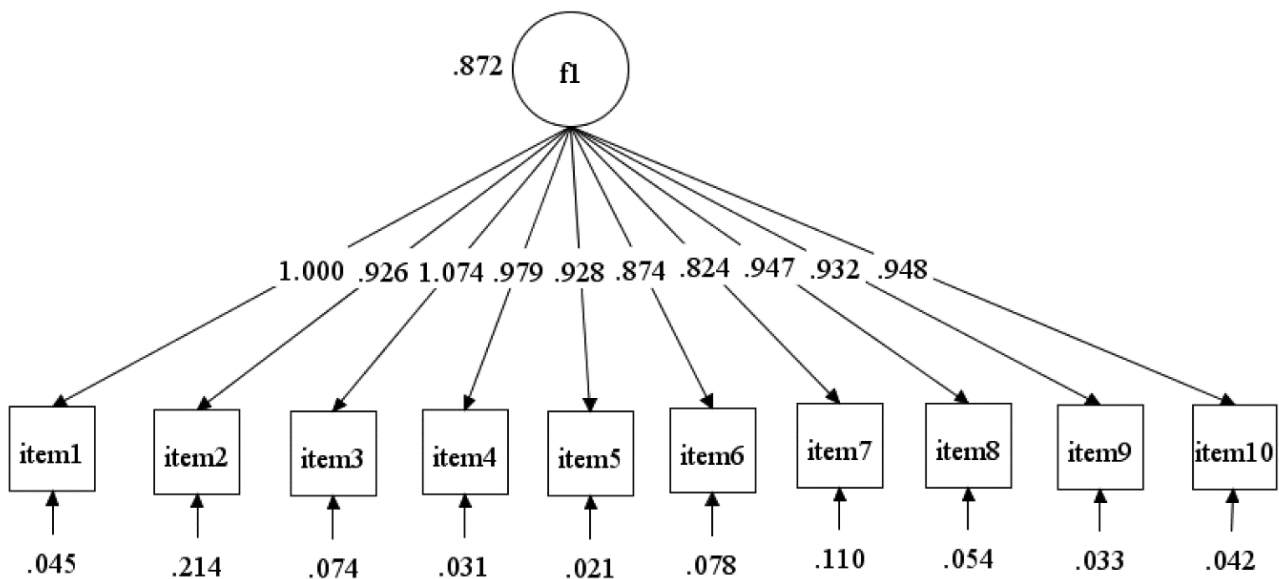


Figure 1. Internal structure between depression factor and items in confirmatory factor analysis model.

Table 5. Statistics in CFA (N = 306)

Model	χ^2	<i>df</i>	CFI	TLI	RMSEA [90% CI]	SRMR
1 factor model	1,982.60	35	.89	.85	.11 [06, .12]	.02

CFI (Bentler, 1990) and TLI (Bentler & Bonett, 1980) higher than .90; RMSEA lower than .08 (Browne & Cudeck, 1993); SRMR lower than .10 (Medsker, Williams, & Holahan, 1994) indicate adequate consistency level.

문항들과 우울 요인 간의 내적 구조를 보여주고 있다.

다른 검사와의 상관관계

한국형 우울장애 선별도구(K-DEP)의 다른 우울 평가도구들과의 수렴타당도를 검증하기 위하여 자료를 수집할 때 같은 피험자들을

Table 6. Correlations of the K-DEP Total Scores with the CES-D, BDI-II, and PHQ-9

K-DEP Total Score	Integrated group (N = 613)			Depressive disorder patient group (n = 90)		
	CES-D	PHQ-9	BDI-II	CES-D	PHQ-9	BDI-II
	.74	.78	.70	.77	.71	.71

대상으로 CES-D, BDI-II, PHQ-9 검사도 함께 실시하였다. Table 6은 문항수준과 검사점수수준에서 각 CES-D, BDI-II, PHQ-9의 총점 간의 상관분석 결과를 보여준다. 총점 간 상관계수를 살펴본 결과, 한국형 우울장애 선별도구는 CES-D, PHQ-9, BDI-II와 통계적으로 유의미하게 높은 상관을 보였다(통합군 .70-.78, 우울장애 환자군 .71-.77).

