

## 근거기반 심리평가의 기본 개념과 국내 적용에 대한 논의: 공황장애와 우울장애의 예시와 함께\*

김 빛 나

서울대학교 심리학과

김 지 혜<sup>†</sup>

삼성서울병원 정신건강의학과

근거기반 심리평가는 특정한 평가 목적을 위하여 평가할 구성개념, 도구, 평가의 전개 방식을 결정하는 데 연구와 이론을 통해 도출된 최선의 근거를 활용하는 심리평가이다. 근거기반 심리평가는 근거기반 심리치료와 더불어 심리학의 근거기반실천이 실현되기 위해 중요한 양대 구성 요소이다. 그러나 국외 및 국내 심리학계에서 근거기반실천은 주로 근거기반 심리치료에 대한 관심이 주를 이루어 왔으며, 근거기반 심리평가에 대한 소개 및 논의는 거의 이루어져 있는 바가 없다. 본 개관 논문에서는 근거기반 심리평가가 어떠한 배경에서 제안되었으며 그 기본 개념이 무엇인가를 소개하고자 한다. 더불어 근거기반 심리평가에 대한 이해를 돕기 위하여, 성인의 공황장애와 우울장애 진단 선별 평가에 있어서 어떠한 방식으로 구체화될 수 있는가를 예시하고자 한다. 마지막으로 국내에서 근거기반 심리평가가 보급되고 확산되기 위한 방안에 대하여 논의하고자 한다.

주요어 : 근거기반 심리평가, 근거기반실천, 공황장애, 우울장애, 반구조화된 진단적 면접, 임상가 평정 척도

\* 본 논문은 2014년 한국임상심리학회 추계학술대회 심포지엄에서 제 1저자에 의하여 발표된 내용을 수정 보완한 것임. 심포지엄에서 발표에 대한 중요한 피드백을 주신 박영숙, 김재환, 김명선, 강연욱 선생님과 초고에 대한 긴요한 코멘트를 주신 이은호 선생님께 감사드립니다.

<sup>†</sup> 교신저자(Corresponding Author) : 김지혜 / 성균관대학교 의과대학 삼성서울병원 정신건강의학과  
서울시 강남구 일원동 / Fax : 02-3410-0050 / E-mail : jihae0931.kim@samsung.com

## 근거기반 심리평가의 출현 배경

### 근거기반 운동의 확산

근거기반 심리평가(evidence-based assessment, EBA)의 출현 배경은 1980년대에 대두한 근거기반 의학(evidence-based medicine, EBM)으로 거슬러 올라간다. 당시 캐나다 McMaster 대학의 Sackett, Haynes 등은 환자에 대한 의학적 의사결정을 함에 있어 가장 좋은 최선의 근거를 사용하는 것을 강조하였으며(안형식, 2000; Sackett et al., 1996), 근거기반 의학이라는 용어는 Guyatt 등(1992)에 의해 제안되었다. 이는 현대에 접어들어 폭발적인 의학 지식의 증가에도 불구하고, 의료 제 분야의 임상적 의사결정에서 과학적 근거의 활용이 미흡하며(안형식, 2000; White, 1995, 1997), 졸업 후 시간이 지날수록 임상 실제에서 사용하는 지식의 수준이 감소하는 반비례 관계에 있다는 자성에서 비롯되었다. 이후 근거기반 운동은 하나의 패러다임으로 의학뿐만 아니라 근거기반 간호(evidence-based nursing), 근거기반 보건의료(evidence-based health care), 근거기반 사회복지(evidence-based social welfare) 등 인접 분야로 적용 범위가 확대되어 왔다.

비록 근거기반 운동이 의학에서 먼저 시작되었으나, 실증주의(pragmatism)에 대한 강조점 및 이상적인 교육과 훈련 모델로서 과학자-실무자모델(scientist-practitioner model)에 대한 지향 등은 심리학이 근거기반 운동과 공유하는 철학적 전통으로 볼 수 있다(안현의, 2003; Shakow et al., 1947). 그러나 심리학 내에서 본격적인 근거기반 운동은 다른 정신건강 관련 분야에 비해 비교적 늦게 시작되었으며(Spring & Neville, 2010), 평가에 비해 치료에

중점을 두고 진행되어 왔다. 심리학 내에서 명시적으로 근거기반 운동과 관련된 언급이 시작된 것은 1990년대 중반으로, 미국심리학회(APA)에서는 1992년 공동 대책 위원회를 출범하여 *Template for Developing Guidelines*(1995)를 발표하였다. 비슷한 시기에 APA의 임상분과(Division 12)에서는 경험적으로 지지된 심리치료(empirically supported treatments, EST)를 발표하였다. 이 시기까지 평가 자체와 관련된 언급은 존재하지 않았으며, 주로 심리치료의 효과 및 약물치료와 비교한 효과와 관련된 가이드라인들이었다. 2004년 영국의 National Institute for Clinical Excellence(NICE)에서는 장애별 가이드라인을 발표하였으나, 역시 평가 부분은 모호하게 처리되었다(Barlow, 2005).

2005년 APA에서는 Presidential Task Force on Evidence-Based Practice를 발족하였는데, 이 시기에 이르러 심리학 내에서의 근거기반 운동을 근거기반실천(evidence-based practice, EBP)이라는 용어로 일컫기에 이르렀다(임민경, 이지혜, 이한나, 김태동, 최기홍, 2013; APA, 2005). APA는 근거기반실천의 특징으로 다음의 3대 요소를 강조하여, (1) 연구를 통해 확보된 가능한 최선의 증거를 (2) 임상적으로 숙련된 치료자가 (3) 환자/내담자의 필요, 가치와 선호 등의 맥락을 고려하여 통합, 적용하는 의사결정 과정으로 정의하였다(APA, 2005, 2006).. 여기에 치료 성과에 대한 평가가 포함되었으나, 근거기반 심리평가 특정한 부분은 그다지 포함되지 못하였다(APA, 2005, 2006). 비슷한 시기인 2005년에 두 개의 학술지(*Psychological Assessment*, *Journal of Clinical Child and Adolescent Psychology*)에 근거기반 심리평가와 관련된 특별호가 출간되었으며, 이를 통해 근

거기반 심리평가의 포괄적 정의 및 임상 장면에서 자주 의뢰되는 장애별 평가 가이드라인들이 제안되었다.

#### 기존 심리검사들에 대한 비판

한편, 근거기반 평가의 필요성이 대두된 또 다른 배경에는 기존 심리검사들의 심리측정적 속성이나 임상적 유용성에 대한 경험적 지지가 부족하다는 비판들이 존재한다(Hunsley & Mash, 2005, 2007; Youngstrom, 2013). 전통적인 풀 배터리(full battery) 검사들을 구성해 온 심리검사들 중 특히 Rorschach 검사의 타당도 및 진단과 예후 예측에의 유용성에 관하여 반복된 의문 제기가 있었다(Hunsley & Bailey, 1999, 2001; Meyer & Archer, 2001; Meyer & Handler, 1997; Stricker & Gold, 1999; Wood, Garb, Lilienfeld, & Garb, 2003). 또한 Draw-A-Person(DAP) 등의 그림검사들과 Thematic Apperception Test(TAT)가 실시, 채점, 해석되는 방식이 일관되지 않으며, 검사로서의 신뢰도와 타당도를 입증하는 증거들이 충분치 않다는 점들에 대해서도 비판이 있었다(Vane, 1981). 나아가 심리측정적 속성이 비교적 입증되었다고 간주되어 온 객관적인 성격 검사인 Minnesota Multiphasic Personality Inventory-2(MMPI-2)의 경우에도 MMPI-2 결과의 피드백 제공 여부가 내담자의 치료 경과 미치는 영향이 미미하다는 보고도 있었다(Lima et al., 2005).

그럼에도 불구하고 지난 30년간 임상심리학자들이 사용 및 교육해온 심리검사의 종류는 관성적으로 거의 변하지 않아 왔다(Camara, Nathan, & Puente, 2000; Cashel, 2002; Childs & Eyde, 2002; Piotrowski, 1999). 심리평가는 심리

학의 필수 기술이자 심리학에 핵심적인 개인차를 개념화하는 방식이었으나, 이러한 배경에서 미국 내에서 심리평가에 대한 훈련과 보험 청구가 최근 10년 사이에 크게 약화되었다는 분석이 있었으며(Merenda, 2007), 대안적인 심리평가 방식의 필요성이 제기되었다.

#### 근거기반 심리평가의 기본 개념

본고에서 소개하고자 하는 근거기반 심리평가의 정의는 Hunsley와 Mash(2007)가 상위랭킹의 임상심리 학술지인 *Annual Review of Clinical Psychology*에 게재한 개관 논문에 실린 것이다. 그들은 “특정한 평가 목적을 위하여 평가할 구성개념, 도구, 평가의 전개 방식을 결정하는 데 연구와 이론을 활용하는 심리평가”라는 정의를 제안하였으며, 근거기반 심리평가가 본질적으로 의사결정 과정이라는 점을 강조하였다. 이들에 의하면 근거기반 심리평가는 다음의 세 가지를 핵심적인 특징으로 한다.

첫째, 정신병리와 정상 발달에 대한 연구결과와 과학적 이론이 평가될 구성 개념과 평가 과정을 선택하는 데 지침이 된다. 근거기반 심리평가는 임상심리학자 훈련의 Boulder Model(1949)과 같이 임상 실제와 과학적 연구가 연결성을 유지하는 것을 강조한다. 이는 임상심리학의 과학적 요소에 대한 강조점을 재확인하는 것이다(Barlow, 2005; McFall, 2005). 이론에 대한 깊이 있는 지식은 임상심리학자가 검사 기술자의 역할을 넘어 전문가로서의 역량을 발휘하기 위해 갖추어야 할 필요조건이라 할 수 있다.

표 1. 심리측정적 건전성의 기준 (Hunsley & Mash, 2008)

