

아동 성폭력 피해 진술에 대한 신빙성 분석도구들의 타당도 연구*

이수정†
경기대학교

본 연구에서는 윈스탑센터에서 경찰조사를 보조하는 진술분석요원들이 사용하고 있는 아동성폭력 피해자 진술분석 도구 중에서 가장 주요한 도구인, 준거기반내용분석(CBCA), 사실성평가(RM), 성폭력발생맥락평가 도구들의 타당도를 검증하였다. 연구의 또다른 목적은 이들 도구들에 대한 현장 활용력을 높이기 위하여 통계분석을 통해 판단의 기준점(cutoff score)을 산출하기 위한 것이기도 하였다. 이를 위하여 본 연구에서는 모사실험설계(Smith, 1997)를 활용하였는데, 예컨대 진짜 피해를 당하지 않은 성폭력 피해를 가장한 집단과 진짜 피해자 집단을 비교하는 방식으로 집단 간 차이를 분석하였다. 이때 집단 간 차이분석은 주로 최혜림과 이수정(2009)의 연구에서 문항 변별도가 적절하다고 판명된 문항을 중심으로 하여 실시되었다. 대부분의 문항은 허위로 피해를 가장한 진술과 진짜 피해 진술을 변별해내는 데에 유용하였으며 총점 상에 있어서도 충분한 변별력을 갖추고 있음이 확인되었다. 판단의 기준점은 3점 척도를 적용하였을 때, 11개 문항으로 구성된 준거기반내용분석(CBCA)은 13점(AUC .87) 이상, 8개 문항으로 구성된 사실성평가(RM)는 9점(AUC .89) 이상, 2점 척도를 적용하였던 성폭력발생맥락평가의 14개 문항 총점은 9점(AUC .85) 혹은 10점(AUC .85) 이상이 적절한 것으로 나타났다. 최종적으로 이들 기준점을 의사결정 시 어떻게 반영하는 것이 적절한지 논의하였다.

주요어 : 진술타당성분석, CBCA, RM, 성폭력발생맥락평가, ROC, 변별력

경찰청은 아동성폭력피해자 조사 시 아동의 진술특성을 이해하고 진술의 신빙성을 분석할 수 있는 민간 전문인력들의 업무가 유용하다는 판단 아래, 2009년도부터 서울 및 경기 지역에 위치한 다섯 곳의 윈스탑센터에서 아동 피해자 진술분석을 실시토록 하고 있다. 이

제도에서는 아동을 대상으로 하는 성폭력사건의 송치 시 진술분석을 훈련받은 심리요원들이 피해진술에 대한 타당성 분석을 실시하여 그에 대한 소견서를 사건 기록과 함께 제출한다. 2010년도에는 다섯 곳에서만 시범 시행하던 이 제도를 전국 18곳 윈스탑센터로 확대하여

* 본 연구는 한국연구재단 2009년도 기초인문사회(단독)B00979의 지원금으로 수행되었음.

† 교신저자: 이수정, (442-760) 경기도 수원시 영통구 의의동 산 94-6 경기대학교 일반대학원 범죄심리학과,
Tel : 031-249-9198, E-mail : suejung@kyonggi.ac.kr

주로 13세 미만의 아동 및 장애인의 진술에 대해서도 진술타당성분석(Statement Validity Analysis)을 실시하도록 하고 있다. 본 연구는 현재 윈스탑센터에서 진술분석요원들이 사용하고 있는 진술분석에의 도구가 피해 진술의 진위 여부를 제대로 변별해낼 수 있는지, 만일 변별에 유용성이 있다면 보다 정확한 변별력을 확보하기 위한 판단의 기준점(cutoff score)은 몇 점일 것인지 탐색하고자 노력하였다. 이는 진술타당성 분석 도구의 타당성 지표를 확보하기 위함이기도 하다.

진술타당성분석(Statement Validity Analysis)은 법정심리 분야에서는 사실상 역사가 아주 오래 된 주제이다. 진술타당성분석에 대한 법적 근거는 1954년 서독의 대법원 판결까지 거슬러 올라간다(Undeutsch, 1982, 1989). 당시 문제가 되었던 사건은 성폭력 사건으로서 진술밖에는 존재하지 않았었고 따라서 법정에서 양측의 진술에 대한 사실성이 다투어졌었다 서독 법원은 당시 사건의 피고인과 증인이었던 피해자의 진술을 법정/임상적 방법으로 분석토록 하여 신빙성을 가리고자 하였는바, 이후 이 같은 진술의 타당성을 가리고자 하는 시도는 독일을 포함한 유럽의 법정에서 일반적으로 받아들여지는 절차가 되었다.

미국의 경우, 1982년 이후 소개가 되어 일부 법원에서 활용되다가 진술의 내용적 사실성을 분석하던 준거기반내용분석(Criterion-Based Content Analysis; 이하 CBCA)에 대한 타당도연구(Boychuk, Esplin, & Raskin, 1988)가 수행되면서 학계에도 알려지기 시작하였다. 당시 19개의 지표를 가지고 있던 CBCA는 이후 많은 실증연구들(Boychuck, 1991; Lamb, Sternberg, Esplin, Hershkowitz, Orbach, & Hovav, 1997; Raskin & Esplin, 1991; Vrij, Akehurst, Soukara, Bull, 2002; Yuille, 1988)을 통해 진술타당성분석에의 유용성이 확인되었고 우리나라에도 소개되기에 이른(김현정, 2009; 박광배, 1995; 조은경, 2004).

최근 우리나라의 형사사법기관 중 경찰청에서는 앞서 언급하였다시피 윈스탑센터에 성폭력 피해자 진술분석요원을 배치하여 13세 미만 아동과 지적장애인의 피해 진술에 대해서는 분석보고서를 제출토록 하고 있다. 이때 진술분석요원은 외국에서 타당성이 입증된 CBCA를 포함한 표 1과 같은 진술내용 및 피해사실의 맥락적 정

황에 대한 훈련(40시간)을 받고 분석을 실시하고 있다. 본 연구는 이 같이 최근 훈련된 인력에 의해 이루어지고 있는 윈스탑센터에서의 진술분석 도구가 피해를 입증하는 데에 과연 변별력을 지니는지 확인하기 위해 수행되었다. 이를 위하여 본 연구에서는 모사준거집단 실험설계라는 방법을 통해 허위로 피해사실을 주장하는 집단과 진실한 피해자 집단과의 진술 상의 특징이 어떻게 다른 것인지 비교 분석하였다.