Table 7. Diagnostic Sensitivity and Specificity for Depressive Disorder (N = 613)

Task	Sensitivity	Specificity
K-DEP	.92	.94
CES-D	.71	.78
PHQ-9	.67	.88
BDI-II	.68	.88

신뢰도, 민감도 및 특이도

한국형 우울장애 선별도구(K-DEP)의 신뢰도를 분석한 결과 Cronbach's alpha는 .99였다. 또한, 한국형 우울장애 선별도구의 진단적 타당도를 검증하기 위해, 개발된 검사의 절단점 13점에서 민감도와 특이도를 613명으로 계산하고, 기존의 다른 검사들인 BDI-II(절단점: 22), PHQ-9(절단점: 9), CES-D(절단점: 25)와 민감도, 특이도를 비교하였다(Table 7). 한국형 우울장애 선별척도의 ROC 곡선 그래프는 Figure 2에 제시하였으며, Area Under the Curve(AUC)가 .827로 양호한 수준이다.

논 의

우울장애는 정신장애 중에서, 국내외적으로 가장 흔하게 발생하는 장애이다. 오늘날의 정신건강정책이 이러한 정신 장애의 조기 선별 및 개입, 예방을 강조하고 있는 만큼, 우울 증상을 효율적이면서도 신뢰도 있고 타당하게 선별할 수 있는 선별도구의 필요성은 점차 증대되고 있는 것이 현실이다. 기존에 활용하던 연구들이 타당화 과정 없이 외국의 선별도구를 번안하여 활용되었던 점, 국내 R&D 등에서 활용 시 외국에 저작권료를 지불해야만 하는 점, 우울장애 진단기준을 통합하지 못하는 점, 한국인의 고유한 문화 반응특성을 충분히 고려하여 개발되지 못한 점, 우울장애의 진단적 특이도와 민감도가 상대적으로 낮았던 점 등이 문제점으로 제기되어 왔다.

이에 본 연구에서는 우울장애를 신뢰도 있고 타당하게 선별할 수 있는 한국형 우울장애 선별도구(K-DEP)를 개발하고 그 타당성을 검증하였다. 검사 개발의 목적은 높은 수준의 민감도와 특이도를 갖추고, 접근이 용이하고 문항이 간명하여 초기 면접 장면에서 간편하게 사용될 수 있도록 하기 위해서였다. 특히, 기존의 우울 평가도구들의 타당화 연구가 사후 검증적인 접근법에 초점이 맞춰져 있었던 것에서 한 걸음 더 나아가, 본 연구에서는 검사 개발과정 자체의 타당성 확보를 위해 표준화된 검사 개발 절차인 Evidence Centered Design(ECD)을 활용하여 체계적이고 구체적인 검사 개발 타당도의 근거를 확보하였다.

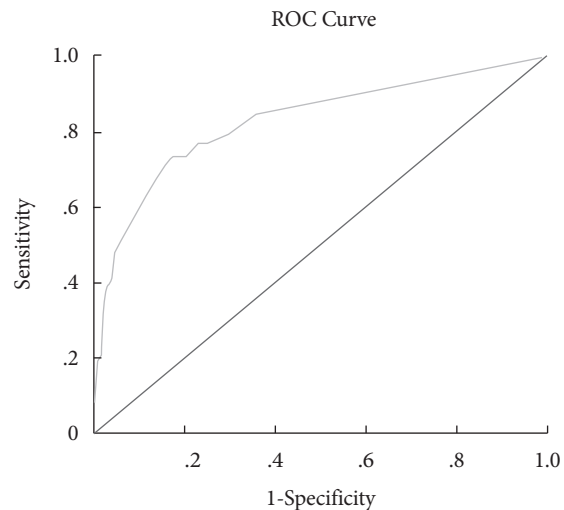


Figure 2. Receiver operating characteristic (ROC) curve of K-DEP.

내용적 측면에서 타당성을 갖추기 위해서 본 연구에서는 척도 제작 과정에서 전문가 자문단 회의, 문헌 검토, 표적집단면접법(Focus Group Interview)을 실시하였다. 그 결과, 개발된 K-DEP는 DSM-5에서 제공하는 주요우울장애의 핵심 진단기준 9개 영역을 모두 포함하며 DSM-V, RDoC에서 제시하는 우울장애 환자들의 진단적 특징을 반영하고 있음을 확인할 수 있었다.

