

성폭력 피해 진술에 대한 신빙성 분석도구의 문항 반응 특성 분석

최혜림 이수정[†]

경기대학교 범죄심리학과

본 연구에서는 2009년 3월에서 8월 사이에 서울 및 경기 지역에 있는 원스탑센터에 배치된 전문가들이 분석한 50 케이스의 진술 분석 보고서에서 진술의 신빙성 분석 구성 내용 중 준거기반내용분석(CBCA), 사실성평가(RM), 성폭력발생맥락평가, 행동변화평가, 타당성 평가에 대하여 분석하였다. 우선 내적 합치도에 대하여 각 구성 내용의 항목들을 분석하여 낮거나 부적합한 내적 합치도를 가진 항목들을 제외한 후 남은 항목들에 대하여 문항 반응 이론(IRT) 분석을 실시하고 변별도와 난이도를 산출하였다. 그 결과 성폭력발생맥락평가의 세 문항을 제외한 모든 항목이 어느 정도 이상의 변별도를 나타내었다. 그러나 이 연구에서는 분석에 사용한 사례의 수가 적고 대부분의 피해 사례가 1회성 혹은 단기 피해인 이유로 모든 항목에 대한 타당한 분석이 이루어지지 못하였다. 이 점에 대해 더 많은 사례를 포함하여 보완하는 것은 후속연구에서 기대할 수 있을 것이다.

주요어 : 진술타당성분석도구, 내적 합치도, 문항반응이론, 문항특성곡선

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

최근 발생한 조두순 사건을 시작으로 아동에 대한 성폭력 사건에 전 국민의 관심이 쏠리고 있다. 또한 국민들은 아동 성폭력 사건의 높은 발생률과 낮은 구속률, 낮은 기소율을 접하고 있다. 실제로 경찰청의 통계 분석에 따르면 한 해 성범죄 발생 건수는 2002년도 6,119건이었으나 2005년도 7,316건, 2008년도 9,883건으로 매해 높은 증가율을 보이고 있는 반면 구속률과 기소율은 점차 감소하는 경향을 보인다. 또한 성폭력 범죄는 다른 범죄에 비해 신고율이 낮은 범죄로 알려져 있다. 이러한 낮은 신고율의 원인이 성범죄의 특성과 우리나라의 정서상 피해사실이 공개되었을 때 도리어 피해자들이 가질 수 있는 수치심이나 사회의 잘못된 눈초리만은 아닐 것이다. 신고를 하더라도 구속·기소에서 낮은 비율을 기록하는 사법 체계도 일정 부분의 책임이 있다 할 수 있다. 그러나 국내 현황을 놓고 볼 때 이 같은 상황이 쉽게 변할 수는 없을 것이다. 국내 경찰 및 검찰과 법원의 업무부하량을 고려해볼 때 성범죄의 피해자가 다만 아동이라 하여 사건의 처리에 무조건 우선성을 둘 수도 없는 노릇일 것이며 그렇다고 범죄 사실에 대한 충분한 입증 절차 없이 피의자를 무조건 구속부터 할 수도 없는 일일 것이다(이수정, 2009).

현재 성폭력 범죄와 관련하여 사법부의 가장 큰 문제는 피고인에 대한 유죄 판단이 과도하게 신체적 상해와 DNA 등 물적 증거에만 의존하고 있다는 점이다(이수정, 2009). 그러나 아동 성폭력 범죄는 적은 수의 사건에서만 의학적 증거가 존재한다(Myers, 1998). 또한 아동 성폭력 범죄는 많은 경우 기존의 성인을 상대로 한 강간죄와는 매우 다른 방식으로 발생(이수정, 2009)하기 때문에 사법부가 주로 의

존하는 물적 증거가 남아있지 않을 가능성이 높다. 그로 인해 많은 아동 성폭력 사건에서 증거의 부족으로 전문가에게 대한 아동의 초기 진술이 가장 강력한 증거를 구성한다(Myers, 1998). 그러나 또 다른 문제는 기존의 수사 방식이 아동 피해자가 진술을 하는 과정에서 이차 피해를 발생시킨다는 것이다. 반복되는 회상 요구로 아동 피해자의 정신적·심리적 고통은 둘째치더라도 성범죄 피해로 인해 이미 위축된 피해자를 일방적으로 다그쳐 진술하게 한다거나, 또는 아동의 피압성성은 전혀 고려치 않은 채 유도신문을 하여 피해사실에 대한 기억조차 왜곡하게 되는 등의 실수가 국내 사법 제도 내에서는 비일비재하게 발생하고 있다(민유숙, 2009). 매우 어린 아동 혹은 약간 더 나이가 많은 아동일지라도 모두 다른 때 경험했거나 본 것과 일치하는 세부 설명이나 행동을 설명하는 잘못된 질문에 쉽게 유도된다(Pipe, Lamb, Orbach, & Esplin, 2004; Faller, 2007). 잘못된 수사 관행으로 인해 왜곡된 피해자의 피해 기억은 판결에서 신뢰성과 진실성을 인정받지 못하여 가해자에게 정당한 처벌을 내리게 하지 못할 가능성이 매우 높아진다.