모사준거집단 설계는 피병 평가연구에서 주로 사용되는 **모사집단 설계(simulation design)**와 **준거집단 설계(known-groups design)**를 혼합한 것이다(Miller, 2001). 일반적으로 모사집단 설계는 참가자들에게 특정 방식(예컨대, 정신병리의 가장)으로 조사자들의 질문에 응답할 것을 지시한 다음, 이들의 수행과 솔직하게 반응한 응답을 분석하는 방식으로 이루어진다(Smith, 1997). 모사집단 설계의 주요 강점은 연구자가 잘 통제된 실험 조건 하에 체계적인 비교연구를 실시할 수 있다는 점이다. 그러나 동시에 모사집단 설계의 가장 큰 단점은 실제 임상 혹은 사법장면에서의 피조사자의 반응과는 거리가 있는 반응만을 수집하게 되므로 일반화의 취약성이 존재한다는 것이다(Bourg, Conner, & Landis, 1995).

한편, 준거집단 설계방법은 실제로 법정임상 실무에서 특정한 임상증세를 지닌 자와 지니지 않은 자를 비교하는 방법이다. Rogers(1984, 1997)는 이와 같은 두 설계를 함께 사용할 때 단점을 보완할 수 있다고 하면서, 모사집단 설계를 통해서 내적 타당도를, 준거집단 설계를 통해서 외적 타당도를 확보할 수 있다고 하였다. 본 연구에서는 Rogers가 제안한 대로 모사집단(피해 가장 허위진술 집단)과 준거집단(실제 성폭력 피해자 집단)이 동시에 존재하는 모사준거집단 설계를 채택하였다.

분석에 포함되었던 세 가지 진술분석도구는 표 1에 서처럼 준거기반 내용분석(Criterion - Based Content Analysis; 이하 CBCA), 사실성 평가(Reality Monitoring; 이하 RM), 성폭력발생맥락평가 체크리스트였다. 이중 RM은 진술의 신빙성을 판단하기 위해 사용되는 CBCA 이외에 현재 가장 널리 알려진 진술타당성 평가 기준으로서 실제 경험한 사건에 대한 기억이 지각

표 1. 아동 성폭력 피해자 진술의 신빙성 분석 구성 내용

조항	구성 내용
준거 기반 내용 분석	<p>준거기반 내용분석(Criterion - Based Content Analysis; CBCA)은 5개의 범주에 19개의 하위 준거로 구성되어 있으며, 아동의 인지적 능력과 동기적 요소를 고려하여 아동 진술의 진실성을 평가한다. 개별 준거들은 진술의 진실성 지표로서 허위로 꾸며낸 진술보다 진실한 진술에서 더 많이 나타날 것임을 전제로 하고 있다. 각 문항에 대하여 0, 1, 2점으로 채점하는데, 현재는 총 38점 중 중앙치인 19점 이상이면 진술에 신빙성이 있다고 판단하고 있다. 본 연구에서 준거기준점 탐색에 포함되었던 문항은 최혜림과 이수정(2009)의 연구 결과 문항변별도가 높다고 판단된 11문항이었고 이들 문항은 '1. 논리적 구성', '3. 세부정보의 양', '4. 맥락적 깊이', '5. 상호작용', '6. 대화 재현', '7. 예기치 않은 상황', '8. 일상적이지 않은 정보', '9. 세부 정보 과잉', '10. 이해하지 못한 정보', '12. 주관적 경험', '13. 가해자 정신상태 추정'이었다.</p>
사실성 평가	<p>사실성 평가(Reality Monitoring; RM)는 지각의 과정에 근거하는 기억과 내부처리에 근거한 기억은 다를 것이라는 가정에서 출발하여 피해자 진술의 내용에 있어 지각적인 정보(시각, 청각, 후각, 미각 등)가 포함되어 있는지를 측정한다. 진술의 생생함, 감각기관의 정보, 공간 정보, 시간 정보, 정서와 감정, 이야기의 재구성, 사실성, 인지적 추론의 8가지 준거에 대해 0, 1, 2점으로 채점하는데, 현재 총 16점 중 중앙치인 8점 이상이면 진술에 신빙성이 있다고 본다. 본 연구에서 준거기준점 탐색에 포함되었던 문항은 최혜림과 이수정(2009)의 연구 결과 문항변별도가 충분하다고 판단된 8문항이었고 문항내용은 '1. 명료성', '2. 지각정보', '3. 공간정보', '4. 시간정보', '5. 재연가능', '6. 정동', '7. 사실적 묘사', '8. 인지적 작동'이었다.</p>
성폭력 발생 맥락 평가	<p>성폭력 발생 맥락 평가 체크리스트는 아동과의 면담, 부모 면담, 기타 정보 등을 종합하여 고려할 뿐만 아니라 각각의 정보가 가지는 중요성, 각 정보에 대한 다른 해석의 가능성들을 체계적으로 고려하도록 고안되었다. 따라서 아동의 진술뿐만 아니라 기타 다른 정보들을 전체적인 맥락 속에서 판단함으로써 아동 진술의 진위여부를 판단하는 데 유용하다. 아동의 진술능력, 성폭력이 발생한 맥락, 아동의 상태, 기타 정보라는 4가지 범주하에 구성된 25개 문항에 대해 0점 혹은 1점으로 평가토록 하는데, 총 25점 중 중앙치인 13점 이상이면 성폭력 발생 가능성이 농후하다고 판단하고 있다. 본 연구에서 준거기준점 탐색에 포함되었던 문항은 최혜림과 이수정(2009)의 연구 결과 문항변별도가 높다고 판단된 14문항이었고 이들 문항은 '1. 아동의 관점 반영', '2. 성폭력에 대한 구체적 진술', '4. 발생장소', '5. 발생시간', '6. 가해자의 유인방법', '7. 가족이 있었던 곳', '8. 아동의 옷', '9. 가해자의 옷', '10. 특징적 사건 존재', '11. 아동의 정서적 상태 진술', '12. 비밀유지의 강요', '13. 아동의 자발적 보고', '14. 사건을 들은 사람의 반응', '16. 진술과정 나타난 아동의 감정'이었다.</p>

적 정보처리과정을 거치기 때문에 실제 경험한 사실은 더 많은 지각정보(perceptual information), 즉 시각적으로 세부적인 정보, 소리, 냄새, 맛, 신체적인 감각들을 포함하고 있다는 전제에서 비롯되었다(Johnson & Raye, 1981; Kapardis, 2002; McNally, 2005). 현재 알려진 RM의 지표는 Sporer(1997)가 제시했던 것으로서 이후 여러 연구자들(Strömwall & Granhag, 2003; Vrij, 2004)에 의해 진술 신빙성 판단에의 유용성이 입증되었다.