내적 구조에 기초한 근거는 탐색적 요인분석과 확인적 요인분석을 실시하였다. 분석 결과 K-DEP의 내적구조는 1요인 구조로 확인되었다. 기존의 우울장애 평가도구들에 대한 연구에서는 우울장애의 요인을 1-4요인으로 측정한다는 것이 보고된 바 있다(Bae & Shin, 2005; Gibbons et al., 2012; Kim et al., 2002; Radloff, 1977; Richardson & Richards, 2008; Whisman, Perez, & Ramel, 2000). Bae와 Shin(2005)이 한국에서 사용되는 CES-D의 요인구조에 대한 연구들을 리뷰한 결과, 원 저자인 Radloff(1977)는 원판 CES-D가 4개의 요인으로 이루어져 있다고 언급하였고, Kim 등(2000)은 요인구조를 3개로 제시하여 요인구조가 원 척도와 상이하거나, Radloff(1977)가 제안한 원 척도와 요인의 개수는 같더라도 요인구조가 다르게 나타난 것으로 밝혀졌다(Cho et al., 1999; Chon & Rhee, 1992; Shin et al., 1991). Lee 등(2016)은 한국판 CES-D에 대해 요인분석을 실시한 결과로 2요인 구조를 보고하였다. 한국판 BDI의 경우 Kim 등(2002)이 확인적 요인분석을 사용하여 위계적 3요인 구조를 보고한 연구가 있다. PHQ-9의 경우 외국의 연구에서는 Richardson과 Richards(2008)이 정서적 요인과 신체적 요인으로 이루어진 2개의 요인구조를 추출하였다. 한편, Kim 등(2017)이 한국의 문화적특징을 반영하여 역기능적 우울을 측정하는 척도를

개발하는 과정에서, 개발된 척도의 요인구조를 분석한 결과, 일반 요인 1개와 특수요인 2개의 구조가 제시되었다. 이 연구에서는 모든 문항이 포함된 일반요인에서 주요 의미를 도출하는 것이 바람직하다는 의견이 제시되었으며, 2요인, 3요인, 4요인 모형의 적합도를 검토한 결과 모두 요인 간 상관이 높거나 여러 개념의 문항이 뒤섞여 있어 내용상 1요인으로 보는 것이 타당하다고 언급하였다(Kim et al., 2017). 1요인 구조의 결과를 보고한 또 다른 연구로는 Baas 등(2011)이 PHQ-9을 활용하여 요인분석을 실시한 연구가 있다. 본 연구에서 개발한 우울장애 측정도구가 선별도구임을 감안하면, BDI와 같이 선별적 기능 이외에 많은 문항으로 증상의 심각도를 측정하는 도구들에서 발견되는 다수의 요인구조와 일치하지 않을 수 있다. 추후 연구에서는 이러한 우울장애 선별도구의 1요인 구조가 선별도구와 증상평가 도구의 차이점을 반영하는 것인지, 한국인의 특수성을 반영한 것인지에 대해 추후 연구에서 다양한 문화의 배경을 지닌 더 많은 참여자를 대상으로 검증되어야 할 것이다. 본 연구에서 개발된 우울선별도구 K-DEP의 Cronbach's alpha 신뢰도는 .99로 높은 수준이었다.

또한, 다른 변수와의 관련성에 기초한 근거를 수집하기 위해 기존에 널리 사용되고 있던 평가 도구들과의 비교 분석을 실시하였다. K-DEP는 기존의 우울장애 평가도구인 BDI-II, PHQ-9, CES-D의 총점 간의 상관에서 비임상군의 경우는 각각 .74, .78, .70의 상관을, 우울 장애 환자집단의 경우는 각각 .77, .71, .71의 높은 수준의 상관관계를 보여주었다. 기존의 연구에서 보고된 임상군에서의 CES-D와 PHQ-9의 상관은 .77이었다(Milette, Hudson, Baron, & Thombs, 2010). PHQ-9과 BDI-II의 임상군에서의 상관은 .72에서 .84 사이로 보고된 바 있다(Dum, Pickren, Sobell, & Sobell, 2008; Hepner, Hunter, Edelen, Zhou, & Watkins, 2009; Kung et al., 2013; Titov et al., 2011). BDI-II와 CES-D의 경우 대학생 집단 대상으로 .86의 상관을 보고한 바 있다(Shan & Baldwin, 2008). 본 연구에서 개발된 K-DEP는 기존에 널리 사용되던 우울진단 검사 도구들인 BDI-II, PHQ-9, CES-D끼리의 상관과 비슷한 수준의 상관을 가지고 있으며, 이를 통해 K-DEP가 기존에 활용되고 있는 우울 진단 검사들과 같은 구성체를 측정하고 있고 공존 타당도가 있음을 확인했다.