성폭력 범죄 관련 사법부에서 발생할 수 있는 또 다른 문제는 법정에서 아동이 피해 사실에 대해 증언하도록 하는 것으로 이는 종종 아동에게 정신적 충격을 줄 수 있다(Ross, Lindsay, & Marsil, 1999). 법정에서 가해자를 마주함으로써 피해 아동은 또다시 공포를 경험할 수 있다. 이러한 문제를 피하기 위한 방법으로는 아동과 가해자 사이에 가림막을 설치하는 것(Coy v. Iowa, 1988), 법정 외의 장소에서 증언을 녹화하여 법정에서 보여주는 것(McGough, 1995; Ross 외, 1999), 아동의 어머

니, 부모, 경찰, 의사, 응급실 간호사 등을 포함한 전문 증인(hearsay witness)을 이용한 증거를 인정하는 것(Quas 외, 1996; Ross 외, 1999) 등이 있다. 미국의 법정에서는 전문 증인을 이용하는 방법이 다른 두 방법보다 더 자주 사용되나, 이 방법은 몇 가지 심각한 법적인 문제를 가져오는데 그 문제점으로는 피고인의 제 6조 수정헌법의 대질 권리를 위반한다는 것, 아동에 대한 반대 심문과 가능한 의혹에 대한 방어 기회 및 판결에 영향을 줄 수 있는 직접적 심리에 대한 배심원의 행동 증거 획득을 박탈한다는 점이며 마지막으로 배심원이 피고인으로부터 아동이 보호받을 필요가 있다고 결정함으로써 유죄 판결을 의도할 수 있다는 점 등이 있다(Ross 외, 1999). 또한 전문 증인은 배심원에게 아동이 연약하다는 인상을 줌으로써 정보의 출처, 즉 아동에 대해 잠재적으로 낮은 신뢰도를 갖게 하여 아동의 증언 시보다 낮은 유죄 판결을 갖게 할 수 있다는 점과 전문 증언의 신뢰도 자체에 대한 의심에서 문제점을 안고 있다(Ross 외, 1999).

이렇듯 여러 문제점을 가진 전문 증언을 보완하기 위한 방법으로 전문가 증언을 사용할 수 있다. 미국의 경우, 검사는 아동 증언의 진실성을 높이기 위한 시도로써 아동 성학대에 따른 아동의 행동적 반응에 대한 전문가 증언을 선호하고 있다(Kovera, Gresham, Borgida, & Gray, 1997) 전문가 증인은 일반적인 경험을 넘어선 대상에 대한 특별한 지식이나 기술을 갖춘 자를 의미한다(캘리포니아 증거 조례 720). Kovera 등(1997)의 연구에 의하면 전문가의 증언은 유죄 판단에 영향을 미친다. 구체적인 전문가 증언은 전문가 부재 조건과 관례적인 전문가 조건과 비교하였을 때 유죄 선고가 유의미하게 증가하였다(Kovera 등, 1997;

Gabora, Spanos, & Joab, 1993).

이처럼 외국의 경우 전문가 증인이 재판 시 중요한 역할을 한다는 것이 입증되고 있다. 한국의 상황을 보면 2009년도 경찰청은 아동 진술의 신빙성을 높여주고 유죄 판단에 도움이 되도록 성폭력 범죄 전문 조사 기관인 윈스탑 센터 중 서울과 경기 지역의 다섯 곳에 피해자 조사 전문가 및 진술분석 전문가를 배치하여 경찰이 13세 미만 아동 피해자와 지적 장애에 피해자를 조사할 경우 개입하여 피해자 진술을 분석하여 사건 송치 시 함께 제공하게 하고 있다.

현재 윈스탑 센터에 배치받아 피해자 조사를 돕고 있는 전문가들은 최소한 해당 분야에 대한 풍부한 전문 지식과 사실에 대한 객관성 및 중립성에 대한 자격 요건을 갖추기 위하여 배치 전 일정 기간 교육을 수료하며 윈스탑 센터 지원 시 성범죄 피해아동 진술에 관하여 신빙성 분석을 실시하고 있다. 전문가는 준거기반내용분석(Criterion-Based Content Analysis; CBCA)과 사실성 평가(Reality Monitoring; RM) 등의 진술타당성분석을 통해 아동의 진술에 대한 신빙성을 분석한다. 2009년 12월 현재 윈스탑 센터에 배치받은 전문가들이 사용하고 있는 아동 성폭력 피해자 진술의 신빙성 분석 구성 내용은 표 1과 같이 제시하였다.

본 연구의 목적은 이러한 전문가의 진술 분석에 쓰이는 이 신빙성 분석 도구의 타당성을 확인하는 것이다. 표 1에서 제시한 각 구성 내용은 그 하위 척도에 관련한 여러 문항을 포함하고 있으며 연구에서 진행된 타당성 분석의 대상은 오리엔테이션을 제외한 준거기반내용분석(CBCA)과 사실성평가(RM), 성폭력발생가능성평가(맥락타당성), 행동변화평가, 타당성 평가의 하위 척도 다섯 개와 각 척도에 속