성폭력발생맥락평가 체크리스트는 앞서 언급한 두 가

지 진술분석 도구들과는 달리 피해자의 진술 이외에 조사과정 상 드러난 정황, 부모와의 면담 등을 토대로 피해 사실의 허위 발고 가능성까지를 판단하도록 제작되어 있다. 이 도구는 Faller(2007)의 체크리스트를 토대로 하고 있는데, 이 체크리스트는 성폭력 피해상황의 전형적인 특징 외에 아동의 피해 사실에 대한 발고기 이혼이나 양육권 논쟁 등 부모의 불화에 의해 허위로 보고되었을 가능성을 평가하도록 구성되어 있다. 연구자들(Faller, 1991; Gardner, 1987; Yuille, 1988)은 아동의 성폭력 피해에 대한 허위 발고 가능성은 북미지

역에서 보고된 전체 사건 중 약 15%에서 25%까지 된다고 지적하였다.¹⁾ Faller(2007)는 이와 같은 사실에 근거하여 아동 진술의 허위 보고 여부를 평가하기 위해서는 부모의 영향, 조사자의 암시 등을 검토함으로써 진술의 진위여부를 검토하고, 피해아동이 평소 성적 환상을 갖고 있었는지, 진술을 상상에 의해 꾸며내는 것인지, 사건과 무관하게 다른 사람을 허위 지목하였는지 등을 종합적으로 고려해 아동진술의 타당성을 평가하고자 시도하였다. 본 연구에서는 그녀가 제시한 성폭력 발생맥락을 평가할 수 있는 26개 기준들 중 타당도 문항을 일부 개편하여 25개 문항으로 구성된 이수정(2009)의 성폭력발생맥락평가 척도에 대하여 총점 및 문항 수준에서 변별타당도를 확인하였다.

세 가지 진술분석 도구의 집단 간 변별력을 확인하기에 앞서 본 연구에 포함되었던 척도들의 개별 문항들에 대한 분석은 최혜림과 이수정(2009)의 연구에서 보고된 적이 있다. 고진검사이론과 문항반응이론 모두를 적용한 결과, 표 1에 설명된 CBCA의 19문항 중에서는 11문항이, RM의 경우에는 8문항 모두가, 성폭력발생맥락평가 25문항 중에서는 14문항이 문항변별력이나 내적합치도 면에서 우수하다고 분석되었다. 세 가지 척도 중 CBCA의 문항특성에 대해서는 외국의 선행연구가 많은 편인데, Vrij(2004)의 CBCA에 대한 메타연구에 따르면, 11문항 중 ‘세부정보의 양’과 ‘맥락적 깊이’, ‘대화 재현’은 외국의 선행 연구에서 거짓 진술보다 참 진술에서 일반적으로 더 많이 나타나는 것으로서 이들 문항은 최혜림과 이수정(2009)의 연구에서도 가장 변별력이 높은 문항들로 확인되었다. 이들 문항에 대한 결과가 국제적인 결과의 흐름을 따르는 데 반해 최혜림과 이수정(2009)의 연구에서는 ‘이해하지 못한 정보’ 문항이 진짜 피해를 당한 국내 아이들의 진술에 있어 일반

적으로 나타나는 특성인 것으로 분석되었다. 이는 Vrij(2005)가 검토한 여덟 개의 선행 연구 중 한 연구에서만 확인되는 결과인 바, 이러한 외국의 선행 연구와 우리나라 연구가 차이는 성에 대한 아동의 지식수준의 차이에 기인하는 것으로 해석해 볼 수 있다. 미국 혹은 유럽의 경우 성에 대해 개방적이고 현실적인 교육을 실시하기 때문에 성지식에 따라 좌우되는 ‘이해하지 못한 정보’ 문항이 참 진술과 거짓 진술을 잘 구분해주지 못하는 반면 성에 대해 폐쇄적이고 대부분 이론적인 성교육에 머물러 있는 우리나라에 있어서는 이와 같은 추세가 진술분석에서는 역작용을 할 수 있을 것이다. 본 연구에서는 국내 피해자들의 특성을 반영하는 최혜림과 이수정(2009)의 연구결과를 토대로 신뢰성 있는 문항이라 평가된 CBCA, RM, 성폭력발생맥락평가의 준거관련 타당성을 확인하였다.

현재 윈스탐센터에서는 최혜림과 이수정(2009)의 연구에서처럼 한국 성폭력 피해아동의 특성에 대해서는 특별한 고려를 하지 않은 채, 모든 문항을 토대로 산출된 총점을 기준으로 하여 아동 혹은 지적장애 피해자들의 진술신빙성을 판단하고 있다. 총점을 기준으로 보았을 때 중앙치를 중심으로 하여 신빙성을 판단하는데, 이에 대해서는 실증적 근거가 존재하지 않는다. 본 연구는 따라서 최혜림과 이수정(2009)의 연구에서 산출된 우수한 심리측정학적 특징을 갖춘 문항들이 모사준거집단설계에서 과연 유용한 변별력을 지니는지, 나아가 진술의 사실성에 대한 판단의 기준점이 어느 정도일 때 최대의 변별력을 지니게 되는지 실증적으로 확인하고자 하였다.