범주	기준
내적 일관성	<p>대규모, 관련 있는, 임상 표본에서의 총점(그리고 필요한 경우 하위 점수의 중심 경향성과 분포 측정치 여럿의 대규모, 관련 있는 (임상 그리고 비임상) 표본에서의 총점(그리고 필요한 경우 하위 점수의 중심 경향성과 분포 측정치 하나 이상의 대규모 '대표' (임상 그리고 비임상) 표본에서의 총점(그리고 필요한 경우 하위 점수의 중심 경향성과 분포 측정치 a 값이 .70-.79라는 증거가 우세</p> <p>a 값이 .80-.89라는 증거가 우세</p> <p>a 값이 .90 이상이라는 증거가 우세</p> <p>k 값이 .60-.74, Pearson 상관 또는 담내상관계수가 .70-.79라는 증거가 우세</p> <p>k 값이 .75-.84, Pearson 상관 또는 담내상관계수가 .80-.89라는 증거가 우세</p> <p>k 값이 .85 이상, Pearson 상관 또는 담내상관계수가 .90 이상이라는 증거가 우세</p>
평가자간 신뢰도	<p>머칠에서 몇 주 간격으로 검사-재검사 상관이 적어도 .70 이상</p> <p>몇 달 간격으로 검사-재검사 상관이 적어도 .70 이상</p> <p>1년 또는 그 이상의 간격으로 검사-재검사 상관이 적어도 .70 이상</p>
검사-재검사 신뢰도	<p>추정되는 구성개념의 영역들이 명확히 정의되고, 신빙된 문항들이 그 영역을 전체적으로 잘 대표하는지가 확인될</p>
내용 타당도	<p>적절 기준에 더하여, 도구와 또는 측연들(지시문, 개별 문항 등이 평가자(인문가 또는 예비(pilot) 피험자에 의해 평가될 양호 기준에 더하여, 다수 평가자 집단의 양적 평정을 통하여 평가될</p>
구성 타당도	<p>독립적으로 반복 검증된 구성 타당도, 즉, 예측 타당도, 공준 타당도, 수렴 타당도, 번별 타당도 중 일부 증거가 있음</p> <p>다수의 타당도 범주에 걸쳐 독립적으로 반복 검증된 구성 타당도의 증거가 우세함</p>
타당도 일반화	<p>양호 기준에 더하여, 다른 임상적 데이터와 비교하였을 시 증분 타당도의 증거가 있음</p> <p>도구의 유용성을 1개 이상의 인구통계학적 집단(연령, 성별, 인종 등) 또는 다양한 맥락(가정, 학교, 입원 등에서 입증하는 일부 증거가 있음</p> <p>도구의 유용성을 1개 이상의 인구통계학적 집단(연령, 성별, 인종 등) 또는 다양한 맥락(가정, 학교, 입원 등에서 입증하는 증거가 우세함</p> <p>도구의 유용성을 1개 이상의 인구통계학적 집단(연령, 성별, 인종 등) 그리고 다양한 맥락(가정, 학교, 입원 등에서 입증하는 증거가 우세함</p>
치료 변화에의 민감도	<p>치료 경과에 따른 변화에 민감하다는 증거가 있음</p> <p>치료 경과에 따른 변화에 민감도를 입증하는 독립적으로 반복 검증된 증거</p>
임상적 유용성	<p>양호 기준에 더하여, 다른 유형의 치료들에 걸친 변화 민감도의 증거</p> <p>활실적 고려(비용, 실시의 용이도, 시간 등을 생각했을 때에도 평가 결과가 임상적으로 유용할 가능성이 있음</p> <p>적절 기준에 더하여, 평가 결과가 임상적 이익(더 나은 치료 성과, 낮은 발작률, 더 높은 환자 만족도)을 준다는 출판된 증거</p> <p>적절 기준에 더하여, 평가 결과가 임상적 이익(더 나은 치료 성과, 낮은 발작률, 더 높은 환자 만족도)을 준다는 출판된 증거가 반복 검증될</p>

주. 적절=appropriate, 양호=good, 우수=excellent

둘째, 평가 목적을 달성하는 데 심리측정적으로 건전한 도구들을 사용한다. 심리측정적 건전성을 갖추기 위해서는 표준화되어 해석 가능한 기준(민감도, 특이도, 절단점 등)을 갖고 있으며, 적절한 신뢰도와 타당도를 가져야 한다. 보다 구체적인 심리측정적 건전성의 기준(“good enough principle”)으로 9개 범주가 제안되었는데, 이상적으로 심리측정적으로 건전한 도구는 기준, 내적 일관성, 평가자간 신뢰도, 검사-재검사 신뢰도, 내용 타당도, 구성 타당도, 타당도 일반화, 치료 변화에의 민감도, 임상적 유용성의 요건을 갖추어야 하며, 비록 이 범주 내의 모든 요건을 만족시키지는 못할지라도 이 중 다수를 갖추어야 한다는 것이다(Hunsley & Mash, 2008). 표 1에 제시된 바와 같이, 예를 들어 내적 일관성  $\alpha$ 의 경우, .70 이상을 적절한, .80 이상을 양호한, .90 이상을 우수한 것으로 판단할 수 있다.

여기서 한 가지 유의할 점은 심리측정적 속성이 “도구 자체”의 속성이 아니라, 본질적으로 “조건적” 속성이라는 것이다. 이는 다음의 두 가지 시사점을 지닌다. 우선 다양성 문제이다. 심리측정적 속성은 특정한 표본에서 특정한 목적으로 사용되었을 때의 속성이기 때문에, 연령, 성별, 인종 등 다양한 표본에서의 반복 검증(replication)을 필수적으로 요구된다. 또한 평가자는 도구 사용의 목적(선별, 진단, 치료 경과 관찰)을 명확히 인지해야 하며, 그 목적에 적합하게 사용하여야 한다. 또 다른 시사점은 근거기반 심리평가는 임상적 의사결정을 위한 것으로, 일반적인 평가라기보다는 특정 문제 또는 장애를 배제 또는 확진(rule out/in)하기 위한 문제 및 장애 특정적 평가라는 점이다.

셋째, 평가과정 전체가 경험적으로 평가된다. 심리검사는 심리평가와 동의어가 아니며, 심리평가는 검사 도구의 선택과 실시, 해석, 그리고 면담, 임상적 관찰 등 다양한 평가 자료들의 통합을 포함한다. 아직까지는 심리검사가 아닌 심리평가의 정확성(타당도)이나 쓸모(임상적 유용성)에 대한 연구는 없으며, 현재까지는 근거기반이라고 할 수 있는 통합된 평가 전략이 존재하지 않는 현실적 한계가 있다. 따라서 지금까지 근거기반 심리평가에 대한 논의가 평가도구 중심의 논의에 지나지 않는다는 비판은 타당한 것이며, 이 세 번째 요건은 아직 미완의 단계라고 볼 수 있다. 그러나 학자들은 지금까지는 개별 평가 도구 중심의 개관과 가이드라인일지라도, 향후 이것이 평가과정 전체로 확장되어야 한다는 인식에는 동의하고 있다(Hunsley & Mash, 2005, 2007, 2008).

이와 관련하여 근거기반 심리평가가 전통적인 심리측정적 속성인 신뢰도와 타당도 이상의 개념으로, 임상적 유용성을 가져야 한다는 주장이 제기되었다. 임상적 유용성은 치료적 유용성(평가 자료가 긍정적인 치료 결과에 기여하는 정도), 진단적 유용성(평가 자료가 정확한 진단에 기여하는 정도), 평가와 관련된 비용(시간, 돈, 잠재적 위험 등)에 비해 얻는 이익을 포괄한다(Hunsley & Mash, 2005). 이와 유사한 견지에서 Youngstrom(2013)도 유용성의 3P인 예측(prediction), 처방(prescription), 과정(process)을 조직 원리로 삼을 때 전통적인 심리평가의 제한점을 극복하고, 심리평가의 교육과 실시가 정당화될 수 있을 것이라고 논하기도 하였다.

근거기반 심리평가를 임상 실제에서  
어떻게 구체화할 것인가:  
성인 공황장애와 우울장애의  
진단과 선별평가의 예시

앞서 상술했듯이 근거기반 심리평가는 특정한 평가 목적 하에 진행되며, 본질적으로 문제 및 장애 특정한 심리평가이다. 각 장애별 가이드라인을 제시하는 것이 본고의 주목적은 아니며 이와 관련해서는 별도의 개관 논문이 필요하겠지만, 근거기반 심리평가를 어떻게 임상 장면에서 구체화할 수 있는가에 대한 이해를 돕고자 공황장애와 우울장애를 들어 예시하고자 한다. 공황장애는 근거기반 평가를 구성하는 요소에 대한 전문가 집단의 합의가 비교적 잘 이루어져 있는 불안장애로, 근거기반 평가의 대표적 사례로 예시될 수 있다(Barlow, 2005; Shear & Maser, 1994). 또한 우울장애는 유병율과 공병율 측면에서 임상 장면에서 가장 자주 의뢰되는 정신장애이다(보건복지부, 2011; American Psychiatric Association, 2013).

심리평가의 목적은 진단, 선별, 예후 예측, 사례 개념화, 치료 계획 및 치료 성과 평가 등 의뢰 문제에 따라 매우 다양할 수 있다(Hunsley & Mash, 2008). 이 모든 평가 목적을 섭렵하는 것은 지면 및 현실적 제약으로 인하여 불가능할 것이며, 다른 평가 목적을 위한 근거기반 심리평가도 가능할 것이나, 본고에서는 우선 진단과 선별이라는 평가 목적에 집중하고자 한다. 즉, 비교적 짧은 시간 내에 각 장애를 진단 내리고 현재의 증상 심각도를 판단하거나 또는 추후에 추가적 평가가 필요한 사례를 선별해내기 위한 심리평가에 포함되어야 할 요소들을 주로 다룰 것이다.

진단은 제시되는 임상적 문제들의 속성 및 원인을 결정하는 가장 기초적인 임상적 의사결정 과정이다. 정확한 진단은 이후 효과적 치료가 이루어지는 출발점이 될 수 있으며, 현재까지 활용 가능한 심리학적 이론과 연구 결과들도 진단을 중심으로 축적되어 있다<sup>1)</sup>. 진

1) 어떠한 학문 분야이든 연구 대상을 정의하는 것은 가장 기본적인이지만, 의의로 쉽게 대답하기 어려운 근본적인 질문이다. 정신장애의 진단이 갖는 의미와 그 신뢰도, 타당도의 문제는 별도의 깊이 있는 고찰을 필요로 하는 주제로, 이를 주제의 중량감에 걸맞은 충분한 깊이로 다루는 것은 본고의 주 목적에서 다소 벗어나는 일이 될 것이다. 다만, 본고에서 근거기반 심리평가의 예시로서 진단의 문제를 다루고 있기 때문에, 그 배경이 되는 논의를 부가적으로 간략히 소개하고자 한다.

정신장애는 독립적인 객관적 기준(cf. lab test)이 부재한 상황에서 주로 심리적, 행동적 표현 수준의 증상과 징후들을 묶은 일종의 증후군(syndrome)이며, 사회문화적 기준이 개인의 경험을 해석하는 데 영향을 준다는 점에 있어, 신체적 질환(disease)의 진단과는 다른 특성을 지닌다(Claridge, 1995; Kleinman, 1988). 정신장애의 병인과 기저 병리과정이 충분히 밝혀지지 않은 상황에서 일관된 진단 분류체계를 확립하고자 하는 노력의 일환으로, DSM-III(1980)에서는 탈이론적, 기술적 진단기준을 도입함으로써 진단의 평가자 간 신뢰도를 향상시키는 데 기여하였다. 그러나 Kleinman(1988)의 용어로 “타당도 없는 신뢰도(reliability without validity)” 문제는 해결되지 않았으며, 진단의 타당도 측면에서는 아직도 여러 가지 문제들이 남아 있다(예: 장애 간 높은 공병율, 장애군 내의 이질성 등). 더욱이 DSM과 같이 정상과 이상의 경계를 질적으로 구분하는 범주적 진단 분류체계의 한계에 대하여 여러 비판들이 노정된 바 있다(for review, see Hyman, 2010). 이는 최근 개정된 DSM-5(2013)에서 지금까지 밝혀진 기저 병인과 병리에 근거하여 일부 진단군의 분류를 수정하고, 연속적인 차원 개념의 양적 평

단과 선별평가를 위해 사용되는 도구들은 필요에 따라 사례개념화와 치료 성과 평가에도 활용될 수 있다.

무엇이 가장 최적의 근거기반 심리평가를 구성하는가 하는 문제는 시간과 비용, 기관 내 의사결정과 업무 분담에 관한 인식, 평가 인력의 숙련도 등 각 개별 기관의 특성에 따라 달라짐을 감안하여야 할 것이다. 그러나 앞으로 소개되는 두 장애의 근거기반 심리평가를 구성하는 요소에 관해서는 선행 연구들에 의한 합의가 비교적 이루어져 있으며, 이를 바탕으로 각 기관의 특성에 적합한 평가과정을 고안하는데 참고할 수 있을 것으로 생각된다.