표 1. 아동 성폭력 피해자 진술의 신빙성 분석 구성 내용

조항	구성 내용
오리엔테이션	오리엔테이션은 성폭력 피해 진술의 핵심이 되는 주요 개념들을 아동이 어느 정도 이해하고 있으며, 어떠한 방식으로 표현하고 있는가를 파악하기 위해 실시하는 도구이다. 아동의 연령 별 발달 특성을 감안하여 유아용과 학령기 아동용으로 구분하였고 모든 내용은 그림으로 보다 재미있게 제시된다는 장점을 가지고 있다. 진술능력을 확인하기 위해 포함된 주요개념은 안과 밖, 신체 부분 명칭, 숫자, 진실과 거짓, 만지다의 총 6개 항목이다.
준거 기반 내용 분석	준거기반 내용분석(CBCA: Criteria - Based Content Analysis)은 5개의 범주에 19개의 하위 준거로 구성되어 있으며, 아동의 인지적 능력과 동기적 요소를 고려하여 아동 진술의 진실성을 평가한다. 개별 준거들은 진술의 진실성 지표로서 하위로 꾸며낸 진술보다 진실한 진술에서 더 많이 나타날 것임을 전제로 하고 있다. 각 0, 1, 2점으로 채점되는데, 총 38점 중 19점 이상이면 진술에 신빙성이 있다고 보아야 한다.
사실성 평가	사실성 평가는 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 채점하여 총 16점 중 8점 이상이면 진술에 신빙성이 있다고 본다.
성폭력 발생 맥락 평가	성폭력 발생 가능성 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 범주 하에 구성된 25개 문항에 대해 0점 혹은 1점으로 반응할 때, 총 25점 중 13점 이상이면 성폭력 발생 가능성이 농후하다고 판단한다.
행동 변화 평가	성폭력 피해 아동의 행동변화에 대한 평가는 성폭력 피해 아동에게서는 또래 아이들보다 더 조속한 성적인 행동이 나타난다. 따라서 아동의 성적 행동에 대한 보호자의 평가를 근거로 9가지 범주 27항목에서의 행동변화를 탐지한다. 문항은 전혀 없으면 0점, 한 달에 한 번이면 1점, 일주일에 한 번이면 2점으로 채점하여 총 27점 중 14점 이상이면 성폭력 피해가 있다고 본다.
타당성 평가	타당성 평가는 이혼이나 재산권 다툼 등으로 인한 부모의 영향, 면담자의 암시 등을 검토함으로써 진술의 진위여부를 검토하고, 피해아동이 평소 성격환상을 갖고 있었는지, 진술을 꾸며내는 것인지, 사건과 무관하게 다른 사람을 허위 지목하는 것인지 등을 종합적으로 고려하여 아동진술의 타당성을 평가하도록 하고 있다. 따라서 피해 아동의 진술 진위 여부, 정신적 상태, 피해 아동의 진술의 의도성 여부, 허위지목 가능성 등 4가지 범주 하에 26개 문항에 대해 0점, 1점, 2점의 3점으로 반응하여 총 52점 중 26점 이하이면 아동의 진술이 외부의 영향을 받지 않았다고 평가한다.

출처: 이수정(2009).

한 문항들이다. 또한 연구에 쓰인 자료는 실제 윈스탑 센터에 배치된 전문가들의 분석 결과로, 내적 합치도와 문항 반응 이론을 사용하였다. 이 결과는 나아가 아직 완전하지 않은 성범죄 피해 진술 분석 도구를 더 효과적이고 적은 문항으로 피해 사실을 예측할 수 있도록 보완하는 데 기여할 것으로 기대한다.

연구방법

연구대상

본 논문에서 사용한 피해자조사 자료는 2009년 3월 이후 아동 성폭력 피해자 조사시 전문가 참여제를 실시하고 있는 서울 및 경기도 지방에 위치한 윈스탑 센터에 배치된 피해자 조사 전문가가 성폭력 피해자에 대한 진술 분석 사례 50건의 조사 보고서에서 모집하였다. 사례 50건의 기관 비율은 표 2와 같다.

피해자 집단은 만 4세에서 만 26세 범위의 성폭력 피해자이며 초기 데이터는 50명이었으나 보고서에 결측값을 포함한 다섯 케이스를 제외한 마흔다섯 케이스를 분석하였다. 분석에 쓰인 케이스의 45명 중 44명은 여성이며 1명의 강간 피해자는 만 12세의 남자아동이다. 지적장애 여부로는 지적장애 1급 1명, 2급 5명, 3급 4명으로 총 지적장애 10명과 일반 35명이며 피해 유형으로는 성폭행 미수가 2명, 강제 추행이 27명, 강간이 16명이었다. 피해를 당한 기간으로는 1회성이 31명, 6개월 미만의 단기 피해 4명, 6개월 이상의 장기 피해 10명이고 가해자의 유형으로 불상자 22건, 편부모의 동거인 포함 지인이 18건, 법적으로 친족에 해당하는 가해자 5건이었다.

표 2. 아동성폭력피해자 조사시 전문가 참여제 시범 실시 기관별 대상 사건 수

	빈도	%
경찰병원	7	14.0
보라매병원	2	4.0
아주대병원	10	20.0
의정부의료원	3	6.0
인천의료원	28	56.0
합계	50	100.0

표 3. 피해자연령

	빈도	퍼센트	유효 퍼센트	누적 퍼센트
만 10세	4	8.9	8.9	8.9
만 11세	6	13.3	13.3	22.2
만 12세	7	15.6	15.6	37.8
만 13세	1	2.2	2.2	40.0
만 14세	2	4.4	4.4	44.4
만 15세	2	4.4	4.4	48.9
만 16세	1	2.2	2.2	51.1
만 17세	1	2.2	2.2	53.3
만 22세	2	4.4	4.4	57.8
만 26세	1	2.2	2.2	60.0
만 4세	1	2.2	2.2	62.2
만 6세	2	4.4	4.4	66.7
만 7세	3	6.7	6.7	73.3
만 8세	6	13.3	13.3	86.7
만 9세	6	13.3	13.3	100.0
합계	45	100.0	100.0	

분석도구

위와 같은 자료로 본 연구에서는 두 가지 분석 방법을 사용하였다. 첫 번째는 각 구성 내용과 사례값을 이용하여 내적 일관성을 알아보기 위해 Cronbach's α 를 구하였다. Cronbach's α 계수는 모든 문항 각각이 동일한 개념을 측정하는지 알아보기 위한 내적 일관성 측정 방법 중의 하나이다. 모든 문항의 α 계수와, 한 문항씩을 제외한 나머지 문항들의 α 계수를 비교하는 것이다. 만일 한 문항을 제외한 값이 전체 문항의 값보다 높다면 제외된 한 문항은 다른 문항들의 내적 일관성 신뢰도를 저하시키고 있다는 결론을 얻을 수 있다. 본 연구에서는 내적 일관성을 위한 전체 문항의 α 계수와 각 문항에 대한 내적 일관성을 분석하였다. 그리고 그 결과치의 α 계수들 중 문항과 총점 간 상관이 부적 관련성을 지니거나 0.1을 기준으로 그보다 낮아 0에 가까운 문항들을 제외하여 Cronbach's α 지수를 상승시키고 내적 일관성이 높은 문항들을 남겨 두었다.