응답반응에 기초한 근거는 문항 반응 분석을 통해 확인할 수 있었다. 각각의 개별 문항 수준에서 분석한 결과 모든 문항은 우울수준에 따라서 높은 정보값과 낮은 측정오차를 보여주며 우울수준을 적절하게 측정하였다. 또한 검사 전체에 대한 분석 결과에서도 검사 정보 곡선(TIF)이 표준편차 1.2 수준에서 가장 높은 지점을 형성하였으며, 표준편차 .8에서 1.6 사이의 지점에서 높은 정보값을

보였다. 이는, 우울 증상이 약 상위 11.5%에 해당하는 인원을 선별할 때 가장 높은 정보값과 낮은 측정오차를 지니고 있는 것을 의미하며, 우울 증상 상위 22%에서 5.5% 사이에 해당하는 집단에 대해 신뢰로운 측정 결과를 제공할 수 있음을 의미한다. 5% 정도 수준의 우울 장애의 평생 유병률과 조기 개입 및 예방을 목적으로 하는 선별 도구로서의 특성을 고려했을 때 이는 적절한 수준이라고 할 수 있다.

개발된 도구의 우울증 선별도구로서의 민감도와 특이도를 검증하였다. 한국형 우울장애 선별도구의 민감도와 특이도의 경우 각각 .92와 .94를 보고하였다. 이는 널리 활용되고 있는 BDI-II나 CES-D, PHQ-9 등의 다른 척도들과 비교했을 때 더욱 높은 수준이며, 최근 Gibbons 등(2013)이 컴퓨터 적응 기법을 사용하여 개발한 우울 선별도구에서 보고한 민감도와 특이도인 .95, .87과 비교하였을 때도 비슷한 수준이었다. 이는 K-DEP의 각 문항들이 한국인들을 대상으로 기존에 사용되던 척도들보다 보다 신뢰되고 있고 효과적으로 우울장애를 선별하는 것으로 보인다. 즉, 우울증이 있는 환자는 잘 선별하고 우울증상이 심하지 않은 정상집단 또한 잘 분별해 내는 것을 의미한다. 특히 본 척도가 선별도구로서 대규모의 연구나 임상 현장에서 활용될 때, 높은 특이도를 유지하는 것이 무엇보다도 중요하다. 예를 들어 전국민을 대상으로 우울장애 선별검사를 실시하고, 실제 유병률이 5% 수준임을 가정했을 때 약 4,750만 명의 사람들은 우울장애에 해당하지 않는다. 이때 특이도가 1% 수준만 떨어져도 47만 5천 명이 추가적으로 진단을 받게 되는 것으로 너무 높은 거짓 양성(false-positive)반응을 양산하게 된다. 이는 정밀진단 비용 등, 높은 추가 의료 비용으로 연결될 것이다. 따라서 선별도구의 특성상 높은 민감도를 유지하면서도 높은 특이도를 가지는 것이 무엇보다도 중요한 요건이라고 할 수 있다. 이에 본 척도는 기존의 선별기능을 갖춘 척도와 비교했을 때 월등히 높은 수준의 민감도와 특이도 수준을 보고하고 있으며 앞으로 선별 도구로서의 활용에도 적합할 것으로 예상된다.

그럼에도 불구하고, 본 연구에는 몇 가지 한계점을 가진다. 본 연구는 초기 타당화 작업에 초점이 맞춰져 예비검사(Pilot Study)를 통해서 선별된 문항들을 대상으로 문은행(Item Bank)을 구축하여, 최종적으로 신뢰도, 타당도, 민감도와 특이도가 높고, 간명한 문항을 선별하는 것을 목적으로 하였다. 따라서 연구의 목적이 문은행구축에 있다 보니 검사의 실용화 및 활용의 적합성에 대한 다양한 연구들이 충분히 진행되고 반영되지 못한 단점을 가진다. 즉, 검사의 채점기준, 다양한 집단에 따른 검사의 결과 비교, 검사의 실제 간명성 여부, 검사 실시에 고려해야 할 점, 검사의 시행에 따른 해석 및 사용 시 주의점 등 검사실행과 결과에 대한 연구가 부족하