#### 반구조화된 진단적 면접

개별 장애의 근거기반 심리평가에 대한 각론에 앞서, 장애의 진단에 공통적으로 사용되는 반구조화된 진단적 면접의 필요성 및 자주 사용되는 도구들에 대한 소개가 선행되어야

할 필요가 있다. 반구조화된 진단적 면접은 임상 장면에서 흔히 편의적으로 사용되는 비구조화된 면담에 비해 진단의 신뢰도를 크게 향상시킬 수 있다(Miller, 2001). 그러한 진단적 면접 중 가장 널리 쓰이는 것이 Structured Clinical Interview for DSM-IV(SCID)이다(First, Spitzer, Gibbon, & Williams, 1996). SCID는 주요 제1축 장애를 DSM-IV 기준에 따라 신뢰도 높게 진단하기 위한 반구조화된 면담 도구로, 한국어판도 번안되어 있다(한오수, 홍진표, 2000). SCID는 현재와 과거에 대한 개방형 질문으로 이루어진 개괄 섹션 및 각 장애별 모듈 설계로 구성되어 있으며, 질문 순서는 경험 많은 임상가들의 감별진단 과정과 유사하게 설계되어 있다. 실시에는 각 사례 당 약 60~90분이 소요되며, 진단기준을 구성하는 각 증상의 존재 여부에 대하여 1(정보 불충분), 2(없음), 3(역치 미만), 4(역치)으로 판단을 내리게 된다.

한 가지 강조하여야 할 점은 SCID가 단순히 질문에 대한 응답자의 대답에 대한 기계적 평가가 아니라, 진단기준 문항에 관한 임상적 판단을 요한다는 점이다(한오수, 홍진표, 2000; First et al., 1996). 만약 A. 기분 삽화, 주요 우울증 삽화에 대한 진단기준 문항 A1(거의 매일, 하루 중 대부분 지속되는 우울감)에 대한 평가에서 응답자가 증상을 부인 하더라도 면담 시 관찰된다면 충분히 3으로 평정할 수 있다. 따라서 SCID를 신뢰롭게 실시하기 위해서는 정신병리에 대한 전문적 지식과 면담 훈련이 필요하다. 현재 SCID는 국제적으로 정신장애 연구에서 표준(the gold standard)으로 받아들여지고 있으며, 연구자 및 임상가의 의사소통을 위한 공유된 용어로서의 유용성을 지니고 있다(각주 1 참조).

정을 함께 고려하는 배경이 되었다(Hyman, 2010; Regier, Narrow, Kuhl, & Kupfer, 2009).

따라서 진단의 문제와 관련하여, 현재로서 합리적인 차선책은 보편적인 진단체계 자체를 폐기하는 것보다는 연구나 치료의 대상을 선택하는 기준으로 구조화된 면접을 통하여 진단의 신뢰도를 높이고, 그 진단의 타당도를 늘 의심하며 신중하게 적용하는 것이다. 즉, 연구자와 임상가들은 DSM 등 정신장애 진단이 전문가 간 ‘공유된 언어’로서 유용한 것임을 인지하고, 진단을 실체처럼 취급하는 오류(“problem of reification”, Hyman, 2010)를 범하지 않도록 유의해야 한다. 본고에서 소개하는 반구조화된 면접을 통한 범주적 진단과 장애별 양적 평정 척도들의 조합은 이러한 배경에 대한 인식 내에서 활용되어야 할 것이다.

### 다중 평가 도구의 활용

전통적인 심리평가와 마찬가지로 근거기반 심리평가에서도 단 하나의 정보에 근거하여 의사결정을 내리기보다는, 다중 평가 도구에서 얻어진 정보들을 통합하고 이를 통해 수렴 증거를 추구하는 방식을 선호한다. 이러한 다중 평가 도구에는 반구조화된 면접, 임상가 평정 척도, 자기보고식 척도, 자기관찰 일지, 행동적 평가 또는 심리생리적 측정<sup>2)</sup> 등이 포함될 수 있다(Antony & Rowa, 2005).

앞으로 예시할 공황장애와 우울장애의 근거기반 심리평가에서도 공통적으로 다중 평가 도구들이 활용되며, 각 평가 도구들은 각기 다른 목적을 위해 사용된다는 점을 분명히 해야 할 것이다. 일단 진단적 의사 결정을 내리기 위해서는 (반)구조화된 면접을 실시한다. SCID는 대부분의 제1축 장애에 공통적으로 사용될 수 있는 기본적인 면접 도구이며, 장애의 종류 또는 실시 시간, 평가자 특성 등에

2) 심리 생리적 측정은 심리평가의 중요한 구성요소가 될 수 있으나, 현실적으로 심리학자에 의해 사용 가능하며 잘 검증된 생리적 측정도구의 선택지는 그리 많지 않은 것으로 보인다. 일례로 공황장애의 평가에서 광장공포증이 있는 상황에 대한 행동적 접근 검사(Behavioral Approach Test) 중 실제 심박율이나 혈압을 기록하여, 불안 증상에 대한 주관적 평정과 객관적 수치 간에 격차가 있음을 보여줄 수 있다(Keller & Craske, 2008). 그러나 생리적 측정치에 대한 연구 결과들은 비일관적인 경우가 있으며, 각 수치가 어떠한 임상적 의미로 해석될 수 있는가에 관해서는 아직 더 많은 연구가 필요한 상황으로 생각되어(Bystritsky, Craske, Maidenberg, Vapnik, & Shapiro, 1995; Keller & Craske, 2008에서 재인용, p.240), 본고에서 각 장애의 근거기반 심리평가 도구에 포함시키지 않았다.

따라 SCID 외 다른 진단적 면접 도구의 사용이 고려될 수 있다. 진단이 내려지고 난 후 증상 심각도를 평가하기 위해서는 각 장애 별로 개발된 임상가 평정 척도(clinician rating scale)가 일차적으로 활용된다. 이는 현재의 증상 심각도, 치료 후 반응 관찰, 관해의 판단 등에 대한 객관적인 정보를 제공하기 위한 목적을 갖고 있다. 더불어 환자 본인에 의하여 작성된 자기보고식 척도들도 평가에 중요한 정보를 제공할 수 있으며, 이를 위해서는 각 장애에 핵심적인 구성개념을 측정하고 심리측정적 속성이 비교적 잘 확립된 것으로 확인된 도구들을 선택하여 실시한다.

### 공황장애의 근거기반 심리평가

공황장애는 예기치 못한, 반복적으로 발생하는 공황발작을 핵심 증상으로 하며, 그에 대한 예기불안이나 회피 행동이 수반되는 불안장애이다(APA, 2013). 공황장애는 평가 전략에 대한 전문가들의 합의가 비교적 도출되어 있는 불안장애로, 미국 국립 보건원(National Institutes of Health, NIH) 컨퍼런스에서 전문가들은 공황장애 평가에 필수적이라고 생각하는 영역 및 각 영역을 평가할 수 있는 도구들의 목록을 표 2와 같이 제안하였다(Antony & Rowa, 2005; Shear & Maser, 1994).

우선 공황장애의 진단에는 반구조화된 진단적 면접 도구를 사용하는 것이 권고되며, 이에 SCID(First et al., 1996)나 ADIS-IV(Anxiety Disorders Interview Schedule for DSM-IV(ADIS-IV; Di Nardo, Brown, & Barlow, 1994)가 사용될 수 있다. 면담의 결과는 SCID의 증상 체크리스트 형태로 제시될 수 있으며, 여기에는 공황발작을 변별하여 정의하는 특성(예기치 못한 갑작



표 2. 공황장애 평가 프로토콜 (Shear & Maser, 1994; Antony & Rowa, 2005에서 재인용)

영역	평가 도구	도구의 유형
진단	SCID-IV(First et al., 1996; 한오수, 홍진표, 2000)	반구조화된 면접
	ADIS-IV(Di Nardo et al., 1994)	반구조화된 면접
	Panic Attack Diary(Rapee et al., 1990)	일지
상황적 단서와 회피행동	Mobility Inventory for Agoraphobia (Chambless et al., 1985)	자기보고식 척도
	상황적 두려움, 회피, 안전 행동을 기록하기 위한 일지	일지
	행동적 접근 검사	행동적 평가
내부감각적 단서	ASI-R(Peterson & Reiss, 1993; 김지혜 등, 2004)	자기보고식 척도
	증상 유발 운동	행동적 평가
인지적 특성 및 기능적 장애	Agoraphobic Cognitions Questionnaire (Chambless et al., 1984; 원호택, 박현순, 권석만, 1995)	자기보고식 척도
	Illness Intrusiveness Rating Scale (Devins, 1994)	자기보고식 척도
	ADIS-IV 또는 SCID-IV	반구조화된 면담
	Sheehan Disability Scale(Sheehan, 1983; 박준영, 김지혜, 2010)	자기보고식 척도
공병 문제	SCID-IV 또는 ADIS-R	반구조화된 면접
	Psychiatric Diagnostic Screening Questionnaire (Zimmerman & Mattia, 1999)	자기보고식 척도
	BDI-II (Beck et al., 1996)	자기보고식 척도

주. 한국판이 번안 및 타당화되어 있는 것으로 확인된 척도는 참고문헌을 추가함.

스러운 발생 및 증상의 점증과 지속에 관한 시간 기준)과 예기불안, 발작으로 인한 생활방식 변화, 상황 회피 여부 및 회피하는 상황의 정보가 함께 제공될 수 있다. 또한 반구조화된 진단적 면접을 통해 확인해야 할 추가적 영역은 공병(comorbidity)이다. 특히 우울은 불안장애와 높은 공병율을 갖고 있으며, 우울이 동반되는 경우 추후 경과와 치료 반응에 부정적인 영향을 줄 수 있기 때문에 관심이 요망된다(Brown, Campbell, Lehman, Grisham, & Mancill, 2001).

일단 진단이 확진되고 나면, 장애의 심각도 정도를 평정하기 위한 임상가 평정 척도가 사용될 수 있다. 공황장애의 경우 Shear 등(1997)

이 개발한 공황장애 심각도 척도(Panic Disorder Severity Scale, PDSS)가 표준적으로 사용되고 있으며, 한국판도 번안되어 있다(김정범, 2001). PDSS는 공황발작의 빈도, 공황발작 동안 경험하는 고통, 예기불안, 광장공포증적 두려움과 회피, 내부감각적(interoceptive) 두려움과 회피, 직무 수행 장애나 고통, 사회적 기능의 장애나 고통의 7개 영역을 평가하고, 각 항목의 점수를 합하여 총점을 구한다. 그 총점에 따른 장애의 심각도를 판단하는 기준이 선행 연구들에 의해 제시되어 있는데, 0-3점을 관해(remission), 4-6점을 부분 관해(partial remission), 7-10점을 경도(mild), 11-15점을 중등도(moderate to marked), 16점 이상을 심각한(severe) 것으로

판단한다(Shear et al., 2001; Yamamoto et al., 2004). PDSS 총점은 치료 반응의 관찰, 관해 여부의 판정 등에도 사용된다. 주지할 것은 PDSS는 진단을 위한 도구가 아니라는 점이다. PDSS는 별도로 진단이 확진된 후, 공황장애의 심각도 또는 관해 여부를 판단하기 위해 사용되는 도구이기 때문에, PDSS 총점을 진단 정보로 활용하는 것은 부적절하다.

아울러 환자에 의해 작성된 자기보고식 척도들도 평가에 부가적으로 고려될 수 있다. 공황장애의 평가에 사용될 수 있는 대표적인 자기보고식 척도들은 불안민감성을 측정하는 ASI-R(Anxiety Sensitivity Inventory-Revised; 김지혜 등, 2004; Peterson & Reiss, 1993), 불안 경험의 결과에 대한 파국적인 사고 내용을 담고 있는 ACQ(Agoraphobic Cognitions Questionnaire; 원호택, 박현순, 권석만, 1995; Chambless, Caputo, Bright, & Gallagher, 1984), 내부감각적 두려움을 별도의 요인으로 갖고 있는 APPQ(Albany Panic and Phobic Questionnaire; 김지혜 등, 2004; Rapee, Craske, & Barlow, 1995)가 있으며, 일반적인 불안 증상을 측정하는 BAI(Beck Anxiety Inventory; Beck, Epstein, Brown, & Steer, 1988), STAI(김정택, 1978; Spielberger, 1970) 등도 있다. 이 중에서 ASI-R은 공황장애에 특정적으로 개발된 척도로, 전체 점수 및 4개 요인(심혈관계 증상에 대한 두려움, 걸으려 드러나는 불안 증상에 대한 두려움, 호흡기계 증상에 대한 두려움, 인지적 통제 불능에 대한 두려움)에서의 점수가 공황장애 집단과 정상 집단을 유의하게 변별해주는 것으로 알려져 있다(김지혜 등, 2004). 또한 우울 증상의 공존 여부를 탐지하기 위하여 BDI-II(Beck Depression Inventory-II; Beck, Steer, & Brown, 1996)가 함께 사용될 수 있다.