연구방법

연구대상

본 연구에 포함되었던 피조사자들로는 두 집단이 존재하였다. 실제로 경기해바라기아동센터를 내원한 성범죄 피해 아동들과 모사준거집단설계에서 또다른 준거집단으로 설정된 통제집단이었다. 이때 통제집단은 성범죄 피해를 가장하라는 지시를 받았는데, 허위진술을 보다 사실적으로 유도하기 위하여 여성부에서 배포한 성

1) 북미지역의 경우 이혼소송 시 양육권 분쟁이 첨예하며 이는 위자료나 자녀양육비에 대한 규모가 커, 조정이 어렵기 때문이다. 허나 국내의 경우 이와 같은 사례는 쉽게 발견하기 어려우며 알려진 바에 따르면 아동 대상 성폭력 사건의 신고율(한국성폭력상담소 비공식 기록)은 전체 사건 중 6%에 해당할 정도로 압수범죄율이 높은 것으로 보고되고 있다. 이는 성폭력 피해에 대한 발고 시 사실상 피해자가 얻을 수 있는 보상의 수준이 경미하기 때문이기도 하며 따라서 허위발고의 가능성 역시 북미지역보다는 현저히 낮을 것으로 판단된다.

표 2. 피해자연령

	실험집단	통제집단
	빈도(%)	빈도(%)
만 11세	4(30.8)	0(0.0)
만 12세	6(46.2)	0(0.0)
만 13세	1(7.7)	8(88.9)
만 14세	1(7.7)	0(0.0)
만 15세	1(7.7)	1(11.1)
합계	13(100.0)	9(100.0)

교육자료를 사전에 시청하고 본인이 성범죄의 피해자가 되었다고 상상하면서 진술을 하도록 유도하였다. 실험 집단의 참가자들은 경찰에 성폭력 피해사실을 이미 신고하여 형사절차가 진행되고 있는 사건들²⁾의 아동 피해자들이었다. 표 2에는 두 집단의 인구학적 특성이 제시되어 있다.

실험집단에 속하였던 실제 성폭력 피해자집단의 피해 관련 범죄명은 강간 6명(46.2%)과 강제추행 7명(53.8%)이 끌고루 포함되어 있었다. 통제집단의 경우에는 강제추행을 연상하며 진술을 하였던 사례가 8명(88.9%) 강간을 연상하였던 사례가 1명(11.1%)이었다.

연구에 포함되었던 모든 아동은 만 11세에서 15세 이하의 초등학교 재학생과 중학교에 재학 중인 여자청소년들로서 평균 수준 이상의 지적 능력을 지녔으며 의사소통에는 별다른 불편함이 없는 자들이었다. 실제 성폭력 피해를 당한 실험집단에 비해 통제집단의 평균연령(모두 중학생)이 약간 높았던 연유는 통제집단의 피조사자들이 성폭력 피해를 모사해야 하였기 때문에 상당한 수준의 성교육 지식을 이미 지니고 있으며 실험으로 인한 부작용이 가능한 존재하지 않아야 한다는 피험자 인권보장과 관련된 전제조건 때문이었다.

2) 일부 사건은 이미 검찰에 의하여 기소가 되었고 나머지 사건들은 당시 고소가 진행되고 있는 사건들이었다. 물론 허위발고율이 15-25%에 이른다는 Faller(1991)의 연구에 따르면 13건 중에서도 2건에서 3건 정도는 허위발고 가능성이 있을 수 있으나, 현 단계에서 그 문제를 입증하기는 불가능하였다. 따라서 담당 경찰의 권고를 받아 허위발고가능성이 낮은 13건을 선정하였다.

측정도구

본 연구는 최혜림과 이수정(2009)의 연구결과를 토대로 아동 성폭력 피해 진술의 타당성을 판단하는 세 가지 분석도구(CBCA, RM, 성폭력발생맥락평가)의 검사중점 수준에서의 타당도를 확인하는 것을 목적으로 하였다.

현재 원스탑센터에서 사용하고 있는 CBCA는 원래 총 19문항으로 구성되어 0점(해당 진술 없음), 1점(한 건의 해당 진술 있음), 2점(두 건 이상의 해당 진술 있음)으로 채점하도록 되어 있다. 2009년도에 서울 경기권의 다섯 곳 원스탑센터에서 수거된 자료를 토대로 한 최혜림과 이수정(2009)의 연구에 따르면 총 19문항의 내적합치도는 .54에 불과하였다. 따라서 연구자들은 고전심리검사이론과 문항반응이론에 따라 문항분석을 실시하여 문제가 있다고 판정된 8문항을 제외하고 11문항만을 사용할 것을 권고하였다. 이들 11문항에 대해 본 연구에서 산출한 내적합치도는 .86까지 향상되었다.

RM의 경우 많은 연구들(Strömwall, 2004; Vrij, 2004)이 8개 문항의 유용성을 보고하고 있으며 최혜림과 이수정(2009)의 연구에서도 이들 문항들은 모두 우수한 문항특성을 지니는 것으로 나타났다. 이 여덟 개 문항은 CBCA와 마찬가지로 0점, 1점, 2점으로 채점³⁾되며, 본 연구에서 산출된 내적합치도는 .90이었다.

성폭력발생맥락평가는 Faller(2007)의 체크리스트를 토대로 이수정(2009)이 구성한 것으로서 최혜림과 이수정(2009)의 연구에 따르면 25개 문항 중 우수한 심리측정학적인 특성을 지니고 있는 문항은 총 14문항이었다. 이들 문항은 해당사항이 있는 경우 1점, 해당 사항이 없는 경우 0점을 주도록 2점 척도로 구성되어 있었다. 25문항 중 일부 문항('아동의 성행동', '물적 증거', '가해자의 자백')의 변별도 지수는 음수이거나 0에 가까운 매우 낮은 수치를 기록하였기에 연구자들은 변별도 지수가 낮은 문항들을 제외하고 변별력이 좋은 문항만을 사용하기를 권고한 바 있다. 본 연구에서는 이들의 권고

3) 마지막 8번은 '인지적 작동'으로서 원래는 역산하여야 하는 문항인데, 혼돈을 피하기 위하여 '추측하는 내용이 포함되어 있으면' 0점을 '추측하는 내용이 없이 객관적 사실에 근거한 진술이면' 2점을 주도록 지시하였다.

에 따라 14문항만을 사용하였으며 이들 문항에 대한 내적합치도는 .84였다. 최종적으로 산출 가능한 CBCA, RM, 성폭력발생맥락평가의 최대 총점은 각 척도의 특성을 반영하였을 때 22점, 16점, 14점이 되었다.