한계를 지닌다. 검사를 개발하고 타당화하는 과정은 검사의 실행, 결과에 대한 해석 및 활용에 대한 연구를 포함해야 한다는 관점을 반영했을 때, 이에 대한 추가 연구를 진행할 필요가 있다. 또한 본 연구는 척도의 개발 과정과 예비타당화에 중점이 맞춰져 있다. 이로 인해서 척도를 구성하고 문항을 선별하는 하나의 방법으로 민감도와 특이도가 가장 높을 수 있는 조합을 고려하여 문항을 선정하였다. 이로 인하여 민감도와 특이도가 과잉 추정되었을 가능성을 배제할 수 없다. 따라서 3년차의 최종 타당화 연구에서는 본 문항 조합의 최종 민감도와 타당도 값을 공개할 예정이다.

K-DEP는 우울장애의 조기진단을 위한 일차적인 선별과 적절한 치료를 제공하기 위한 정책에 따라 보건복지부 정신건강기술개발사업단의 지원을 받아 개발되는 평가도구이다. 위에서 언급한 제한점에도 불구하고, 3년차 연구결과를 통합한다면 추후 지역사회, 정신보건센터, 연구장면 등에서 널리 활용될 척도도 개발될 것으로 기대된다. 또한 한국인을 대상으로 개발한 우리나라의 검사도구로서, 임상 장면과 연구장면에서 기존의 외국에서 개발된 우울 평가도구를 사용함으로써 소요되는 저작권료와 관련된 비용을 크게 절감하고 한국의 실정에 맞는 평가도구를 배포할 수 있을 것으로 기대된다. 더 나아가, K-DEP는 주요 우울 장애를 조기에 적절히 선별하여 잠재적인 우울증 환자에게 적절한 개입을 제공하는 목적으로 활용되어 우울증으로 인해 초래되는 사회적 비용을 절감할 수 있을 것으로 기대된다. K-DEP는 다년간의 연구과제를 통해 한국인들을 대상으로 우울장애의 모든 구성개념에 대하여 방대한 양의 예비문항으로 검사를 실시하고, 이에 대한 최신 심리측정이론을 적용하여 최종 문항을 선별하는 체계적이고 구체적인 검사개발 과정을 따르고 있다. 본 연구를 통해서 우울장애를 구분하기 위한 문항들을 선별해내어 갖출 수 있었으며, 우울장애 선별 문항들에 대한 한국인의 문항반응 특성을 수집할 수 있었고, 문항반응 특성에 대해 수집된 자료들은 추후 한국형 우울장애 선별도구의 채점에 반영하여 활용될 것이다. 우울장애 선별도구로서의 유용성과 타당성을 갖추고있으며 한국인의 특수성을 고려한 K-DEP에 대한 지속적인 관심이 필요하며, 이를 활용하여 현재까지의 연구를 보다 확장시키는 추후 연구가 활발히 이루어져야 할 것으로 여겨진다.