## 우울장애의 근거기반 심리평가

우울장애는 정신장애 중 단일 장애로서 유병율이 가장 높으며, 타 장애와의 공병율도 높은 장애로, 임상 장면에서 자주 의뢰되는 문제 중 하나이다. 우울장애의 근거기반 심리평가의 가이드라인으로 Joiner, Walker, Pettit, Perez와 Cukrowicz(2005)는 (1) 주요 우울증의 증상을 충분히 포괄하고, (2) 특히 중요성을 갖는 우울한 기분, 무쾌감(anhedonia), 자살의 문제를 포함하며, (3) 증상의 변화를 탐지할 수 있고, (4) 자살과 관련해서는 단순한 자살 욕구나 생각과 단호한 자살 계획이나 준비를 구분하여 파악하고, (5) 비전형적(atypical) 하위 유형 그리고 가능하다면 멜랑콜릭(melancholic) 하위유형을 평가할 것을 제안하였다. 이들은 위와 같은 목적을 달성하기 위하여, 결론적으로 구조화된 임상적 면접을 사용하되, 증상 심각도에 있어서는 타당화가 잘 되어 있는 자기보고식 척도들을 보충적으로 사용하는 방식을 권고하고 있다.

우선 진단을 위해서는 SCID(한오수, 홍진표, 2000; First et al., 1996)나 MINI(MINI International Neuropsychiatric Interview; 유상우 등, 2006; Sheehan et al., 1997) 등의 (반)구조화된 진단적 면접을 사용할 것이 권장된다. 시간과 노력이 많이 들지만, 장애에 대한 상세한 기술, 비전형적 양상, 경과, 만성 정도 등의 정보를 포함한다는 점에서 전반적으로 SCID의 장점이 더 많은 것으로 알려져 있다(Joiner et al., 2005). 대신 MINI는 실시 시간이 평균 15분으로 더 짧고, 좀 더 구조화된 형식을 갖추고 있어 임상가들뿐만 아니라 훈련받은 비임상가들도 실시할 수 있다는 장점이 있다(Lecrubier et al., 1997; Pinninti, Madison,

Musser, & Rissmiller, 2003).

우울의 증상 심각도를 평가할 수 있는 도구 들로는 크게 자기보고식 척도들과 임상가 평정 척도가 있다. 우선 자주 사용되며 심리측정적 속성이 비교적 잘 확립된 자기보고식 척도들로는 BDI-II(Beck et al., 1996), CES-D(Center for Epidemiologic Studies Depression Scale; 전경구, 최상진, 양병찬, 2001; Radloff, 1997), IDD (Inventory to Diagnose Depression: Zimmerman & Coryell, 1987)가 있다. 이 중에서 국내에서 주로 사용되는 척도는 CES-D와 BDI-II이다. CES-D는 본래 일반 인구에서 역학 연구를 위하여 개발된 척도로 무료이며 대규모 실사가 용이하여 국내외의 역학 연구에 다수 사용되어 왔다(박준혁, 김기웅, 2011; 전경구 등, 2001; Cho & Kim, 1998; Cho, Nam, & Suh, 1998; Kim et al., 2007). 한편, BDI-II는 DSM-IV 진단기준에 잘 부합하도록 기간과 문항이 수정되었고, 수면과 섭식 상의 변화에서 비전형적 양상과 자살 항목을 포함하는 장점이 있으며, Joiner 등(2005)은 외래에서 증상 심각도 및 변화를 탐색하기 위한 최고의 척도로 추천하고 있다. 단, BDI-II를 진단을 위해 사용하거나 임의적 절단점으로 진단 범주화하는 것은 부적절한 것으로 지적하고 있다. 국내에서는 그동안 원판 BDI(Beck, Ward, Mendelson, Mock, & Erbaugh, 1961)를 주로 사용해 왔으나, 최근 몇 년 간 BDI-II를 국문으로 번안하여 신뢰도와 타당도를 확인하고자 하는 연구들이 출판되어 신뢰도와 타당도에 대한 근거들이 축적되고 있다(김명식, 이임순, 이창선, 2007; 성형모 등, 2008; 송열매, 이해경, 김준원, 이건설, 2012; 유병관, 이해경, 이건설, 2011; 임선영 등, 2011; 송열매, 이해경, 김준원, 이건설, 2012). 또한 원저자로부터 저작권을 득하고, 연령, 성

별, 학력, 지역을 고려하여 층화 표집된 대규모 표본을 대상으로 한 정식 타당화 연구가 진행되었다(김지혜, 이은호, 황순택, 홍상황, 2015; Lim, Lee, Whang, Hong, & Kim, 2014).

한편, 우울 증상의 심각도를 평가하기 위한 임상가 평정 척도로 대표적인 것은 Hamilton 우울 평정 척도(Hamilton Rating Scale for Depression; 이중서 등, 2005; Hamilton, 1967)이다. Hamilton 우울 평정 척도는 여러 가지 버전 중 17개 문항 버전이 가장 널리 사용되고 있으며, 훈련받은 평가자가 각 항목에 대한 평가 기준에 따라 임상적 판단을 내리게 되어 있다. Hamilton 우울 평정 척도는 임상 시험 및 약물 치료에 대한 반응 연구에서 치료 반응과 관해를 판단할 때 표준적으로 사용되는 도구로, 일반적으로 17점 이상을 경도, 20-24점을 중등도, 25점 이상을 심각한 우울 증상이 있는 것으로 판단하며, 7점 이하를 관해, 8-16점을 부분 관해로 간주한다(Shelton et al., 2007). Hamilton 우울 평정 척도의 경우에도 역시 우울장애로 이미 진단된 환자의 증상 심각도를 양적으로 평가하기 위한 것으로, 진단적 도구가 아님을 유의할 필요가 있다. 임상가 평정 방식의 Hamilton 우울 평정 척도는 자기보고식 척도에 비해 치료 변화에 더 민감하며, 연령, 학력 수준, 성별과 같은 환자 특성에 따라 점수가 달라질 수 있는 자기보고식 척도에 비하여 우울 심각도를 더 객관적으로 평가할 수 있는 장점이 있다(이중서 등, 2005).

그러나 양극성(bipolarity)과 관련된 문항이 포함되어 있지 않은 것은 Hamilton 우울 평정 척도의 중요한 한계로 지적되어 왔다(Sung et al., 2012). SCID를 사용하여 동일하게 주요우울삽화로 진단된 경우일지라도, 제2형 양극성 장애 환자 집단은 주요우울장애 환자 집단에

비하여 Hamilton 우울 평정 척도에서 평균 3점 가량 더 낮은 점수를 얻는 것으로 밝혀졌다 (Sung, Kim, Lee, Yu, Hong, & Kim, 2012). 따라서 양극성 우울증의 경우에는 BDRS(Bipolar Depression Rating Scale; Berk et al., 2007), MADRS(Montgomery- Asberg Depression Rating Scale; 안용민 등, 2005; Montgomery & Asberg, 1979) 등 이러한 문제를 보완한 다른 임상가 평정 척도를 대안으로 고려해야 할 필요성이 있다. 이러한 임상가 평정의 실시가 가능하지 않은 경우에는 양극성 장애를 선별하기 위해 개발된 자기보고식 척도로 MDQ(Mood Disorder Questionnaire; 전덕인 등, 2005; Hirshfeld et al., 2000) 또는 HCL-32 (Hypomania Symptom Checklist-32; 오미영 등, 2009; Angst et al., 2005)를 사용하여 절단점 이상의 점수를 얻은 환자들에 대하여 양극성 장애 병력 또는 추후 발병 가능성에 대한 추가적 면담을 실시하는 것이 도움이 될 것이다.

마지막으로 우울장애의 근거기반 심리평가

에서 중요하게 다루어져야 할 문제는 자살이다. 우울증은 정신장애 중 비교적 흔하지만, 자살을 초래할 수 있다는 점에서 치명적이며 (권석만, 2013), 따라서 자살사고 및 자살행동 가능성에 대한 민감하고 철저한 평가가 요망된다. 국내에서 자살 관련 척도들에 대한 고찰에 따르면, 자살 예측 타당도의 측면에서 SSI(Scale for Suicide Ideation; 신민섭, 박광배, 오경자, 김중술, 1990; Beck, Kovacs, & Weissman, 1979), BHS(Beck Hopelessness Scale; 김지혜 등, 2015; 신민섭 등, 1990; Beck & Steer, 1988) 등이 신뢰할 만한 척도로 제안된 바 있다(박승진, 임아영, 박수빈, 나리지, 홍진표, 2013). 또한 앞서 소개된 도구 중에서는 BDI-II와 Hamilton 우울 척도가 자살 관련 문항을 포함하고 있으며, 단 1개의 자살 문항이기는 하지만 해당 문항에서의 응답이 2점 이상일 경우는 자살률이 각기 6.9배, 4.9배 증가한다는 결과들을 고려할 때, 이를 잘 활용하는 것의 유용성을 시사한다고 볼 수 있다(Brown, 2001;

표 3. 우울장애의 근거기반 평가 도구

영역	평가 도구	도구의 유형
진단	SCID-IV(First et al., 1996; 한오수, 홍진표, 2000) MINI(Sheeahan et al., 1997; 유상우 등, 2006)	반구조화된 면접 구조화된 면접
증상 심각도	BDI-II(Beck et al., 1996; 김명식 등, 2007; 성형모 등, 2008; 김지혜, 이은호, 황순택, 홍상황, 2015) CES-D(Radloff, 1997; 전경구 등, 2001) IDD(Zimmerman & Coryell, 1987) Hamilton 우울 평정 척도(Hamilton, 1967; 이중서 등, 2005)	자기보고식 척도 자기보고식 척도 자기보고식 척도 임상가 평정
양극성 장애 선별	MDQ(Hirshfeld et al., 2000; 전덕인 등, 2005) HCL-32(Angst et al., 2005; 오미영 등, 2009)	자기보고식 척도 자기보고식 척도
자살	SSI(Beck et al., 1979; 신민섭 등, 1990) BHS(Beck & Steer, 1988; 신민섭 등, 1990; 김지혜 등, 2015)	자기보고식 척도 자기보고식 척도

박승진 등, 2013에서 재인용, p. 8).

## 논 의

본 개관 논문에서는 근거기반 심리평가가 출현하게 된 배경 및 근거기반 심리평가의 정의와 기본 개념에 대하여 소개하였다. 요약하자면, 근거기반 심리평가는 특정한 평가 목적을 염두에 두는 임상적 의사결정 과정으로, 무엇을(구성개념), 무엇으로(평가도구), 어떻게(평가 과정) 평가할지를 이론과 연구 결과에 기초하는 것이다(Hunsley & Mash, 2005; 2007). 근거기반 심리평가는 심리학의 근거기반실천의 중요한 요소이지만, 근거기반 심리치료에 비하여 충분한 관심을 받지 못해 왔으며, 아직까지 그 발전은 초기 단계에 있는 것으로 보인다. 또한 최근의 연구에 따르면, 임상가들이 아직까지는 연구 결과보다 직관에 의존하거나, 표준화된 평가 도구보다 신뢰도가 낮은 비구조화된 임상 면담을 주로 사용하는 경향이 있는 것으로 나타났다(Jensen-Doss & Hawley, 2010).