두 번째로 문항 분석을 통해 내적 일관성을 검증한 후 남은 문항들을 다시 문항 반응 이론(Item Response Theory; IRT)으로 분석하였다. 문항 반응 이론(Item Response Theory; IRT)은 문항과 척도를 분석하고 심리학적 측정 도구를 개발, 실시하며 심리학의 개념(예를 들어 우울 등)을 개별적으로 평가하기 위해 사용하는 수리학적 모델과 통계적 방법의 집합이다 (Reise, Ainsworth, & Haviland, 2005). 그러나 고전검사이론(Classical Test Theory; CTT)처럼 모든 문항에 동일한 비중을 두는 것이 아니라 IRT는 측정하고자 하는 주제에 따라 각 문항 간의 비중과 측정비율을 달리하는 이론이다. 간

단히 말하면 IRT는 하위척도의 각 문항들이 얼마나 하위척도의 주제를 잘 측정하는가를 분석하는 것이다. IRT의 장점은 크게 다섯 가지로 구분할 수 있는데 시행한 문항의 수와 관계없이 응답자의 특질을 평가한다는 점과 더 큰 모집단의 응답자 표본과 관계없이 문항 지수(예를 들어 변별도, 난이도)를 평가하는 특징, 다른 방식의 검사 수행을 비교가능하다는 점, 시행하지 않은 문항에 대한 피험자의 수행을 예측할 수 있는 점, 그리고 각각의 검사 점수의 정확도를 예측할 수 있다는 것이다 (Gibbons, Immekus, & Bock, 2007).

IRT에서는 각 문항의 비중을 보여주는 난이도와 변별도를 그래프로 제시하는데 이를 문항 특성 곡선(Item Characteristic Curve; ICC)이라고 한다. 이 문항 특성 곡선(ICC)은 각 문항에 대한 반응 비율과 개인의 특질의 수준을 보여주는데 X축은 잠재 특질의 수준이며 Y축은 각 개인이 그 문항을 맞추거나 조건을 충족할 확률이다. 이 ICC는 각 항목 당 구성된 그래프이며 여기서 얻을 수 있는 정보로는 난이도와 변별도가 있다. ICC에서 난이도(difficulty)란 어려운 수준에 따른 각 단계에 관한 개인의 수치이다(Reise 외, 2005). 즉, 한 항목에 대하여 세로축이 50%의 확률일 때 가로축의 값이 어느 수준인가를 말해주는 것이다. 다음으로 변별도(discrimination)란 ICC에서 난이도 값의 기울기를 일컫는다. 변별도가 더 높은 항목들은 한 항목의 난이도 주변 특질 범위 내에 속한 개인들을 구별하기 더 좋다(Reise 외, 2005). 즉, 한 항목이 특질에 관하여 개인들을 얼마나 잘 구별하는가에 관한 것이다. 만약 그래프의 기울기가 완만하다면 그 항목을 맞추거나 조건을 충족하는 것이, 측정하길 원하는 주제와 연관이 있는지와 별로 관련이 없다는 것이다.

전형적으로 난이도는 -2에서 2 사이의 값을 가지며 변별도는 0에서 2 사이의 값을 가진다 (Hagman, Kuerbis, Morgenstern, Bux, Parsons, & Heidinger, 2009). 본 연구에서 다루는 진술타당성 분석 도구는 성취도 평가처럼 정답이 따로 있는 성격의 도구가 아니기 때문에 분석 결과에서 난이도보다는 변별도를 더 중시하였다. 또한 난이도와 변별도의 p 값은 보통 0.05 이하일 때 유의하다고 보지만 연구의 사례 수가 45개로 많은 수가 아님을 고려하여 0.10 내외

의 경우에도 어느 정도 유의하다고 해석하였다.

연구 결과

문항 분석

준거기반내용분석(CBCA)의 문항에 대해 내적 일관성을 분석한 결과는 표 4에 제시하였다.

표 4. CBCA에 대한 문항 분석

	문항 삭제 시 척도 평균	문항 삭제 시 척도 분산	문항 총점 간 상관	문항 삭제 시 α
논리적 구성	16.00	17.70	.239	.515
진술 비구조화	16.89	17.79	.035	.549
세부정보의 양	16.04	17.04	.346	.498
맥락적 깊이	16.09	16.21	.445	.477
상호작용	16.38	15.85	.446	.470
대화 재현	16.51	14.78	.533	.441
예기치 않은 상황	17.11	16.53	.216	.511
일상적이지 않은 정보	16.60	16.07	.279	.497
세부 정보 과잉	16.89	16.76	.125	.532
이해하지 못한 정보	17.23	16.88	.164	.522
외적 사실 연합	16.60	17.46	.067	.544
주관적 경험	16.30	16.00	.445	.473
가해자 정신상태 추정	17.51	17.08	.269	.505
자발적 수정 혹은 첨가	17.15	18.00	-.004	.559
기억 부족 시인	16.70	18.95	-.122	.578
자신의 진술 의심	17.79	18.74	-.014	.537
자기 비난	17.83	18.75	.000	.535
가해자 용서	17.62	18.81	-.075	.553
전형적 범죄 특징 진술	17.70	18.48	.020	.539

분석 결과 19문항 중 문항과 총점 간 상관이 부적 관련성을 지니거나 0.100보다 작은 여덟 개의 문항(진하게 표시)을 제외한 후 내적 일관성을 분석하자 Cronbach α 지수가 제외 전 .54에서 제외 후 .70까지 향상하였다. 이 결과에 따라서 이후 문항 반응 이론 분석에서는 19문항 대신 내적 일관성이 높은 11문항을 중심으로 분석하였다. 11개의 문항으로 계산한 새 총점의 평균은 13.47이었으며 표준편차는 3.97이었다.