실험절차 및 분석방법

측정도구 상에서 두 준거집단의 변별을 보다 객관적으로 집행하기 위하여 실험은 두 부분으로 나누어 블라인드 설계로 구성되었다. 즉 앞 단계에서는 경기해바라기센터에 근무하는 한 명의 면담전문가가 진짜 혹은 위장 피해자를 모두 조사하도록 하였다. 이때 면담은 기본적으로 Fisher와 Geiselman(1992)이 개발한 인지면담기법(Cognitive Interview, CI)을 적용하여 집행되었다. 인지면담을 포함하여 면담은 7단계로 구성되었는데, ① 조사자 소개와 진술녹화 설명, ② 라포형성, ③ 신체적 이완, ④ 진술능력을 평가하기 위한 오리엔테이션, ⑤ 인지면담으로 이루어진 사건관련 질문 및 피해진술, ⑥ 아동에 대한 격려 및 보상, ⑦ 종결로 구성되었다.

이렇게 수거된 면담 녹화자료는 동 기관에 근무하는 10건 이상의 진술분석을 적용해 본 적이 있는 다른 조사자에 의하여 평가되었다. 이 조사자는 총 22건의 녹화된 진술이 어느 준거집단에 속하는 것인지 알지 못한 채 진술분석에 임하였다. 결과적으로 최종 분석에 포함되었던 자료는 두 번째 조사자가 평가한 진술분석의 결과치였다.

애초 본 연구는 세 가지 분석도구(CBCA, RM, 성폭력발생맥락평가)의 총점 상에서의 타당도를 확인하고 성폭력 피해와 관련된 진술신빙성에 대한 의사결정의 기준을 마련하는 것을 목적으로 하였다. 따라서 진짜 성폭력 피해자 집단이었던 실험집단과 성교육 비디오를 시청한 후 피해를 가장하는 집단과의 변별기준점을 산출하기 위하여 ROC(Receiver Operating Characteristic) 분석을 적용하였다. ROC 분석이 제공하는 AUC(Area Under Curve)는 변별력의 지표로 널리 활용되는데, 대상자에 대한 완벽한 변별이 가능한 경우 1.00의 값을, 전혀 변별의 유용도가 없는 우연 수준의 경우 .50의 값을 지닌다. 참고로 Swets(1988)는 AUC가 .90 이상의 값을 지닌 때는 매우 정확한 수준의 변별력을, .70 이상 .90 이

하의 AUC를 지니면 적당한 수준의 변별 정확성을, .60 정도 이하이면 정확성 차원에서 수용되기 힘든 변별력을 지니는 것으로 판단할 수 있다고 하였다. ROC 분석 이외에 추가로 세 척도에 있어 개별 문항 상의 실험집단과 통제집단의 차이검정도 실시하였다. 본 연구의 경우 사례수를 충분하게 확보하지 못하였기에 개별문항에 대한 분석결과는 문항의 변별도 기준으로 활용하지는 않았다.

결 과

실제로 성폭력 피해를 당하여 경기해바라기아동센터에 내원하였던 실험집단의 피해 진술과 성교육 비디오를 시청한 후 허위로 피해를 가장한 집단의 진술내용상의 차이를 확인하기 위하여 각 척도의 총점에 대해 t 집단 간 차이검정을 실시하였다.

t 검정 결과, 유의도 .001 수준에서 두 준거집단은 세 척도의 총점 모두에서 진술 상의 유의한 차이를 지니는 것이 확인되었다.

세 척도에 포함되었던 개별문항에 있어서의 집단 간 차이를 확인한 결과, 우선 CBCA에서는 유의도 .05 수준에서 세 개 문항이 집단 간 차이를 반영하지 못하고 있는 것으로 나타났다. 이들 문항은 ‘예기치 않은 상황’, ‘세부 정보 과잉’, ‘이해하지 못한 정보’ 등이었다. 이들 문항의 타당성에 대해서는 추후 연구에서 다시 검토하기로 하고 본 연구에서는 일단 이들 세 문항을 포함하여 총 11문항에 대한 총점을 산출하였다.

RM의 경우 오로지 한 문항이 준거집단 간 차이를 반

표 3. 각 척도별 준거집단에 대한 차이검정

	실험집단	통제집단	t
	평균(SD)	평균(SD)	
CBCA	15.36(2.54)	7.44(4.39)	5.44***
RM	13.54(2.15)	6.89(3.06)	6.01***
성폭력맥락	12.00(2.24)	6.22(3.73)	4.55***

***p<.001

표 4. CBCA 문항별 차이검정

	실험집단 (n = 13)		통제집단 (n = 9)		t
	M	SD	M	SD	
1번	2.00	.00	1.11	.78	4.15***
3번	1.85	.38	1.44	.53	2.09*
4번	1.92	.28	1.11	.60	4.29***
5번	1.69	.63	.67	.71	3.57**
6번	1.62	.51	.56	.73	4.05***
7번	.77	1.01	.44	.53	.88
8번	1.54	.78	.56	.73	3.00**
9번	1.15	.90	.56	.53	1.79
10번	.46	.66	.33	.71	.44
12번	1.69	.48	.56	.53	5.25***
13번	.77	.73	.11	.33	2.53*

*p < .05, **p < .01, ***p < .001

영하지 못하는 것으로 나타났다. 이는 ‘정동’으로서 허위로 피해상황을 진술한 통제집단도 비교적 풍부한 진술을 하였던 것으로 확인되었다.

성폭력발생맥락평가 척도의 개별문항에 있어서의 집단 간 차이검정에서는 최혜림과 이수정(2009)의 연구결

표 5. RM 문항별 차이검정

	실험집단 (n = 13)		통제집단 (n = 9)		t
	M	SD	M	SD	
1번	1.69	.48	.89	.60	3.48**
2번	1.46	.66	.78	.44	2.71**
3번	1.85	.38	1.22	.44	3.57**
4번	1.69	.48	.89	.78	2.99**
5번	2.00	.00	1.00	.71	5.16***
6번	1.46	.52	1.00	.71	1.77
7번	2.00	.00	1.00	.50	7.29***
8번	1.38	.77	.11	.33	4.65***