### 근거기반 평가의 국내 현황

심리평가는 한국 임상심리학의 태동기부터 임상심리학자의 주요 역할 중 하나로 여겨져 왔다(염태호, 1996). 최근 자료들을 참조하면, 임상심리전문가들이 현재 관여하는 활동으로 심리진단 및 평가가 33.0%로 직업적 활동 중 2위를 차지하였고, 심리진단 및 평가에 50% 이상의 시간을 보내는 임상심리전문가는 26.0%로 집계되었다(권정혜, 2008). 또한 1967년부터 2013년까지 46년간 한국심리학

회지: 임상에 게재된 총 1154 편의 논문 중 일관되게 높은 게재 빈도수를 유지한 항목도 심리검사의 개발과 타당화였다(이혜림, 김수현, 박은혜, 김나현, 최승원, 2014).

한편, 근거기반 심리평가가 세계적으로도 아직 초기 단계임을 감안할지라도, 국내의 경우에는 그와 관련된 학문적 논의와 담론이 거의 전무한 상태이다. 또한 현재 국내에서의 심리평가 교육과 실무는 전통적인 풀 배터리 방식의 종합검사의 실시와 해석이 주류를 이루고 있으며, 근거기반 심리평가는 대부분의 임상가들에게 낯선 주제이다. 이러한 상황은 심리평가와 관련된 국내 논문들이 주로 MMPI, K-WAIS, Rorschach 등 풀 배터리에 포함되는 심리검사 도구들에 대한 것이라는 연구동향에도 반영된다(이혜림 등, 2014). 이러한 논문들은 주로 개별 평가도구의 신뢰도와 타당도를 보고하기 위한 목적에서 진행된 연구들이며, 심리평가의 타당도와 유용성에 대한 상위 인지적인 비판적 고찰은 Rorschach와 MMPI의 진단적 예측 불일치 문제에 대한 김중술, 홍강의, 조수철, 신민섭(2000)의 논문 정도가 예외적인 사례이다. 더욱이 이혜림 등(2014)이 지적하듯, 다수의 평가 연구들은 특정 도구를 소개하는 데 그칠 뿐 새롭게 소개된 도구의 임상적 유용성을 입증할 후속 연구가 보고되고 있지 않아, 임상 현장에서의 경험이 연구 논문의 형태로 환류되어 축적되지 못하고 있다는 한계가 있으며, 국내 근거기반 실천의 발전을 위하여 해결되어야 할 문제라고 할 수 있다.

이러한 연유로 현재 근거기반 심리평가에 관하여 참고할 수 있는 국내 문헌은 극히 드물다. 근거기반 심리평가의 상위 개념인 근거기반실천과 관련된 첫 개관 논문이 국내에서

게재된 것은 2013년에 이르러서였으며(임민경 등, 2013), 국내의 근거기반실천에 대하여 존재하는 소수의 논의는 대부분 심리치료에 중점을 두었다. 비록 한국 임상심리 대학원의 교육훈련모델의 방향성과 관련하여, 근거기반 심리평가의 중요성이 대두될 것이라는 전망이 최근에 제안되었던 바 있으나(임혜진, 전은경, 윤희정, 최승원, 2013), 현 시점에서 근거기반 심리평가와 직접적으로 관련된 국내 논문은 단 1편이 발표되어 있을 뿐이다. Ebesutani와 신수현(2014)은 국내 정신건강 분야 종사자들에게 근거기반 심리평가 및 심리치료에 대한 태도, 지식수준과 사용 정도를 조사하였다. 연구 결과, 구조화된 진단적 면접은 모든 집단에서 가장 적게 사용되는 평가 방법으로 나타났으며, 국내에서 근거기반 심리평가의 보급이 매우 저조한 수준임을 보여주었다. 이들의 논문은 근거기반 심리평가에 대한 전문가 집단의 태도 및 근거기반 심리평가 활용 현황을 국내에서 처음으로 제시하였다는 의의가 있으나, 근거기반 심리평가의 기본 개념이 무엇이며 국내 상황에서 어떻게 적용될 수 있는가에 관해서는 직접적인 답을 제공하지 못하였으며 이를 후속 연구들의 몫으로 남겨 놓았다.

#### 근거기반 평가의 국내 보급을 위한 논의 필요성

이와 같은 상황에서 국내에서 근거기반 심리평가가 보급되고 확산되기 위해서는 근거기반 심리평가의 필요성과 국내 적용 가능성에 대한 토론과 논의의 장이 우선 마련되어야 할 것으로 보인다. 이는 본질적으로 소수 의견의 개진에 의해 결정내릴 수 있는 문제가 아니며, 심리평가를 실시하고 연구하는 많은 국내 전

문가들의 토론을 통하여 중지(衆智)를 모아야 할 쟁점이라고 생각된다. 따라서 본고의 목적은 해답을 제시하기 위한 것이라기보다는 그러한 토론의 마중물을 제공하고자 하는 것이다. 다만 저자들의 사견으로는 이 토론의 목표가 반드시 양자택일에 있는 것은 아닐 것이라는 생각이 든다. 이는 그간 시행되어 온 전통적인 심리평가 접근을 전면적으로 폐기하기 위한 것이라기보다는, 각 접근이 지니고 있는 장단점을 비교하고, 양자를 통합 또는 어떠한 상황에서 어떠한 부분을 취사선택할 것인가 하는 문제에 대한 국내 전문가들의 활발한 의견개진과 공감대 형성이 필요하다는 것이다. 신경심리평가 내에서 점화되었던 고정된(fixed) 대 유연한(flexible) 배터리 논쟁에서 Kane(1991)이 지적하듯, 각 평가 방식은 나름의 장단점을 지니고 있으며 지지하는 이론적 입장에 따라 평가의 개념과 목적은 달라질 수 있다. 문제는 어떠한 하나의 방식만이 옳다고 간주하는 선입견일 수 있으며, 어떠한 방식의 심리평가가 타당하며 유용한가에 대한 논쟁은 그 자체로 심리평가의 정교화와 발전을 위하여 유의한 과정이라고 생각된다.

현재까지 선행 연구들이 공통적으로 지적하는 점은 평가 의뢰 목적과 여타 임상적 유용성에 대한 고려가 선행되지 않은 상황에서 관성적으로 모든 환자와 내담자들에게 동일하게 풀 배터리의 검사를 실시하는 것이 타당하지도 유용하지도 않다는 것이다. 따라서 무엇을, 무엇으로, 어떻게 평가할지에 따라 필요할 것으로 판단되는 접근을 유연하게 적용할 필요가 있을 것으로 사료된다. 물론 이러한 판단을 위해서는 평가자의 전문성이 더욱 요구되며, 풀 배터리에 포함되는 검사뿐만 아니라 그 외의 선택지를 가용할 수 있는 훈련이 필

요하다.

더불어 이 시점에서 한 가지 재고해야 할 것은 ‘폴 배터리’라는 용어 자체이다. 폴 배터리는 용어는 우리에게 상당히 익숙한 나머지 마치 지능 검사와 MMPI, 그리고 Rorschach 등 투사적 검사의 조합을 일컫는 고유명사처럼 사용되어 오고 있다. 사실 배터리는 해석을 위하여 두 개 이상의 검사들을 연합하여 함께 사용한다는 의미이며(Russell, Russell, & Hill, 2005), 단일 검사 도구만을 사용하였을 때 얻을 수 없는, 검사들 간의 조합과 패턴으로부터 도출되는 정보들을 제공한다는 점에서 장점을 갖고 있다. 본고에 상술된 근거기반 심리평가의 예시 또한 다중 평가 도구로부터 수렴 증거를 추구한다는 점에서 일종의 배터리 방식을 따른다고 볼 수 있다. 다만 의뢰되는 문제에 따라 배터리에 무엇을 포함할지가 달라져야 하며, 그 구성이 고정되어 있는 것이 아니라 최신의 이론과 연구 결과에 근거하여 갱신되어야 한다는 점을 강조한다는 점에서 차이가 있다. 폴 배터리이든 근거기반 심리평가에 따른 배터리이든, 배터리 검사 구성의 타당성과 유용성에 대해서는 끊임없는 검증이 필요하며, 각 검사도구의 이론적 근거에 대한 검증 절차 및 임상적 유용성에 대한 근거를 제시할 수 있는 경험적 연구가 계속 누적되어야 할 것이다.

두 가지 평가 전통을 임상 현장에서 절충하는 한 가지 방안은 의뢰 목적과 문제에 따라 다양한 심리평가 프로토콜을 마련하는 것이다. 만일 진단만을 요하는 사례라면, 본고에 소개된 진단과 선별평가 도구만으로도 제한된 시간 내에 필요로 하는 정보를 제공할 수 있을 것이다. 이를 위해 각 기관에서 자주 의뢰되는 장애 및 문제의 비율을 파악하여, 장애별

평가 가이드라인을 합의해 둘 필요가 있다. 그러나 환자/내담자의 특성에 따라 추가적인 평가가 도움이 되는 사례들 또한 존재하며, 정서 및 성격에 대한 심층 평가가 필요한 경우에는 MMPI, Rorschach 등을 진단적 도구와 적절히 조합한 일종의 폴 배터리 검사를 사용할 수 있다.<sup>3)</sup> 또는 진단 외에 인지 및 신경심

3) 여기서 전통적인 심리평가 도구들은 진단적 용도로 사용된다기보다는 한 개인의 인지, 정서 및 성격에 대한 다양하고 개별 사례적(idiographic) 정보를 얻기 위해 사용될 수 있으며, 법칙 발견적(nomothetic) 측정에서 얻어진 정보들을 개인 수준에서 개념화하거나 또는 보완하는 기능을 할 수 있을 것으로 생각된다(김중술 등, 2000; Stricker & Gold, 1999). 김중술 등(2000, p. 396)의 ‘치료에 반응이 저조한 우울증 환자’의 사례를 들어 보자. 이 환자를 평가할 때, SCID 등 반구조화된 진단적 면접을 실시하여 정신과적 진단을 확진함으로써, 이 환자가 통상적 치료에 대한 반응성이 떨어지는 것이 불안 등 공존 병리의 존재 여부 때문인지 또는 단극성 우울증이 아닌 양극성 우울증에 속하여 항우울제 치료에 반응성이 저조한 것인지 등의 가능성을 탐색할 수 있다. 또한 HDRS나 BDI-II는 현재 증상 심각도가 어떠한 수준인지를 파악하고, 추후 치료 반응을 추적 관찰할 때 반복 실시하여 증상의 증감, 관해 여부를 판단하는 데 유용하게 사용될 수 있다. 그런데 이 환자가 Rorschach에서 Depression Index는 상승하지 않았으며, 여러 개의 반사 반응을 나타내어 Egocentricity Index가 상승해 있고, 대인관계 관련 내용 척도들에서 그에 조용하는 특기할만한 반응을 나타낸다면? 이러한 개별 사례적 자료들은 앞서 법칙 발견적 자료를 통하여 확인된 진단명과 우울 증상들이 이 개인에게 어떤 의미인가, 그 저변에 존재하며 우울 증상에 영향을 주는 것들이 무엇인가에 대한 잠정적 가설을 세우는데 도움을 줄 수 있다. 일례로 김중술 등(2000)이 기술하듯 이러한 경우 우울 증상 기저에 있는 자아도취적 성격, 자기애적 상처 또

리적 기능에 대한 평가가 요망되는 경우에는 WAIS와 신경심리평가 도구들을 활용할 수 있다. 그리고 어떠한 평가 프로토콜을 활용하든, 검사 자료 외에 과거력, 면담, 행동 관찰 및 다수 원천으로부터의 자료들을 이론적 틀 속에서 통합하는 임상적 판단이 중요한 역할을 함은 부동의 원칙이다.