표 5는 RM, 즉 사실성 평가의 하위 문항들에 대하여 문항 분석한 결과이다. 분석 결과 사실성 평가의 여덟 항목 중에 문항과 총점 간 상관이 부적 관련성을 지니거나 0에 가까웠던 문항은 없었으므로 이후 문항 반응 이론 분석에서도 여덟 개 문항 모두 그대로 사용하였다. 총 8문항인 사실성 평가의 Cronbach α 지수는 .64였으며 총점의 평균은 12.15이며, 표준편차는 2.57이었다.

다음에 제시한 표 6은 아동과의 면담, 부모 면담, 기타 정보 등을 종합하여 전체적인 맥락 속에서 판단하여 아동 진술의 진위 여부를

가능하는 데 유용한 성폭력발생가능성평가의 하위 문항들에 대한 문항 분석 결과이다.

성폭력발생가능성평가에 속한 25문항 중 분석 결과 문항과 총점 간 상관이 부적 관련성을 지니거나 거의 0에 가까운 문항들은 총 8개(진하게 표시)였다. 이 문항들을 제외하자 Cronbach α 지수는 .63에서 .77까지 향상하였다. 따라서 이후 문항 반응 이론 분석에서는 총 25문항 중에서 높은 내적 일관성을 지닌 17문항을 중심으로 분석하였다. 새로운 17문항에 대한 데이터 결과에서 평균은 9.84이었으며 표준편차는 3.10을 기록하였다.

성폭력 피해 아동의 행동변화에 대한 평가 분석은 표 7에서 제시하였다. 행동변화평가의 문항은 총 27문항이나 분석에 쓰인 50 케이스에서 모두 0점을 기록한 여덟 문항을 우선 제외하고 19개의 문항만을 가지고 문항 분석을 실시하였다.

우선 제외한 문항 외 19개의 문항 분석 결과 문항과 총점 간 상관이 부적 관련성을 지니거나 거의 0에 가까운 문항들은 6개였다. 이 문항들을 추가로 배제한 후 다시 분석한

표 5. RM에 대한 문항 분석

	문항 삭제 시 척도 평균	문항 삭제 시 척도 분산	문항 총점 간 상관	문항 삭제 시 α
명료성	10.43	5.12	.47	.58
지각정보	11.04	4.96	.34	.61
공간정보	10.34	5.49	.51	.59
시간정보	10.53	5.47	.25	.63
재연가능	10.34	5.40	.41	.60
정동	10.79	4.95	.34	.61
사실적 묘사	10.34	5.66	.40	.61
인지적 작동	11.38	5.02	.22	.66

표 6. 성폭력발생가능성평가에 대한 문항 분석

	문항 삭제 시 척도 평균	문항 삭제 시 척도 분산	문항 총점 간 상관	문항 삭제 시 α
아동의 관점 반영	10.29	8.58	.52	.60
성폭력에 대한 구체적 진술	10.39	8.33	.43	.60
발달단계 보다 많은 성지식	11.12	9.15	.08	.63
발생장소	10.27	8.74	.51	.60
발생시간	10.39	8.28	.45	.60
가해자의 유인방법	10.55	8.38	.29	.61
가족이 있었던 곳	10.82	8.53	.21	.62
아동의 옷	10.61	8.03	.40	.59
가해자의 옷	10.55	8.17	.37	.60
특징적 사건 존재	10.82	8.40	.26	.61
아동의 정서적 상태 진술	10.39	8.24	.47	.59
비밀유지의 강요	10.94	8.39	.30	.61
아동의 자발적 보고	10.41	8.46	.34	.60
사건을 들은 사람의 반응	10.73	8.66	.16	.63
아동이 사건을 부인	11.14	9.50	-.11	.64
진술과정 나타난 아동의 감정	10.39	8.12	.53	.58
아동의 성행동	11.16	9.18	.10	.63
아동의 발달단계	11.16	9.60	-.18	.65
의학적 증거	11.02	9.10	.05	.64
물적 증거	11.12	8.86	.24	.62
가해자의 자백	11.16	9.14	.13	.63
목격자의 진술	11.02	9.27	-.02	.64
또 다른 피해자의 존재	11.00	9.29	-.03	.65
부모가 관찰한 내용	10.78	9.30	-.06	.66
전문가에 의한 증거	11.16	9.89	-.37	.66

Cronbach α 지수는 제외하기 전 .76에서 제외 후 분석에서 .81으로 상승하였다. 따라서 이후에는 27문항 대신 13문항을 중심으로 문항 반

응 이론 분석하였다. 13문항의 데이터로 계산한 새 총점의 평균은 .96이었으며 표준편차는 2.50이었다.