*p < .05, **p < .01, ***p < .001

표 6. 성폭력발생맥락평가 문항별 차이검정

	실험집단 (n = 13)		통제집단 (n = 9)		t
	M	SD	M	SD	
1번	.92	.28	.44	.53	2.78*
2번	.92	.28	.11	.33	6.22***
4번	1.00	.00	.56	.53	3.10**
5번	1.00	.00	.44	.53	3.84***
6번	.85	.38	.44	.53	2.09*
7번	.62	.51	.11	.33	2.61*
8번	.69	.48	.44	.53	1.14
9번	.85	.38	.78	.44	.39
10번	.38	.51	.11	.33	1.42
11번	1.00	.00	.56	.53	3.08**
12번	.54	.52	.22	.44	1.49
13번	.92	.28	.56	.53	2.14*
14번	.54	.52	.56	.53	-.08
16번	.77	.45	.67	.50	.51

*p < .05, **p < .01, ***p < .001

과와는 달리 여러 개의 문항이 유의한 결과를 산출하지 못하였다. 8번, 9번, 10번, 12번, 14번, 16번 문항이 이 들어있는데, 이 문항들에 대해서는 보다 큰 표본을 확보하여 문항변별력을 추후 재분석해보기로 하고 본 연구에서는 일단 14개 문항 모두를 합산하여 총점을 산출하였다.

다음으로는 ROC 분석을 적용하여 각 척도 상에 있어 두 준거집단의 차이를 가장 잘 변별해낼 수 있는 총점 상의 준거기준점은 무엇인지 탐색하였다. 표 7, 표 8, 표 9는 각기 CBCA, RM, 그리고 성폭력발생맥락평가의 ROC 분석 결과치를 보여주고 있다.

표 7에 제시되었다시피 3점 척도로 평가하였던 CBCA의 11문항에 대하여 준거집단에 대한 총점 상의 변별기준은 13점이 적합한 것으로 판단되었다. 이때의 AUC는 .87인 것으로 나타나 Swets(1988)가 제안한 기준으로 보았을 때 적절한 수준의 정확성을 지니는 것으

표 7. CBCA의 ROC 분석 결과

	AUC	민감도	특이도	PPP	NPP
7이상	.778	1.000	.555	.765	1.000
8이상	.778	1.000	.555	.765	1.000
9이상	.833	1.000	.667	.813	1.000
10이상	.833	1.000	.667	.813	1.000
11이상	.833	1.000	.667	.813	1.000
12이상	.812	1.000	.778	.846	1.000
13이상	.868	.846	.890	.917	.800
14이상	.829	.770	.890	.910	.728
15이상	.791	.692	.890	.901	.667
16이상	.808	.616	1.000	1.000	.643

로 판단할 수 있었다. 이 변별기준점은 피해사실의 진위성을 91.7%(positive predictive power)까지 정확하게 변별하였다. 이때의 민감도(sensitivity; true positive rate)는 84.6%, 특이도(specificity; 1-false positive rate)는 89.0%로 특정 지표상에 불균형은 존재하지 않는 것으로 나타났다.

표 8에는 RM 총점에 대한 ROC 분석결과가 제시되어 있다. 변별 정확성의 기준이 되는 AUC 지표를 살펴

표 8. RM의 ROC 분석 결과

	AUC	민감도	특이도	PPP	NPP
6이상	.722	1.000	.445	.722	1.000
7이상	.722	1.000	.445	.722	1.000
8이상	.722	1.000	.445	.722	1.000
9이상	.889	1.000	.778	.867	1.000
10이상	.850	.924	.778	.857	.876
11이상	.868	.846	.890	.917	.800
12이상	.868	.846	.890	.917	.800
13이상	.885	.770	1.000	1.000	.750
14이상	.808	.616	1.000	1.000	.643

본 결과 8문항으로 구성된 RM 총점 상 9점을 기준으로 하였을 때 피해의 진위성을 밝히는 데 최상의 변별력을 지니는 것이 확인되었다. 이때 AUC는 .89로, Swets(1988)의 기준으로 볼 때, 적절한 수준의 변별 정확성을 지녔다. 이 기준을 적용하여 보았을 때 피해 사실의 진위성은 86.7%까지 정확하게 변별이 되었다. 이때의 민감도는 100.0%, 특이도는 77.8%로 특이도에 비해 민감도가 더 높은 것으로 나타났다.

표 9에는 성폭력발생맥락평가 총점에 대한 ROC 분석결과가 제시되어 있다. 2점 척도로 구성된 14문항의 총점에 대한 AUC 지표를 살펴 본 결과 9점 혹은 10점을 기준으로 하였을 때 최상의 변별력을 지니는 것이 확인되었다. 이들 두 준거기준점에 대한 AUC는 .85에 이르는 것으로 나타났으며 이 기준은 피해사실의 진위성을 85.7% 정확하게 변별해내었다. 진술타당성분석은 단순한 양적 기준에만 근거하여 의사결정을 하는 것이 아니기에 실무적으로는 25문항 중 선택된 14문항에 있어, 분석평가의 결과치가 9점이나 10점 이상으로 산출되는 경우에는, 피해진술이 성폭력사건의 전형적인 발생맥락과 높은 일관성을 보인다고 평가하면 될 것이다. 또한 결과를 해석할 때에는 다른 척도들의 결과도 참조하여 최종적인 진술 신빙성에의 척도 간 일관성이 있는 판단을 내리면 될 것이다. 참고로 CBCA와 RM 등 진술 타당성 분석 도구들을 하나 이상 사용하는 경우에 하나의 도구만을 적용하는 때보다 통계적으로 유의한 수준에서 변별 정확율을 높일 수 있다고 한다(Vrij, Akehurst, Soukara, & Bull, 2002).

표 9. 성폭력발생맥락평가의 ROC 분석 결과

	AUC	민감도	특이도	PPP	NPP
7이상	.722	1.000	.444	.722	1.000
8이상	.778	1.000	.556	.765	1.000
9이상	.850	.923	.778	.857	.875
10이상	.850	.923	.778	.857	.875
11이상	.791	.692	.889	.900	.667
12이상	.714	.538	.889	.875	.571

결론 및 논의

서론에서 언급했듯이 본 연구에서는 현재 윈스탑센터에서 사용하고 있는 진술분석 도구들 중 세 개의 도구, CBCA, RM, 그리고 성폭력발생맥락평가의 타당도에 대한 증거를 수집하는 것이었다. 또한 진술타당성을 현실적으로 의사결정 함에 있어 그 근거가 될 수 있는 판단의 기준을 총점 상에서 산출하여 보는 것을 목적으로 하였다. 집단 간 차이분석과 ROC분석에서는 현재 선택된 문항들의 총점이 피해 진술의 진위성 판단에 비교적 우수한 변별력을 지니는 것이 확인되었다. 이는 성폭력을 직접 경험한 아동의 경우 풍부한 내용적, 지각적 기억 요소를 지니지만, 성폭력을 경험하지 않은 아동은 실제 사건에서 무엇이 일어났는지를 이야기할 때 상세히 재현할 가능성이 적기 때문이라는 Raskin과 Esplin(1991)의 주장을 다시 한 번 확인하여 준다.