만약 토론과 논의를 통하여 일정한 공감대가 이루어진다면, 그에 따른 근거기반 심리평가의 교육과 실시를 위한 환경의 조성이 필요하다. 심리치료의 측면에서 근거기반실천에 대하여 고찰한 임민경 등(2013)이 한국의 근거기반실천 보급을 위해 제안한 심리학회의 관련 분과와 단체, 대학교들의 협력적인 공동 노력과 관련된 내용은 근거기반 심리평가의 경우에도 유효하게 적용될 것이다. 특히 근거기반 심리평가와 관련하여 다음의 두 가지 점이 강조될 필요가 있을 것이다. 첫째, 사람들이 교육 및 수련을 받은 대로 실무를 하는 경향이 있다는 외국의 설문조사 결과는 국내의 임상 수련 및 대학원 수준의 심리평가 교육 커리큘럼을 재고할 필요성을 제기한다 (Youngstrom, 2013). 둘째, 임상 현장에서 실제로 근거기반 심리평가가 실현되기 위해서는 각 기관 및 장면의 특성에 적합한 근거기반 심리평가의 적용 방안들이 모색되어야 한다. 앞서 언급하였듯, 시간과 비용, 기관 내의 의사결정 체계 및 업무 분담에 관한 인식, 평가 인력의 숙련도 등의 요소에 따라서 가능하고 최적화된 근거기반 심리평가의 형태는 변화될 수 있다.

비록 본 개관 논문은 근거기반 심리평가와 관련된 전 영역을 충분히 포괄하고 있지는 못

는 대인관계 상의 실패 경험들이 우울 증상의 완화를 위해 접근해야 할 타깃이 될 수 있다.

하지만, 저자들이 아는 한도 내에서는 국내에서 근거기반 심리평가를 소개하는 첫 개관 논문으로서의 의의가 있을 것으로 생각된다. 부족하나마 본 논문이 향후 국내에서 근거기반 심리평가에 대한 논의가 공론화되고 활성화될 수 있는 작은 계기가 될 수 있기를 희망한다.

## 참고문헌

- 권석만 (2013). 현대 이상심리학. 서울: 학지사.
- 권정혜 (2008). 한국 임상심리학자들의 역할과 활동: 2007년 조사 보고서. 한국심리학회지: 임상, 27(2), 571-579.
- 김명식, 이임순, 이창선 (2007). 한국판 BDI-II의 타당화 연구 I: 여대생을 대상으로. 한국심리학회지: 임상, 26(4), 997-1014.
- 김정범 (2001). 한국판 공황장애 심각도 척도의 개발. 정신병리학, 10(2), 140-151.
- 김정택 (1978). 특성불안과 사회성과의 관계 - Spielberger의 STAI를 중심으로. 고려대학교 석사학위 논문.
- 김중술, 홍강의, 조수철, 신민섭 (2000). 왜 진단적 심리검사 결과가 일치하지 않는가 - Rorschach와 MMPI의 경우. 한국심리학회지: 임상, 19(3), 393-407.
- 김지은, 이선영 (2011). 불안민감성이 다양한 불안증상에 미치는 영향에 대한 마음챙김의 매개효과 - 일반 성인을 대상으로. 인지행동치료, 11(2), 23-38.
- 김지혜, 유범희, 오강섭, 양종철, 김율리, 이소영, 임영진 (2004). 한국어 확장판 불안 민감도 지수의 타당화 연구. 신경정신의학, 43(1), 54-61.
- 김지혜, 양종철, 김정범, 임기영, 이소영, 유범



- 희 (2004). 한국판 알바니 공황-공포 질 문지 타당화 연구. *신경정신의학*, 43(3), 329-336.
- 김지혜, 이은호, 황순택, 홍상황 (2015). 한국판 백우울척도 2판 지침서. 대구: 한국심리주식회사.
- 박승진, 임아영, 박수빈, 나리지, 홍진표 (2013). 자살성향 측정척도들의 자살예측력에 대한 고찰. *대한불안학회지*, 9(1), 10-18.
- 박준영, 김지혜 (2010). 한국판 Sheehan Disability Scale의 신뢰도와 타당도 연구. *한국심리학회지: 임상*, 29(1), 73-81.
- 박준혁, 김기웅 (2011). 한국의 우울증 역학에 대한 고찰. *대한의사협회지*, 54(4), 362-369.
- 박현순, 이창인, 김영철, 김종원 (1997). 공황장애 환자의 임상양상에 관한 조사연구 - 인지특성을 중심으로. *한국심리학회지: 임상*, 16(2), 1-13.
- 박혜원, 이은호, 김도관, 유범희, 이동수, 김지혜 (2009). 한국판 해밀턴 우울증평가척도 (K-HDRS)의 확인적 요인 분석. *신경정신의학*, 48, 21-28.
- 보건복지부 (2011). 2011년도 정신질환실태 역학조사 요약보고서.
- 성형모, 김정범, 박영남, 배대석, 이선희, 안현희 (2008). 한국어판 백 우울 설문지 2판의 신뢰도 및 타당도 연구. *생물치료정신의학*, 14(2), 201-212.
- 송열매, 이혜경, 김준원, 이건설 (2012). 인터넷으로 시행한 백 우울 척도 2판의 신뢰도와 타당도: 대학생을 대상으로. *신경정신의학*, 51, 402-408.
- 신민섭, 박광배, 오경자, 김중술 (1990). 고등학생의 자살성향에 대한 연구: 우울-절망-자살간의 구조적 관계에 대한 분석. *한국심리학회지: 임상*, 9(1), 1-19.
- 안용민, 이규영, 이중서, 강민희, 김도훈, 김정란, ..., 김용식 (2005). 한글판 Montgomery-Asberg 우울증 평가 척도의 표준화 연구. *신경정신의학*, 44(4), 466-476.
- 안현의 (2003). 과학자-실무자 모델의 국내 적용 가능성에 관한 논의. *한국심리학회지: 상담 및 심리치료*, 15(3), 461-475.
- 안형식 (2000). Evidence-based Healthcare의 개념과 발전방안. *한국의료QA학회: 가을학술대회*, 611-626.
- Ebesutani, C., 신수현 (2014). Knowledge, attitudes, and usage of evidence-based assessment and treatment practices in the Korean mental health system: current status and future directions. *한국심리학회지: 임상*, 33(4), 891-917.
- 염태호 (1996). *임상심리학회 30년사*. 한국임상심리학회. 서울: 하나의학사.
- 오미영, Angst, 성태훈, 이은호, 홍경수, 이동수, 김지혜 (2009). 한국판 HCL-32의 타당화. *한국심리학회지: 임상*, 28(2), 321-338.
- 원호택, 박현순, 권석만 (1995). 한국판 공황장애 척도 개발 연구. *한국심리학회지: 임상*, 14(1), 95-110.
- 유병관, 이혜경, 이건설 (2011). 한국어판 백 우울 설문지 2판의 신뢰도 및 요인분석: 대학생을 대상으로. *생물정신의학*, 18, 126-133.
- 유상우, 김영신, 노주선, 오강섭, 김찬형, 남궁기, ..., 김세주 (2006). 한국판 Mini International Neuropsychiatric Interview 타당도 연구. *대한불안학회지*, 2(1), 50-55.

- 이중서, 배승오, 안용민, 박두병, 노경선, 신현균, ..., 김용식 (2005). 한국판 Hamilton 우울증 평가 척도의 신뢰도, 타당도 연구. *신경정신의학*: 44, 456-465.
- 이지현, 장정기, 박지홍, 안준호, 이 철, 김창윤 (2003). 해밀턴 우울증 평가척도의 요인구조. *정신병리학*, 12, 3-16.
- 이혜림, 김수현, 박은혜, 김나현, 최승원 (2014). 한국 임상 심리학의 연구 동향: 1967~2013년 ‘한국심리학회지: 임상’ 학회지 게재 논문 분석. *한국심리학회지: 임상*, 33(4), 919-942.
- 이효윤, 이인혜, 성주현 (2012). 성인 쌍둥이 및 쌍둥이 가족을 대상으로 한 불안민감성의 유전적 영향력에 관한 연구: 우울, 음주, 흡연과의 관계. *한국심리학회지: 건강*, 17(1), 143-163.
- Lim, S. U., Lee, E. H., Whang, S. T., Hong, S. H., & Kim, J. H. (2014). Psychometric properties of the Beck Depression Inventory-II in Korea. 14' 한국임상심리학회 추계학술대회 포스터 발표.
- 임선영, 이은정, 정성원, 김희철, 정철호, 전태연, ..., 김정범 (2011). 한국판 백 우울 척도 2판의 타당화 연구. *대한불안학회지*, 7(1), 48-53.
- 임민경, 이지혜, 이한나, 김태동, 최기홍 (2013). 근거기반실천과 심리치료. *한국심리학회지: 일반*, 32(1), 251-270.
- 임영진, 이소영, 김지혜 (2005). 불안민감도와 특질불안의 차별성과 공통성: 부적정서, 정적정서, 생리적 과각성 간의 관계. *한국심리학회지: 임상*, 24(2), 439-449.
- 임혜진, 전은경, 윤효정, 최승원 (2013). 한국 임상심리대학원의 교육훈련 프로그램 연구 - 전문가 집단을 중심으로. 13' 한국심리학회 연차학술발표대회 논문집, 166.
- 전겸구, 최상진, 양병찬 (2001). 통합적 한국판 CES-D 개발. *한국심리학회지: 건강*, 6(1), 59-76.
- 전덕인, 윤보현, 정한용, 하규섭, 신영철, 박원명 (2005). 한국형 기분장애 질문지의 타당화 연구. *신경정신의학*: 44, 583-590.
- 조맹제, 김계희 (1993). 주요우울증환자 예비평가에서 the Center for Epidemiological Studies Depression Scale (CES-D)의 진단적 타당성 연구. *신경정신의학*, 32(3), 381- 399.
- 최명애, 방경숙, 박연환, 강현주 (2011). 국내 근거기반간호의 현황과 발전 방향. *간호학의 지평*, 8(2), 129-138.
- 최영희, 김경미, 박기환, 윤혜영 (2003). 공황장애 환자에서 인지행동치료의 약물 감량 효과. *대한정신약물학회지*, 14(4), 367-376.
- 최영희, 김지혜, 임영진, 채정호, 우종민, 류한옥, 유범희 (2004). 공황장애에서 단기약물 치료가 불안민감도에 미치는 영향. *대한정신약물학회지*, 15(3), 333-338.
- 최영희, 최윤정, 우종민, 윤혜영 (2006). 광장공포증 유무에 따른 공황장애 환자의 치료 결과 비교. *인지행동치료*, 6(2), 163-178.
- 한오수, 안준호, 송선희, 조맹제, 김장규, 배재남, ..., 홍진표 (2000). 한국어 판 구조화 임상면담도구 개발: 신뢰도 연구. *신경정신의학*, 39(2), 362-372.
- 한오수, 홍진표 (2000). DSM-IV의 제1축 장애의 구조화된 임상적 면담. 서울: 하나의학사.
- Angst, J., Adolfsson, R., Benazzi, F., Hantouche, E., Meyer, T., Sjeppar, P., ... & Scott, J. (2005). The HCL-32: towards a