References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)* (5th ed.). Arlington, VA: American Psychiatric Publishing.
- An, J. Y., Seo, E., Lim, K. H., Shin, J. H., & Kim, J. B. (2013). Standardization of the Korean version of screening tool for depression (Patient Health Questionnaire-9, PHQ-9). *Journal of the Korean Society of Biological Therapies in Psychiatry*, 19, 47-56.
- Arroll, B., Khin, N., & Kerse, N. (2003). Screening for depression in primary care with two verbally asked questions: Cross sectional study. *BMJ*, 327, 1144-1146.
- Baas, K. D., Cramer, A. O., Koeter, M. W., van de Lisdonk, E. H., van Weert, H. C., & Schene, A. H. (2011). Measurement invariance with respect to ethnicity of the Patient Health Questionnaire-9(PHQ-9). *Journal of Affective Disorders*, 129, 229-235.
- Bae, S. W., & Shin, W. S. (2005). The factor structure of the CES-D Scale(the Center for Epidemiologic Studies-Depression Scale): An application of confirmatory factor analysis. *Health and Social Science*, 18, 165-19.
- Beck, A., Steer, R., & Brown, G. (2014). *Manual for the Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238-246.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.
- Cai, L., Thissen, D., & du Toit, S. (2011). IRTPRO 2.1 for Windows [Computer software]. Chicago, IL: Scientific Software International.
- Cho, M. J., Ha, Y. S., Han, K. J., Park, S. A., Song, M. S., Kim, Y. I., . . . Hong, J. P. (1999). Depressive symptoms in a Korean rural sample: Risk factor and factor structure. *Journal of Korean Neuropsychiatric Association*, 38, 266-277.
- Cho, M. J., & Kim, K. H. (1993). Diagnostic validity of the CES-D(Korean version) in the assessment of DSM-III-R major depression. *Journal of Korean Neuropsychiatric Association*, 32, 381-399.
- Cho, M. J., & Kim, K. H. (1998). Use of the Center for Epidemiologic Studies Depression(CES-D) Scale in Korea. *The Journal of Nervous and Mental Disease*, 186, 304-31.
- Choi, H. S., Choi, J. H., Park, K. H., Joo, K. J., Ga, H., Ko, H. J., & Kim, S. R. (2007). Standardization of the Korean version of Patient Health Questionnaire-9 as a screening instrument for major depressive disorder. *Journal of the Korean Academy of Family Medicine*, 28, 114-119.
- Chon, K. K., & Rhee, M. K. (1992). Preliminary development of Korean version of CES-D. *Korean Journal of Clinical Psychology*, 11, 65-76.
- Demirchyan, A., Petrosyan, V., & Thompson, M. (2011). Psychometric value of the Center for Epidemiologic Studies Depression (CES-D) scale for screening of depressive symptoms in Arme-