본 논문에서 소개하고 있는 진술타당성 분석도구들은 결국에 가서는 법정에서 유용성이 다투어질 수밖에 없다. 이는 이와 같은 법정평가도구들이 증거에 대한 Daubert 기준을 충족시키는지 하는 문제와 관련되어 있다. 영미법정의 경우 Daubert 기준은 전문증거의 수용 여부를 판가름하는 매우 중요한 근거가 되고 있다(Honts, 1994). 특히 미국 대법원에서는 전문증거에 대한 다섯 가지 기준을 제시하였다. 첫째, 과학적인 기술에 대한 가설검증이 가능한가? 둘째, 전문증거로서 특정 기술이 전제하는 내용이 검증될 수 있는 사안인가? 셋째, 오류율이 알려져 있는가? 넷째, 전문증거를 산출하기 위한 방법이 동료 전문가들이 심사하는 학회지에 게재된 적이 있는가? 다섯째, 전문증거를 산출하는 기술이 과학공동체에 의해 받아들여지고 있는가? 미국의 법정은 이 같은 다섯 가지 기준을 근거로 전문증거의 수용 여부를 판단한다. Vrij(2005)는 이 기준을 적용하자면 마지막 질문에 대해 진술타당성분석은 여전히 논쟁의 소지가 있다고 판단한다. 외국의 법정에서 이 같은 논의가 깊이 있게 이루어지고 있는데 반해, 우리나라의 경우 법정평가(forensic evaluation) 분야는 여전히 취약하다. 전문가들이 사용하는 평가도구는 형사사법 분야에서 검증되지 않은 채 남용되고 있으며, 더욱이 이에 실증적 결과가 전혀 존재하지 않는 전문가의

감정이 재판부에 의해 받아들여지기도 한다. 이와 같은 국내 상황을 고려하여 볼 때 본 논문과 같은 실증적 연구결과는 법정에서의 심리학 활용을 위해 매우 필요한 연구일 수 있다.

여러 가지 의미있는 결과에도 불구하고 몇 가지 점에서 본 연구의 결과에 대한 해석에는 제한점을 두어야 한다. 가장 먼저 고려해야 할 부분은 연구에 포함되었던 피조사자의 숫자가 매우 적었다는 점이다. 따라서 이 논문에서 산출된 변별기준점은 현장에서 참고할 보조자료 정도로 고려하되, 실제로는 개별 피해자의 진술 내용과 피해가 발생한 환경적인 특징에 더 많은 주의를 기울여야 할 것이란 점이다. 즉 이 연구에서 산출된 변별기준점을 피해 진술의 진위판단에 있어 절대적인 근거로 사용해서는 안 될 것이다. 또 한 가지 고려해야 할 사항은 본 연구에 포함되어 분석되었던 문항들은 최혜림과 이수정(2009)의 연구에 근거하여 문항분석에 있어 가장 변별력이 있었던 문항들만을 토대로 하였다는 사실이다. 물론 후속연구에서는 이처럼 일부 제한된 문항 뿐 아니라 나머지 문항들을 대상으로 하여서도 문항특성을 분석해야 할 것이며 이와 함께 척도들의 변별기준점에 대한 탐구는 지속적으로 검증이 되어야 할 것이다.

세 번째 고려해야 할 사항은 본 연구에 포함되었던 준거집단에 관한 문제이다. 두 집단은 평균 연령 상에 차이가 있었는데, 허위 진술 집단이었던 통제집단에 포함되었던 피조사자의 연령이 실제 피해자 집단의 피조사자들보다 조금 더 높았다. 이런 차이가 발생했던 이유는 모사집단의 대상자에서 초등학생을 배제하고자 했기 때문인데, 이 같은 연령상의 차이는 이 나이 또래의 아동들에게 있어 언어의 유창성 상에 의미있는 차이를 유발한다고 한다(송명자, 2008). 허나 집단 간의 연령차가 연구의 의미있는 결과를 희석하지는 않는다고 판단되는데, 그 이유는 실제 피해를 당했던 집단의 피조사자들의 연령이 통제집단의 연령보다 더 어렸다는 사실이다. 즉 실제 피해자들은 연령이 어려 언어적인 유창성이 부족함에도 불구하고 진술의 내용 상 더욱 풍부한 사실을 전달할 수 있었다. 이는 상상이나, 고의적인 왜곡에 의해서 성폭력 피해를 허위 진술하는 것이 진짜 성폭력 피해에 대한 진술과 유의하게 구분이 가능하

는 사실을 보다 강력히 시사한다.

그러나 무엇보다도 우선적으로 신중하게 고려해야 하는 사항은 본 연구에서 사용된 진술분석의 도구는 아동의 진술내용을 토대로 하여 실제적 진실에 접근하고 있다는 사실이다. 만일 만에 하나 진술을 하는 아동의 기억이 보호자나 그 이외의 원인에 의해 원천적으로 왜곡이 되었고 그런 사실에 대해 아동이 전혀 인지하지 못하는 상황(Loftus & Davies, 1984)이라면 왜곡된 기억에서 유래한 진술의 신빙성을 진술 내용에 대한 분석만으로 판단하기에는 한계가 있을 수밖에 없다. 이와 같은 형태의 기억 왜곡 가능성은 진술에 대한 내용분석만으로는 완전히 배제할 수 없다. 본 연구의 결과를 해석함에 있어 이 문제는 꼭 유의해야 하는 지점일 것이다.