- self-assessment tool for hypomanic symptoms in outpatients. *Journal of Affective Disorders*, 88, 217-233.
- Antony, M. M., & Rowa, K. (2005). Evidence-based assessment of anxiety disorders in adults. *Psychological Assessment*, 17(3), 256-266.
- American Psychiatric Association. (1980). *Diagnostic and Statistical Manual of Mental Disorders-3rd edition*. Washington, DC: Author.
- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders-5th edition(DSM-5)*. Washington, DC: Author.
- American Psychological Association. (1995). *Template for developing guidelines: Interventions for mental disorders and psychological aspects of physical disorders*. Washington, DC: APA.
- American Psychological Association. (2005). *Report of the 2005 Presidential Task Force on evidence-based practice*. Washington, DC: APA.
- American Psychological Association. (2006). Evidence-based practice in psychology. *American Psychologist*, 61(4), 271-285.
- Barlow, D. H. (2005). What's new about evidence-based assessment? *Psychological Assessment*, 17(3), 308-311.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4, 561-571.
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory Manual* (2nd ed.) New York: Guilford Press.
- Beck, A. T., Kovacs, M., & Weissman, A. (1979). Assessment of suicidal ideation: the scale for suicide ideation. *Journal of Consulting and Clinical Psychology*, 47(2), 343-352.
- Berk, M., Malhi, G. S., Cahill, C., Carman, A. C., Hadzi Pavlovic, D., Hawkins, M. T., ... & Mitchell, P. B. (2007). The Bipolar Depression Rating Scale (BDRS): its development, validation and utility. *Bipolar Disorders*, 9(6), 571-579.
- Brown, G. K. (2001). *A review of suicide assessment measures for intervention research with adults and older adults*. USA National Institute of Mental Health.
- Brown, T. A., Campbell, L. A., Lehman, C. L., Grisham, R. J., & Mancill, R. B. (2001). Current and lifetime comorbidity of the DSM-IV anxiety and mood disorders in a large clinical sample. *Journal of Abnormal Psychology*, 110, 585-599.
- Camara, W. J., Nathan, J. S., & Puente, A. E. (2000). Psychological test usage: implications in professional psychology. *Professional Psychology: Research and Practice*, 31, 141-154.
- Cashel, M. L. (2002). Child and adolescent psychological assessment: current clinical practices and the impact of managed care. *Professional Psychology: Research and Practice*, 33, 446-453.
- Chambless, D. L., Caputo, G. C., Bright, P., & Gallagher, R. (1984). Assessment of fear of fear in agoraphobics: The Body Sensation Questionnaire and Agoraphobic Cognition Questionnaire. *Journal of Consulting and Clinical Psychology*, 52(6), 1090-1097.
- Charter, R. A. (2003). A breakdown of reliability coefficients by test type and reliability

- method, and the clinical implications of low reliability. *Journal of General Psychology*, 130, 290-304.
- Childs, R. A., & Eyde, L. D. (2002). Assessment training in clinical psychology doctoral programs. *Journal of Personality Assessment*, 78, 130-144.
- Claridge, G. (1995). *Origins of Mental Illness*. Cambridge, MA: Malor Books.
- 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(5), 304-310.
- Cho, M. J., Nam, J. J., & Suh, G. H. (1998). Prevalence of symptoms of depression in a nationwide sample of Korean adults. *Psychiatry Research*, 81, 341-352.
- Di Nardo, P. A., Brown, T. A., & Barlow, D. H. (1994). *Anxiety Disorders Interview Schedule for DSM-IV: Lifetime version*. Boulder, CO: The Psychological Corporation.
- First, M. B., Spitzer, R. L., Gibbon, M., Williams, J. B. W. (1996). *User's Guide for the Structured Clinical Interview for DSM-IV Axis I disorders - Research Version*. New York: Biometric Research.
- Guyatt, G., Cairns, J., Churchill, D., Cook, D., Haynes, B., Hirsh, J., ... & Tugwell, P. (1992). Evidence-based medicine: a new approach to teaching the practice of medicine. *JAMA*, 268(17), 2420-2425.
- Hamilton, M. (1967). Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology*, 6, 278-296.
- Hirschfeld, R. M., Williams, J. B., Spitzer, R. L., Calabrese, J. R., Flynn, L., Keck Jr, P. E., ... & Zajecka, J. (2000). Development and validation of a screening instrument for bipolar spectrum disorder: the Mood Disorder Questionnaire. *American Journal of Psychiatry*, 157(II), 1873-1875.
- Hunsley, J., & Bailey, J. M. (1999). The clinical utility of the Rorschach: unfulfilled promises and an uncertain future. *Psychological Assessment*, 11, 266-277.
- Hunsley, J., & Bailey, J. M. (2001). Whither the Rorschach? An analysis of the evidence. *Psychological Assessment*, 13, 472-485.
- Hunsley, J., & Mash, E. J. (2008). *A Guide to Assessments That Work*. New York: Oxford Univ. Press.
- Hunsley, J., & Mash, E. J. (2007). Evidence-based assessment. *Annual Review of Clinical Psychology*, 3, 29-51.
- Hunsley, J., & Mash, E. J. (2005). Introduction to the special section on developing guidelines for the evidence-based assessment (EBA) of adult disorders. *Psychological Assessment*, 17(3), 251-255.
- Hyman, S. E. (2010). The diagnosis of mental disorders: the problem of reification. *Annual Review of Clinical Psychology*, 6, 155-179.
- Jensen-Doss, A., & Hawley, K. M. (2010). Understanding barriers to evidence-based assessment: clinician attitudes toward standardized assessment tools. *Journal of Clinical Child & Adolescent Psychology*, 39(6), 885-896.
- Joiner, T. E., Walker, R. L., Pettit, J. W., Perez, R. L., & Cukrowicz, K. C. (2005).

- Evidence-based assessment of depression in adults. *Psychological Assessment*, 17(3), 267-277.
- Kane, R. L. (1991). Standardized and flexible batteries in neuropsychology: an assessment update. *Neuropsychology Review*, 2, 281-339.
- Keller, M. L., & Craske, M. G. (2008). Panic Disorder and Agoraphobia. In J. Hunsley & E. J. Mash (Eds.), *A Guide to Assessments That Work*. (pp.229-253) New York: Oxford Univ. Press.
- Kim, M-D., Hong, S-C., Lee, C-I., Kwak, Y-S., Shin, T-K., Jang, Y-H, ... & Whang, S-E. (2007). Prevalence of depression and correlates of depressive symptoms for residents in the urban part of Jeju Island, Korea. *International Journal of Social Psychiatry*, 53(2), 123-134.
- Kleinman, A. (1988). *Rethinking psychiatry: From cultural category to personal experience*. New York: Free Press.
- Leclercq, Y., Sheehan, D. V., Weiller, E., Amorim, P., Bonora, I., Harnett Sheehan, K., ... & 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(5), 224-231.
- Lee, E-H., Kim, J-H., & Yu, B-H. (2009). Reliability and validity of the self-report version of the Panic Disorder Severity Scale in Korea. *Depression and Anxiety*, 26, E120-E123.
- Lima, E. N., Stanley, S., Kaboski, B., Reitzel, L. R., Richey, A., Castro, Y., ... & Joiner Jr, T. E. (2005). The incremental validity of the MMPI-2: When does therapist access not enhance treatment outcome? *Psychological Assessment*, 17, 462-468.
- McFall, R. M. (2005). Theory and utility - key themes in evidence-based assessment: comment on the special section. *Psychological Assessment*, 17(3), 312-323.
- Merenda, P. F. (2007). Update on the decline in the education and training in psychological measurement and assessment 1. *Psychological Reports*, 101(1), 153-155.
- Meyer, G. J., & Archer, R. P. (2001). The hard science of Rorschach research: What do we know and where do we go? *Psychological Assessment*, 13, 486-502.
- Meyer, G. J., & Handler, L. (1997). The ability of the Rorschach to predict subsequent outcome: A meta-analysis of the Rorschach Prognostic Rating Scale. *Journal of Personality Assessment*, 69(1), 1-38.
- Miller, P. R. (2001). Inpatient diagnostic assessments: 2. Interrater reliability and outcomes of structured vs. unstructured interviews. *Psychiatry Research*, 105(3), 265-271.
- Montgomery, S. A., & Asberg, M. (1979). A new depression scale designed to be sensitive to change. *The British Journal of Psychiatry*, 134, 382-389.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory*. (3rd ed.) New York: McGraw-Hill.
- Persons, J., & Fresco, D. M. (2008) Adult Depression. In J. Hunsley & E. J. Mash (Eds.), *A Guide to Assessments That Work*. (pp. 96-120) New York: Oxford Univ. Press.
- Pinninti, N. R., Madison, H., Musser, E., & Rissmiller, D. (2003). MINI International

- Neuropsychiatric Schedule: clinical utility and patient acceptance. *European Psychiatry*, 18, 361-364.
- Piotrowski, C. (1999). Assessment practices in the era of managed care: current status and future directions. *Journal of Clinical Psychology*, 55, 787-796.
- Radloff, L. S. (1997). The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1, 385-401.
- Regier, D. A., Narrow, W. E., Kuhl, E. A., & Kupfer, D. J. (2009). The conceptual development of DSM-V. *The American Journal of Psychiatry*, 166(6), 645-650.
- Russell, E. W., Russell, S. L. K., & Hill, B. D. (2005). The fundamental psychometric status of neuropsychological batteries. *Archives of Clinical Neuropsychology*, 20, 785-794.
- Sackett, D. L., Rosenberg, W. M., Gray, J. A. M., Haynes, R. B., Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't. *British Medical Journal*, 312(7023), 71-72.
- Shakow, D., Hilgard, E. R., Kelly, E. L., Luckey, B., Sanford, R. N., & Shaffer, L. F. (1947). Recommended graduate training program in clinical psychology. *American Psychologist*, 2, 539-558.
- Sheehan, D. V., Lecrubier, Y., Harnett Sheehan, K., Janavs, J., Weiller, E., Keskiner, A., ... & Dunbar, G. C. (1997). The validity of the Mini International Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability. *European Psychiatry*, 12(5), 232-241.
- Shear, M. K., & Maser, J. D. (1994). Standardized assessment for panic disorder research: A conference report. *Archives of General Psychiatry*, 51, 346-354.
- Shear, M. K., Brown, T. A., Barlow, D. H., Money, R., Sholomskas, D. E., Woods, S. W., ... & Papp, L. A. (1997). Multicenter collaborative panic disorder severity scale. *American Journal of Psychiatry*, 154(11), 1571-1575.
- Shear, M. K., Rucci, P., Williams, J., Frank, E., Grochocinski, V., Vander Bilt, J., ... & Wang, T. (2001). Reliability and validity of the Panic Disorder Severity Scale: replication and extension. *Journal of Psychiatric Research*, 35(5), 293-296.
- Shelton, R. C., Prakash, A., Mallinckrodt, C. H., Wohlreich, M. M., Raskin, J., Robinson, M. J., & Detke, M. J. (2007). Patterns of depressive symptom response in duloxetine treated outpatients with mild, moderate or more severe depression. *International Journal of Clinical Practice*, 61(8), 1337-1348.
- Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *Manual for the state-trait anxiety inventory*. Palo Alto, CA: Consulting Psychologists Press, Inc.
- Spring, B., & Neville, K. (2010). Evidence-based Practice in Clinical Psychology. In D. H. Barlow (Ed.), *The Oxford handbook of clinical psychology*. Oxford Univ. Press.
- Steiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, 80, 99-103.