표 7. 행동변화평가에 대한 문항 분석

	문항 삭제 시 척도 평균	문항 삭제 시 척도 분산	문항 총점 간 상관	문항 삭제 시 α
어머니 또는 다른 여성의 가슴을 만지려고 함	1.10	5.33	.51	.73
자신의 몸을 사람이나 물체에 비벼댐	1.19	6.33	.22	.75
자신의 입을 어머니나 다른 여성의 가슴에 갖다 댐	1.15	5.32	.74	.71
어른에게 생식기를 보여줌	1.17	6.18	.17	.76
여자(남자)아이가 남자(여자)의 옷을 입으려고 함	1.17	6.10	.23	.75
손으로 자신의 생식기를 만짐(자위행위)	1.13	5.57	.44	.73
집에 있을 때 자신의 생식기를 만짐	1.13	6.03	.21	.75
어른이 뽀뽀하거나 안아줄 때 화들짝 놀람	1.08	5.40	.53	.73
성행위소리를 냄(신음소리, 거친 숨소리)	1.17	6.14	.33	.75
유혹적으로 말함	1.19	6.33	.22	.75
다른 아이의 생식기를 만짐	1.06	5.17	.47	.73
다른 아이나 어른과 성관계를 가지려고 함	1.19	6.50	-.01	.76
어른의 생식기를 만짐	1.17	6.44	.03	.76
동물의 생식기를 만짐	1.15	6.34	.04	.77
뽀뽀할 때 자신의 혀를 다른 사람의 입 속에 넣으려고 함	1.15	6.43	.02	.76
어른의 옷을 벗기려고 함(바지나 셔츠 등)	1.13	5.09	.69	.71
포레아이에 비해 성에 대해 많이 알	1.19	6.54	-.07	.76
벗은 몸의 사람을 뚫어지게 쳐다봄	1.10	4.90	.76	.70
알몸 또는 반쯤 벗은 사람의 사진을 뚫어지게 쳐다봄	1.17	6.44	.03	.76

타당성 평가의 분석에서도 행동변화평가에 서처럼 총 26문항 중에 모든 케이스가 0점을 기록한 11문항을 우선 제외하였다. 그리고 나머지 15문항에 대한 분석 결과는 표 8에서 제시하였다.

26문항 중 모든 케이스가 0점인 11문항을

제외한 15문항의 분석 결과에서, 문항과 총점 간 상관이 부적 관련성을 지니거나 거의 0에 가까운 문항은 ‘부모의 피학대경험으로 아동의 경험을 지나치게 부풀림’이라는 문항 하나였다. 이 문항을 제외하고 14문항만으로 다시 분석하자 Cronbach α 지수는 .73에서 .75로 높

표 8. 타당성 평가에 대한 문항 분석

	문항 삭제 시 척도 평균	문항 삭제 시 척도 분산	문항 총점 간 상관	문항 삭제 시 α
부모의 생각과 느낌을 강요	1.24	3.56	.64	.69
가족구성원의 생각과 느낌을 강요	1.29	4.13	.22	.73
아동이 부모의 눈치를 보며 진술	1.12	3.49	.40	.72
관련 없는 신체적 증상을 성폭력으로 오해	1.31	4.26	.12	.74
부모의 피해대경험으로 아동의 경험을 지나치게 부풀림	1.33	4.43	-.09	.75
부모가 아동에게 특정한 내용진술을 의식적으로 강요	1.29	3.88	.49	.71
의도는 나타나지 않으나 부모가 특정한 내용 진술 요구	1.31	3.84	.66	.70
아동이 부모로부터 진술에 대한 훈련/연습 흔적	1.31	3.84	.66	.70
조사자가 피해아동에게 특정 내용 진술을 유도	1.10	3.64	.35	.72
면담자가 피해아동에게 특정 내용 진술을 유도	1.27	3.95	.34	.72
피해아동이 조사자의 눈치를 보며 진술	1.22	3.89	.31	.72
아동의 과거/현재 정신 병력	1.31	3.68	.59	.69
아동이 학대경험으로 자신의 경험을 지나치게 부풀림	1.33	4.31	.12	.74
소통불능	1.18	3.65	.29	.73
비교의 거짓말	1.27	4.03	.18	.74

아졌다. 이 결과에 따라 이후 분석에서는 26 문항 대신 14문항을 중심으로 분석하였다. 데이터에서 타당성 평가의 걸러진 14문항에 대한 새 총점의 평균은 1.33이었으며 표준편차는 2.10이었다.

내적일관성 분석 결과 다섯 개의 하위 척도는 각각 높은 상관을 가진 문항만을 남겨 두었으며 그 문항과 각 하위 척도 총점 간

Cronbach's α 점수를 새로 산출하였다. 각 척도 간 상관관계와 내적 일관성에 대한 결과는 표 9에 정리해두었다.

ICC 분석

각 하위 척도의 문항들에 대한 IRT 분석을 실시하기 전에 각 척도의 새로운 총점을 문항

표 9. 척도 간 상관관계 및 내적 일관성

		CBCA	RM	성폭력 발생 맥락	피해아동의 성행동 변화	타당성 평가
CBCA	Pearson's <i>r</i>	.70				
	<i>n</i>	47				
RM	Pearson's <i>r</i>	.610***	.64			
	<i>n</i>	47	47			
성폭력 발생가능성 평가	Pearson's <i>r</i>	.615***	.471**	.77		
	<i>n</i>	47	47	49		
피해아동의 성행동 변화	Pearson's <i>r</i>	.193	.026	-.322	.81	
	<i>n</i>	45	45	47	48	
타당성 평가	Pearson's <i>r</i>	-.143	-.455***	-.081	-.086	.75
	<i>n</i>	46	46	48	48	49

※ 각 척도의 신뢰도 대각행렬에 제시

으로 간주하여 각 척도가 아동 진술의 신빙성을 얼마나 잘 측정하는지 M-PLUS를 통해 분석하였으며 그 결과는 표 10, 그림 1에 제시하였다. 분석을 위하여 각 하위 척도는 내적 일관성이 높은 나머지 항목의 평균을 중심으로 0과 1로 나누어 이원 변수로 재코딩하였다. 각 척도값은 해당 문항의 평균보다 높은 경우 1로, 낮은 경우 0으로 나누었으나 타당성 평가는 문항 내용의 특성상 높은 경우 0으로, 낮은 경우 1로 역코딩하였다.