참 고 문 헌

- 김현정 (2010). 성폭력피해아동의 진술분석 도구의 상대적 유용성 비교연구. 경기대학교 박사학위 청구논문.
- 박광배 (1995). 법심리학. 정민사.
- 송명자 (2008). 발달심리학. 학지사.
- 이수정 (2009). 아동피해자조사매뉴얼. 2009년도 진술분석전문가 교육 자료집. 경기대학교 · 여성부.
- 이수정 (2009). 아동 대상 성범죄 조사절차의 미비점 및 개선방안 연구: 진술타당성분석(Statement Validity Analysis)의 형사사법적 활용가능성, 한국경찰연구, 8(3), 141-172.
- 조은경 (2004). 성폭력 피해 아동의 진술타당성 분석 및 활용방안에 관한 연구. 한국형사정책연구원.
- 최혜림, 이수정 (2009). 성폭력 피해 진술에 대한 신빙성 분석도구의 문항반응특성분석. 한국심리학회지: 여성, 14(4), 591-616.
- Bourg, S., Connor, E., & Landis, E. (1995). The impact of expertise and sufficient information on psychologists' ability to detect malingering. *Behavioral Sciences and the Law*, 13, pp.505-515
- Boyчук, T. D. (1991). *Criteria-based content analysis of children's statements about sexual abuse: A field-based validation study*. Unpublished doctoral dissertation, Arizona State University, Tempe.
- Boyчук, T. D., Esplin, P. W., & Raskin, D. C. (1988, June 18). Children's statements of sexual abuse: Validity of criteria-based content analysis. Presented at the NATO Advanced Study Institute on Credibility Assessment, Maratea, Italy. Op. Cit. in M. Steller (1989), Recent developments in statement analysis. In J. C. Yuille (Ed.), *Credibility assessment* (pp. 135-154), NATO Advanced Science Institute Series, Boston: Kluwer Academic Publishers.
- Faller, K. C. (1991) Possible explanations for child sexual abuse in divorce. *American Journal of Orthopsychiatry* 61(1). 86-91.
- Faller, K. C. (2007). *Interviewing children about sexual abuse* Oxford: Oxford University Press.
- Fisher R., & Geiselman R. E. (1992), *Memory-enhancing techniques for investigative interviewing: The cognitive interview*. Charles Thomas Publisher, U.S.A.
- Gardner, R. A. (1987). *The parental alienation syndrome and the differentiation between fabricated and genuine child sex abuse* Creskill, NJ: Creative Therapeutics.
- Honts, C. R. (1994). Assessing children's credibility: Scientific and legal issues in 1994. *North Dakota Law Review*, 70, 879-903.
- Kapardis, A. (2006). *Psychology and law*. New York: Cambridge University Press.
- Johnson, M. K. and Raye, C. L. (1981). Reality monitoring. *Psychological Review*, 88(1), 67-85.
- Lamb, M. E., Sternberg, K. J., Esplin, P. W., Hershkowitz, I., Orbach, Y., & Hovav, M. (1997). Criterion-based content analysis: A field validation study. *Child Abuse and Neglect*, 21, 255-264.
- Loftus, E. F., & Davies, G. M. (1984). Distortions in the memory of children. *Journal of Social Issues*, 40, 52-67.
- Miller, H. A. (2001). *M-FAST: Miller-forensic assessment of symptoms test professional manual*. Odessa, FL: Psychological Assessment Resources.

- Raskin, D. C., & Esplin, P. W. (1991). Statement validity assessment: Interview procedures and content analysis of children's statements of sexual abuse. *Behavior Assessment*, 13, 265-291.
- Rogers, R. (1984). Towards an empirical model of malingering and deception. *Behavioral Sciences and the Law*, 2, 93-112.
- Rogers, R. (1997). Current status of clinical methods. In Richard Rogers (Ed.), *Clinical assessment of malingering and deception* (2nd ed. pp.301-327). New York: Guilford.
- Smith, G. P. (1997). *Assessment of malingering with self-report instruments*. In R. Rogers (Ed.), *Clinical assessment of malingering and deception (2nd ed)* (pp 351-372). New York: Guilford.
- Sporer, S. L. (1997). The less travelled road to truth: verbal cues in deception detection in accounts of fabricated and self-experienced events. *Applied Cognitive Psychology*, 11, 373-397.
- Strömwall, L. A., & Granhag, P. A. (2003). How to detect deception? Arresting the beliefs of police officers, prosecutors and judges. *Psychology, Crime and Law*, 9, 19-36.
- Swets, J. A. (1988) Measuring the accuracy of diagnostic systems. *Science*, 240, 1285 - 1293.
- Undeutsch, U. (1982). Statement reality analysis. In A. Trankell(ed.), *Reconstructing the past: The role of psychologists in criminal trials*. Stockholm: P.A. Norsted and Sons.
- Undeutsch, U. (1989). The development of statement reality analysis. Paper presented at the NATO Advanced Study Institute on credibility assessment in Maratea, Italy, June 1988.
- Vrij, A. (2004). Criterion-based content analysis: A Qualitative review of the first 37 studies. *Psychology, Public Policy and Law*, 11(1), 3-41.
- Vrij, A. (2005). Criteria-based content analysis: A qualitative review of the first 37 studies. *Psychology, Public Policy, and Law*, 11, 3-41.
- Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2002). Will the truth come out? The effect of deception, age, status, coaching, and social skills on CBCA scores. *Law and Human Behavior*, 26, 261 - 283.
- Yuille, J. C. (1988). The systematic assessment of children's testimony. *Canadian Psychology*, 29, 247 - 262.

A Validation Study of Statement Validity Tests for Child Victims of Sexual Molestation

Lee Soo Jung
Kyonggi University

This study validated statement validity test tools such as CBCA, RM, and a contextual validity test of sexual molestation, used at ONE-STOP centers in South Korea. Furthermore, this study attempted to search for the cutoff scores of these instruments for judicial decision making. This study applied a simulation study with two criterion groups (Smith, 1997), a real sexually victimized group and feigned group. These 2 criterion groups were compared based on the scores of three dependent measures. Total scores of items were analysed, which were proven to have good item characteristics based on the study of Choi and Lee (2009). Total scores of three scales differentiated real sexual victims from feigned subjects. The most effective cutoff scores were found to be 13 of CBCA consisted of 11 items, 9 of RM consisted of 8 items, and 9 or 10 of a contextual validity test consisted of 14 items.

Keywords : statement validity analysis, CBCA, RM, contextual validity test, ROC, differentiating power

1차원고 접수일 : 2010년 3월 12일
수정원고 접수일 : 2010년 5월 18일
게재 확정일 : 2010년 5월 28일