- Stricker, G., & Gold, J. R. (1999). The Rorschach towards a nomothetically based, idiographically applicable configural model. *Psychological Assessment, 11*, 240-250.
- Sung, G., Kim, B-N., Lee, E-H., Yu, B-H., Hong, K. S., Kim, J-H. (2012). Underestimating the severity of bipolar depression: a comparison of the Hamilton depression rating scale items. *Journal of Affective Disorders, 136*(3), 425-429.
- Vane, J. R. (1981). The Thematic Apperception Test: a review. *Clinical Psychology Review, 1*(3), 19-36.
- White, K. L. (1995). Evidence-based medicine. *The Lancet, 346*(8978), 837-838.
- White, K. L. (1997). Archie Cochrane's legacy: An American perspective. In Maynard A., Chalers, I. (Eds.) *Non-random Reflections on Health Services Research: On the 25th anniversary of Archie Cochrane's Effectiveness and Efficiency*. London: BMJ Publishing Group.
- Wood, J. M., Garb, H. N., Lilienfeld, S. O., & Garb, H. N. (2003). *What's Wrong with the Rorschach?* San Francisco: Jossey-Bass.
- Yamamoto, I., Nakano, Y., Watanabe, N., Noda, Y., Furukawa, T. A., Kanai, T., ... & Kamijima, K. (2004). Cross-cultural evaluation of the Panic Disorder Severity Scale in Japan. *Depression and Anxiety, 20*(1), 17-22.
- Youngstrom, E. A. (2013). Future directions in psychological assessment: combining evidence-based medicine innovations with psychology's historical strengths to enhance utility. *Journal of Child & Adolescent Psychology, 42*(1), 139-159.
- Zimmerman, M., & Coryell, W. (1987). The Inventory to Diagnose Depression (IDD): A self-report scale to diagnose major depressive disorder. *Journal of Consulting and Clinical Psychology, 55*, 55-59.

원고접수일 : 2015. 02. 23.

수정원고접수일 : 2015. 05. 08.

게재결정일 : 2015. 05. 08.

**Basic concepts of evidence-based assessment(EBA) and  
discussion for its application in Korea:  
with examples of panic disorder and depression**

**Bin-Na Kim**

Department of Psychology  
Seoul National University

**Ji-Hae Kim**

Department of Psychiatry  
Samsung Medical Center

Evidence-based assessment(EBA) is defined as psychological assessment that uses the best available evidence drawn from research and theory to guide the selection of constructs to be assessed, methods/measures to be used and the process in which the assessment unfolds (Hunsley & Mash, 2007). EBA is one of the major cornerstones of what constitutes evidence-based practice in psychology(EBPP). However, the emphasis has been on the evidence-based treatment(EBT) component since the beginning of EBPP. Thus, EBA has remained underappreciated and it is a rarely discussed topic especially in Korea. Therefore, our primary goal of this review article is to introduce from which backgrounds EBA has emerged and what is the basic concepts of EBA are. Also, we would like to illustrate how the concepts of EBA can be applied to psychological evaluation of panic disorder and depressive disorder in adults. Lastly, we discussed necessary steps to disseminate and expand EBA in Korea.

*Key words : Evidence-Based Assessment, Evidence-Based Practice, Panic Disorder, Depression, Semistructured Diagnostic Interview, Clinician Rating Scale*



부록 I. 국내에서 사용되는 공황장애 · 우울장애 진단적 선별 평가 도구들의 심리측정적 속성의 평가

해당 장애	도구	기준	내적 일관성	평가자간 신뢰도	검사- 재검사 신뢰도	내용 타당도	구성 타당도	타당도 일반화	임상적 유용성
공황장애	SCID-IV	NA	NA	G/E	NA	A/G	A/U	G/G	A/G
	PDSS	A/A	G/G	G/G	NA	A/A	A/A	G/G	A/G
	ASI-R	G/G	G/G	NA	A/A	A/A	E/E	E/G	A/A
	ACQ	G/G	G/E	NA	A/A	A/A	G/G	G/G	A/A
	APPQ	A/G	G/E	NA	A/A	A/A	A/A	A/A	A/A
우울장애	SCID-IV	NA	NA	G/E	NA	G/G	G/U	G/G	E/G
	MINI	NA	NA	E/U	NA	A/A	G/A	G/A	G/A
	HDRS	A/A	A/A	E/E	NA	G/G	G/G	G/G	E/G
	BDI-II	G/G	E/E	NA	E/G	E/G	E/G	E/E	E/G
	CES-D	E/E	G/E	NA	A/U	G/G	G/G	E/E	G/G

주. SCID-IV=Structured Clinical Interview for DSM-IV, PDSS=Panic Disorder Severity Scale, ASI-R=Anxiety Sensitivity Inventory-Revised, ACQ=Agoraphobic Cognitions Questionnaire, APPQ=, MINI=MINI International Neuropsychiatric Interview. HDRS=Hamilton Depression Rating Scale, BDI-II=Beck Depression Inventory-II, CES-D=Center for Epidemiologic Studies Depression Scale.

평가 기준은 Hunsley와 Mash(2008)의 제안을 따름: A(adequate), G(good), E(excellent). 또한 해당 기준의 평가 자료가 존재하지 않는 경우를 U(unavailable), 해당 기준이 평가가 적용되지 않는 경우를 NA(not applicable)로 표기함. (단, 진단적 평가와 관련된 심리측정적 속성이므로 ‘치료 변화에의 민감도’ 항목은 기준에서 제외하였음) 표에서 /를 기준으로 왼쪽이 외국의 연구들에 의한 결과이며(c.f. Keller & Craske, 2008; Persons & Fresco, 2008), 오른쪽은 RISS 및 Google Scholar 검색을 통한 국내 연구 결과임. 상기 평가 결과는 잠정적인 것으로 향후 추가적인 연구가 진행된다면 달라질 수 있으며, 각 장애별 가이드라인에 관해서는 별도의 심도 있는 고찰이 필요할 것으로 생각됨.

부록 II. 부록 I에서 각 도구의 속성을 평가하기 위하여 검색된 참고문헌

- 김명식, 이임순, 이창선 (2007). 한국판 BDI-II의 타당화 연구 I: 여대생을 대상으로. 한국심리학회지: 임상, 26(4), 997-1014.
- 김정범 (2001). 한국판 공황장애 심각도 척도의 개발. 정신병리학, 10(2), 140-151.
- 김지은, 이선영 (2011). 불안민감성이 다양한 불안증상에 미치는 영향에 대한 마음챙김의 매개효과 - 일반 성인을 대상으로. 인지행동치료, 11(2), 23-38.
- 김지혜, 양종철, 김정범, 임기영, 이소영, 유범희 (2004a). 한국판 알바니 공황-공포 질문지 타당화 연구. 신경정신의학, 43(3), 329-336.
- 김지혜, 유범희, 오강섭, 양종철, 김울리, 이소영, 임영진 (2004b). 한국어 확장판 불안민감도 지수의 타당화 연구. 신경정신의학, 43(1), 54-61.
- 김지혜, 이은호, 황순택, 홍상황 (2015). 한국판 벡우울척도 2판 지침서. 대구: 한국심리주식회사.
- 박준혁, 김기웅 (2011). 한국의 우울증 역학에 대한 고찰. 대한의사협회지, 54(4), 362-369.
- 박현순, 이창인, 김영철, 김종원 (1997). 공황장애 환자의 임상양상에 관한 조사연구 - 인지특성을 중심으로. 한국심리학회지: 임상, 16(2), 1-13.
- 박혜원, 이은호, 김도관, 유범희, 이동수, 김지혜 (2009). 한국판 해밀턴 우울증평가척도 (K-HDRS)의 확인적 요인 분석. 신경정신의학, 48, 21-28.
- 성형모, 김정범, 박영남, 배대석, 이선희, 안현희 (2008). 한국어판 벡 우울 설문지 2판의 신뢰도 및 타당도 연구. 생물치료정신의학, 14(2), 201-212.
- 송열매, 이해경, 김준원, 이건설 (2012). 인터넷으로 시행한 벡 우울 척도 2판의 신뢰도와 타당도: 대학생을 대상으로. 신경정신의학, 51, 402-408.
- 원호택, 박현순, 권석만 (1995). 한국판 공황장애 척도 개발 연구. 한국심리학회지: 임상, 14(1), 95-110.
- 유병관, 이해경, 이건설 (2011). 한국어판 벡 우울 설문지 2판의 신뢰도 및 요인분석: 대학생을 대상으로. 생물정신의학, 18, 126-133.
- 유상우, 김영신, 노주선, 오강섭, 김찬형, 남궁기, ..., 김세주 (2006). 한국판 Mini International Neuropsychiatric Interview 타당도 연구. 대한불안학회지, 2(1), 50-55.
- 이중서, 배승오, 안용민, 박두병, 노경선, 신현균, ..., 김용식 (2005). 한국판 Hamilton 우울증 평가 척도의 신뢰도, 타당도 연구. 신경정신의학: 44, 456-465.
- 이지현, 장정기, 박지홍, 안준호, 이철, 김창윤 (2003). 해밀턴 우울증 평가척도의 요인구조. 정신병리학, 12, 3-16.
- 이효운, 이인혜, 성주현 (2012). 성인 쌍둥이 및 쌍둥이 가족을 대상으로 한 불안민감성의 유전적 영향력에 관한 연구: 우울, 흡연과의 관계. 한국심리학회지: 건강, 17(1), 143-163.
- Lim, S. U., Lee, E. H., Whang, S. T., Hong, S. H., & Kim, J. H. (2014). Psychometric properties of the Beck Depression Inventory-II in Korea. 14' 한국임상심리학회 추계학술대회 포스터 발표.

- 임선영, 이은정, 정성원, 김희철, 정철호, 전태연, ..., 김정범 (2011). 한국판 백 우울 척도 2판의 타당화 연구. *대한불안학회지*, 7(1), 48-53.
- 임영진, 이소영, 김지혜 (2005). 불안민감도와 특질불안의 차별성과 공통성: 부적정서, 정적정서, 생리적 과각성 간의 관계. *한국심리학회지: 임상*, 24(2), 439-449.
- 전경구, 최상진, 양병찬 (2001). 통합적 한국판 CES-D 개발. *한국심리학회지: 건강*, 6(1), 59-76.
- 조맹제, 김계희 (1993). 주요우울증환자 예비평가에서 the Center for Epidemiological Studies Depression Scale (CES-D)의 진단적 타당성 연구. *신경정신의학*, 32(3), 381-399.
- 최영희, 김경미, 박기환, 윤혜영 (2003). 공황장애 환자에서 인지행동치료의 약물 감량 효과. *대한정신약물학회지*, 14(4), 367-376.
- 최영희, 김지혜, 임영진, 채정호, 우종민, 류한욱, 유범희 (2004). 공황장애에서 단기약물 치료가 불안민감도에 미치는 영향. *대한정신약물학회지*, 15(3), 333-338.
- 최영희, 최윤정, 우종민, 윤혜영 (2006). 광장공포증 유무에 따른 공황장애 환자의 치료 결과 비교. *인지행동치료*, 6(2), 163-178.
- 한오수, 안준호, 송선희, 조맹제, 김장규, 배재남, ..., 홍진표 (2000). 한국어 판 구조화 임상면담도구 개발: 신뢰도 연구. *신경정신의학*, 39(2), 362-372.
- 한오수, 홍진표 (2000). DSM-IV의 제1축 장애의 구조화된 임상적 면담. 서울: 하나의학사.
- Cho, M. J., & Kim, K. H. (1998a). Use of the center for epidemiologic studies depression (CES-D) scale in Korea. *The Journal of Nervous and Mental Disease*, 186(5), 304-310.
- Cho, M. J., Nam, J. J., & Suh, G. H. (1998b). Prevalence of symptoms of depression in a nationwide sample of Korean adults. *Psychiatry Research*, 81, 341-352.
- Kim, M-D., Hong, S-C., Lee, C-I., Kwak, Y-S., Shin, T-K., Jang, Y-H, ... & Whang, S-E. (2007). Prevalence of depression and correlates of depressive symptoms for residents in the urban part of Jeju Island, Korea. *International Journal of Social Psychiatry*, 53(2), 123-134.
- Lee, E-H., Kim, J-H., & Yu, B-H. (2009). Reliability and validity of the self-report version of the Panic Disorder Severity Scale in Korea. *Depression and Anxiety*, 26, E120-E123.