- nian population. *Journal of Affective Disorders*, 133, 489-498.
- de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford Press
- Dum, M., Pickren, J., Sobell, L. C., & Sobell, M. B. (2008). Comparing the BDI-II and the PHQ-9 with outpatient substance abusers. *Addictive Behaviors*, 33, 381-387.
- Ghio, L., Gotelli, S., Marcenaro, M., Amore, M., & Natta, W. (2014). Duration of untreated illness and outcomes in unipolar depression: A systematic review and meta-analysis. *Journal of Affective Disorders*, 152, 45-51.
- Gibbons, R. D., Hooker, G., Finkelman, M. D., Weiss, D. J., Pilkonis, P. A., Frank, E., . . . Kupfer, D. J. (2013). The CAD-MDD: A computerized adaptive diagnostic screening tool for depression. *The Journal of Clinical Psychiatry*, 74, 669.
- Gibbons, R. D., Weiss, D. J., Pilkonis, P. A., Frank, E., Moore, T., Kim, J. B., & Kupfer, D. J. (2012). Development of a computerized adaptive test for depression. *Archives of General Psychiatry*, 69, 1104-1112.
- Hepner, K. A., Hunter, S. B., Edelen, M. O., Zhou, A. J., & Watkins, K. (2009). A comparison of two depressive symptomatology measures in residential substance abuse treatment clients. *Journal of Substance Abuse Treatment*, 37, 318-325.
- Jung, S. Y., Kim, S. H., Park, K. H., Jaekal E. J., Lee, S. W., Choi, Y. Y., . . . Choi, K. H. (2017). A systematic review of validation studies on depression rating scales in Korea, with a focus on diagnostic validity information: Preliminary study for development of Korean screening tool for depression. *Anxiety and Mood*, 13, 1-7.
- Kim, J. H., Cho, Y. R., Park, S. H., Kim, H. R., Kim, S. H., & Pyo, K. S. (2002). Factor structure of the Korean version of the Beck Depression Inventory(BDI): An application of confirmatory factor analysis in clinical samples. *Korean Journal of Clinical Psychology*, 21, 247-258.
- Kim, J. Y., Choi, S. M., Seo, G. H., Cho, S. J., Kim, J. K., & Cho, M. J. (2000). Trans-cultural difference in expressing depressive symptoms in Koreans—Factor structure of the Center for Epidemiologic Studies Depression Scale(CES-D). *The Korean Journal of Psychopathology*, 9, 78-87.
- Kim, J. N., Lee, S. M., Choi, S. W., Chae, J. M., Seo, D. G., Choi, Y. K., & Cho, C. H. (2017). Development and analysis of a Dysfunctional Depression Scale based on Korean culture. *The Korean Journal of Health Psychology*, 22, 849-882.
- Kim, M., Lee, I., & Lee, C. (2007). The validation study 1 of Korean BDI-2: In female university students sample. *Korean Journal of Clinical Psychology*, 26, 997-1014.
- Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). New York, NY: Springer.
- Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9. *Journal of General Internal Medicine*, 16, 606-613.
- Kroenke, K., Wu, J., Bair, M. J., Krebs, E. E., Damush, T. M., & Tu, W. (2011). Reciprocal relationship between pain and depression: A 12-month longitudinal analysis in primary care. *The Journal of Pain*, 12, 964-973.
- Kung, S., Alarcon, R. D., Williams, M. D., Poppe, K. A., Moore, M. J., & Frye, M. A. (2012). Comparing the Beck Depression Inventory-II (BDI-II) and Patient Health Questionnaire (PHQ-9) depression measures in an integrated mood disorders practice. *Journal of Affective Disorders*, 145, 341-343.
- Lecrubier, Y., Sheehan, D. V., Weiller, E., Amorim, P., Bonora, I., Sheehan, K. H., . . . Dunbar, G. C. (1997). The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: Reliability and validity according to the CIDI. *European Psychiatry*, 12, 224-231.
- Lee, S., Oh, S. T., Ryu, S. Y., Jun, J. Y., Lee, K., Lee, E., . . . Choi, W. J. (2016). Validation of the Korean version of Center for Epidemiologic Studies Depression Scale-Revised (K-CESD-R). *Korean Journal of Psychosomatic Medicine*, 24, 83-93.
- Lim, S. Y., Lee, E. J., Jeong, S. W., Kim, H. C., Jeong, C. H., Jeon, T. Y., . . . Kim, J. B. (2011). The validation study of Beck Depression Scale 2 in Korean version. *Anxiety and Mood*, 7, 48-53.
- Lim, K., Park, Y., Kim, D., Shin, I., Lee, W., & Kim, J. (2009). A preliminary study of the standardization of the Korean version of the Patient Health Questionnaire-9. *Korean Society For Health Promotion And Disease Prevention*, 9, 275-28.
- Marsella, A. J., & Yamada, A. M. (2007). Culture and psychopathology: Foundations, issues, and directions. In S. Kitayama & D. Cohen (Eds.), *Handbook of cultural psychology* (pp. 797-818). New York, NY: Guilford Press.
- Medsker, G. J., Williams, L. J., & Holahan, P. J. (1994). A review of current practices for evaluating causal models in organizational behavior and human resources management research. *Journal of Management*, 20, 439-464.
- Milette, K., Hudson, M., Baron, M., & Thombs, B. D. (2010). Comparison of the PHQ-9 and CES-D depression scales in systemic sclerosis: Internal consistency reliability, convergent validity and clinical correlates. *Rheumatology*, 49, 789-796.
- Ministry of Health and Welfare & Samsung Medical Center. (2017). *The Survey of mental disorders in Korea*. Retrieved from <http://www.ssis.or.kr>
- Muthén, B., & Muthén, L. (1998-2015). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.
- Park, K. H., Yoon, S. W., Lee, W. H., & Choi, K. H. (2018, April). *Psychometric features and utility as a screening tool of Korean version of BDI-II*. Poster session presented at the 2018 annual symposium of the Korean Clinical Psychology Association, Seogwipo, Jeju.
- Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1, 385-401.