표 10은 ICC의 변별도와 난이도를 숫자로 나타낸 것이다. 그래프의 기울기인 변별도 지수를 보면 CBCA와 RM, 맥락타당성은 값이 1 이상으로 높은 변별도값을 보여주고 있으며 CBCA의 *p*값은 0.131, RM은 0.049, 맥락타당성은 0.083으로 유의미하다. 타당성평가의 변별도 지수는 1에는 미치지 못하며 0.290의 *p*값을 갖는다. 행동평가의 변별도 지수는 0.089, *p*값은 0.732로 낮은 변별도를 보이며 유의미하지 않다. 한편 난이도의 값을 보여주는 난이도

표 10. 하위 척도 다섯 개의 변별도 및 난이도

	변별도 지수 (S.E.)	변별도 <i>p</i>	난이도 지수 (S.E.)	난이도 <i>p</i>
CBCA	1.837 (1.218)	0.131	-0.159 (0.214)	0.456
RM	1.452 (0.738)	0.049	0.034 (0.227)	0.882
맥락타당성	1.401 (0.808)	0.083	-0.765 (0.292)	0.009
행동변화	0.089 (0.259)	0.732	8.654 (25.193)	0.731
타당성평가	0.296 (0.280)	0.290	-2.192 (2.037)	0.282

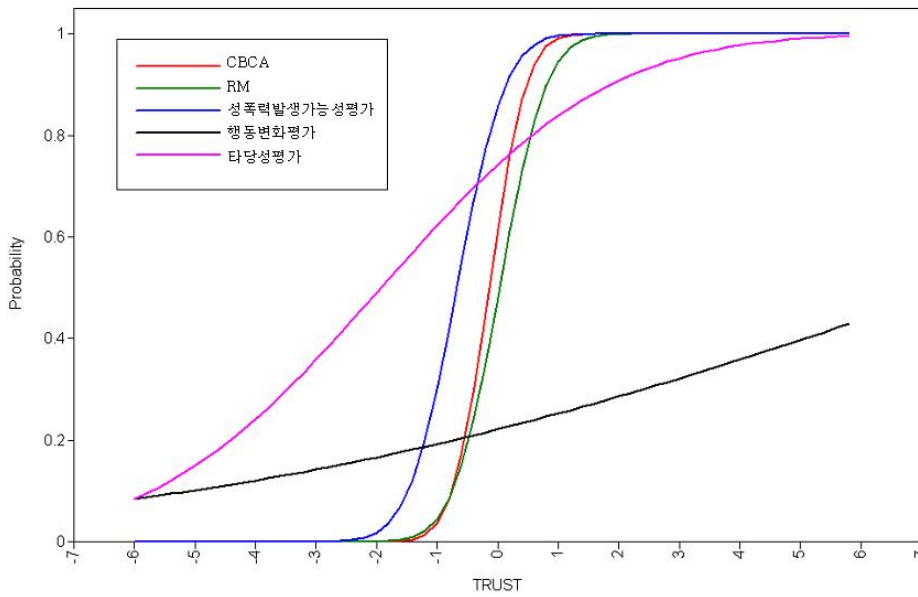


그림 1. 하위 척도 다섯 개의 ICC

지수를 보면 CBCA와 RM, 맥락타당성은 0을 중심으로 근처에 머물러 있으며 맥락타당성의 p 값 0.009만이 유의미하다. 타당성평가는 난이도 지수가 -2.192이고 0.282의 p 값으로 유의미하지 않다. 행동변화는 8 이상의 난이도와 0.731의 p 값을 가져 유의미하지 않다. 이러한 결과들은 그림으로 보는 것이 더 분명하게 파악할 수 있다. 하위 척도의 ICC는 그림 1에 제시하였다. 표에서 해석한 것과 동일하게 그림에서 보이듯이 CBCA와 RM, 성폭력발생가능성평가(맥락타당성)은 가파른 기울기, 즉 좋은 변별도를 지니지만 타당성평가는 그보다 완만한 기울기의 그래프이며 행동변화평가척도의 곡선은 매우 완만하다.

각 하위척도를 한 문항으로 하는 분석을 마친 후 각 척도의 해당 문항들을 가지고 ICC를 분석해 보았다. 우선 CBCA(준거기반 내용분석)는 총 19문항에서 내적 합치도를 통해 남은 11문항에 대한 자료만을 사용했다. 자료에는

각 문항에 대해 0, 1, 2점으로 값을 매겨져 있기 때문에 M-PLUS에 사용하기 위해 없음을 0 점, 있음을 1점으로 이원 분류하여 재코딩 하였다. 이에 대한 변별도와 난이도, ICC는 표 11과 그림 2로 제시하였다.

표 11은 CBCA의 11문항에 대한 변별도와 난이도를 표로 제시한 것이다. ‘대화 재현’, ‘주관적 경험’, ‘상호작용’은 1이 넘는 높은 변별도 지수를 보여주고 있으나 세 항목 중 ‘주관적 경험’만이 0.079의 p 값으로 유의미하다. ‘가해자 정신상태 추정’, ‘세부 정보 과잉’, ‘이해하지 못한 정보’의 변별도 지수는 0.3 이하로 낮은 편이며 p 값은 세 항목 모두 0.2 이상으로 유의미하지 않다. 난이도 지수는 대부분의 항목이 0의 근처에 분포하여 있으나 간혹 ± 3 의 범위 밖의 결과가 있으며 난이도에 있어 유의미한 p 값을 지닌 항목은 11항목 중 여섯 개이다. 그림 2는 CBCA에서 높은 내적 일관성을 보인 11문항에 대한 문항 반응 곡선이다.