- Richardson, E. J., & Richards, J. S. (2008). Factor structure of the PHQ-9 screen for depression across time since injury among persons with spinal cord injury. *Rehabilitation Psychology, 53*, 243-249.
- Russell, L. B. (1994). *Educated guesses: Making policy about medical screening tests*. Oakland, CA: University of California Press.
- Sharp, L. K., & Lipsky, M. S. (2002). Screening for depression across the lifespan. *American Family Physician Journal, 66*, 1001-1008.
- Shean, G., & Baldwin, G. (2008). Sensitivity and specificity of depression questionnaires in a college-age sample. *The Journal of Genetic Psychology, 169*, 281-292.
- Sheline, Y. I., Gado, M. H., & Kraemer, H. C. (2003). Untreated depression and hippocampal volume loss. *American Journal of Psychiatry, 160*, 1516-1518.
- Shin, S. C., Kim, M. K., Yun, K. S., Kim, J. H., Lee, M. S. Moon, S. J., . . . Yoo, K. J. (1991). The Center for Epidemiologic Studies-Depression Scale (CES-D): Its use in Korea—Standardization and factor structure of CES-D. *Journal of Korean Neuropsychiatric Association, 30*, 752-767
- Sung, H., Kim, J., Park, Y., Bai, D., Lee, S., & Ahn, H. (2008). A study on the reliability and the validity of Korean version of the Beck Depression Inventory-II (BDI-II). *Journal of the Korean Society of Biological Therapies in Psychiatry, 14*, 201-212.
- Titov, N., Dear, B. F., McMillan, D., Anderson, T., Zou, J., & Sunderland, M. (2011). Psychometric comparison of the PHQ-9 and BDI-II for measuring response during treatment of depression. *Cognitive Behaviour Therapy, 40*, 126-136.
- Whisman, M. A., Perez, J. E., & Ramel, W. (2000). Factor structure of the Beck Depression Inventory—Second edition (BDI-II) in a student sample. *Journal of Clinical Psychology, 56*, 545-551.
- World Health Organization. (2008). *The global burden of disease: 2004 update*. Retrieved from http://www.who.int/healthinfo/global_burden_disease/2004_report_update/en/ World Health Organization. (2018). Depression fact sheet. Retrieved from <http://www.who.int/mediacentre/factsheets/fs369/en/>
- Yoo, S. W., Kim, Y. S., Noh, J. S., Oh, K. S., Kim, C. H., NamKoong, K., . . . Kim, S. J. (2006). Validity of Korean version of the Mini-International Neuropsychiatric Interview. *Anxiety and Mood, 2*, 50-55.
- Yoon, S. Y., Lim, J. H., & Han, C. (2012). Rating scales for measurement-based clinical practice of depression. *Korean Journal of Psychopharmacology, 23*, 136-146.
- Yu, B., Lee, H. K., & Lee, K. (2011). Validation and factor structure of Korean version of the Beck Depression Inventory Second edition (BDI-II): In a university student sample. *Korean Journal of Biological Psychiatry, 18*, 126-133.

국문초록

한국형 우울장애 선별도구 개발과정: 예비 타당화 연구

윤서원¹·이빛나²·박기호¹·정수연¹·김신향¹·박연수³·백승연³·김초롱³·이원혜⁴·최윤영²·이승환^{3,5}·최기홍¹

¹고려대학교 심리학과, ²한양사이버대학교 청소년상담학과, ³인제대학교 임상 감정인지기능 연구소, ⁴국립정신건강센터 임상심리과, ⁵인제대학교 일산백병원 정신건강의학과

현대의 정신건강 정책은 정신질환 예방과 조기개입을 강조함에 따라서 비전문가도 쉽게 평가할 수 있으며, 신뢰도와 타당도가 높은 선별도구가 필수적으로 요구되고 있다. 그러나 현재까지는 한국인을 대상으로 우울을 선별하는 진단적 민감도와 특이성이 높으며, 신뢰할만한 타당도의 근거를 충분히 제시하는 도구가 개발되어 있지 않거나, 로열티 비용을 지불해야 한다. 본 연구는 한국형 우울장애 선별도구(Korean screening assessment for depressive disorders, K-DEP)를 개발하는 과정을 공유하고 타당도와 관련된 수집된 근거들과 신뢰도를 제시함으로써 한국형 우울장애 선별도구의 타당화 연구를 진행하는 것을 목적으로 한다. 이에 본 연구는 DSM-5 및 RDoC 진단기준에 따른 우울선별도구의 측정 영역을 포괄하고 우울장애를 민감하게 선별하여 연구 뿐 아니라 병원 및 지역사회 정신건강 기관들에서 사용할 수 있으며, 동시에 선별도구의 특성에 맞게 민감도와 특이도가 높고 간결한 우울장애 선별도구의 개발을 목적으로 한다. 연구결과로 내용에 기초한 근거(Evidence based on contents), 내적구조에 기초한 근거(Evidence based on internal structure), 다른 변수와의 관련성에 기초한 근거(Evidence based on relation with other variables), 응답반응에 기초한 근거(Evidence based on response process), 결과에 기초한 근거(Evidence based on consequence)를 제시하였고 신뢰도, 민감도 및 특이도 또한 제시하였다. 마지막으로 한국형 우울장애 선별 도구의 실효성과 제3년차 연구과제에 대해 논의하였다.

주요어: 우울장애, 선별검사, 한국형 우울장애 선별도구, 타당도, 신뢰도, 타당화, DSM-5