표 11. CBCA의 11문항에 대한 변별도 및 난이도

	변별도 지수 (S.E.)	변별도 <i>p</i>	난이도 지수 (S.E.)	난이도 <i>p</i>
논리적 구성 (문항 1)	0.739 (0.512)	0.148	-3.381 (2.186)	0.122
세부정보의 양 (문항 3)	0.730 (0.387)	0.059	-2.886 (1.259)	0.022
맥락적 깊이 (문항 4)	0.747 (0.478)	0.118	-2.252 (1.081)	0.037
상호작용 (문항 5)	1.847 (1.485)	0.213	-1.532 (0.475)	0.001
대화 재현 (문항 6)	2.428 (2.139)	0.256	-0.827 (0.262)	0.002
예기치 않은 상황 (문항 7)	0.557 (0.243)	0.022	0.287 (0.411)	0.484
일상적이지 않은 정보 (문항 8)	0.313 (0.291)	0.282	-1.865 (1.728)	0.280
세부 정보 과잉 (문항 9)	0.274 (0.265)	0.302	-0.529 (0.858)	0.537
이해하지 못한 정보 (문항10)	0.291 (0.325)	0.371	1.327 (1.527)	0.385
주관적 경험 (문항12)	1.880 (1.070)	0.079	-1.526 (0.416)	0.000
가해자 정신상태 추정 (문항13)	0.213 (0.180)	0.237	3.672 (3.495)	0.293

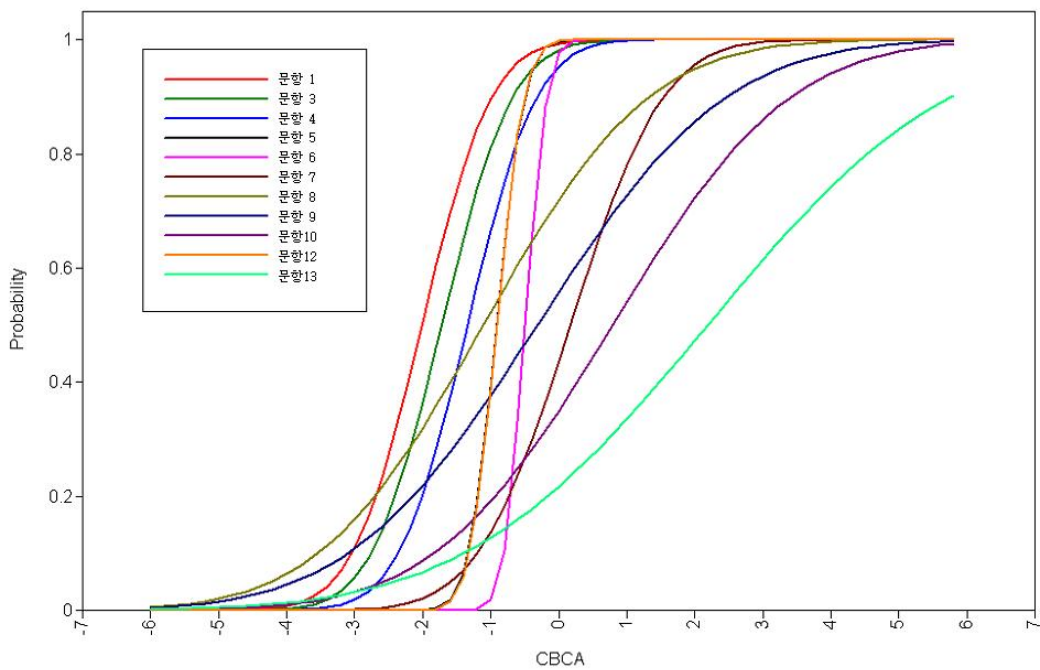


그림 2. CBCA의 11문항에 대한 ICC

표에 정리한 결과를 한 눈에 확인할 수 있으며 변별도 지수가 높은 ‘대화 재현(B223)’과 ‘주관적 경험(B235)’, ‘상호작용(B222)’은 가파른 기울기를 가지고 있음을 볼 수 있다. ‘주관적 경험’과 ‘상호작용’의 경우 변별도 지수와 난이도 지수가 아주 근소한 차이를 가지고 있

을 뿐이어서 그래프 상에서 거의 겹쳐서 구분이 어렵다.

다음으로 RM(사실성 평가)의 여섯 문항에 대한 변별도와 난이도, ICC는 표 12와 그림 3으로 제시하였다. RM은 내적 일관성을 검증하기 위한 문항 분석에서 여덟 개의 문항 모두

표 12. RM의 6문항에 대한 변별도 및 난이도

	변별도 지수 (S.E.)	변별도 <i>p</i>	난이도 지수 (S.E.)	난이도 <i>p</i>
명료성 (문항 1)	1.272 (0.249)	0.000	-2.164 (0.346)	0.000
지각정보 (문항 2)	0.363 (0.255)	0.154	-2.467 (1.594)	0.122
시간정보 (문항 4)	1.126 (0.514)	0.029	-2.008 (0.363)	0.000
재연가능 (문항 5)	0.996 (0.413)	0.016	-2.411 (0.403)	0.000
정동 (문항 6)	0.837 (0.403)	0.038	-1.731 (0.639)	0.007
인지적 작동 (문항 8)	0.367 (0.334)	0.272	-0.243 (0.580)	0.676

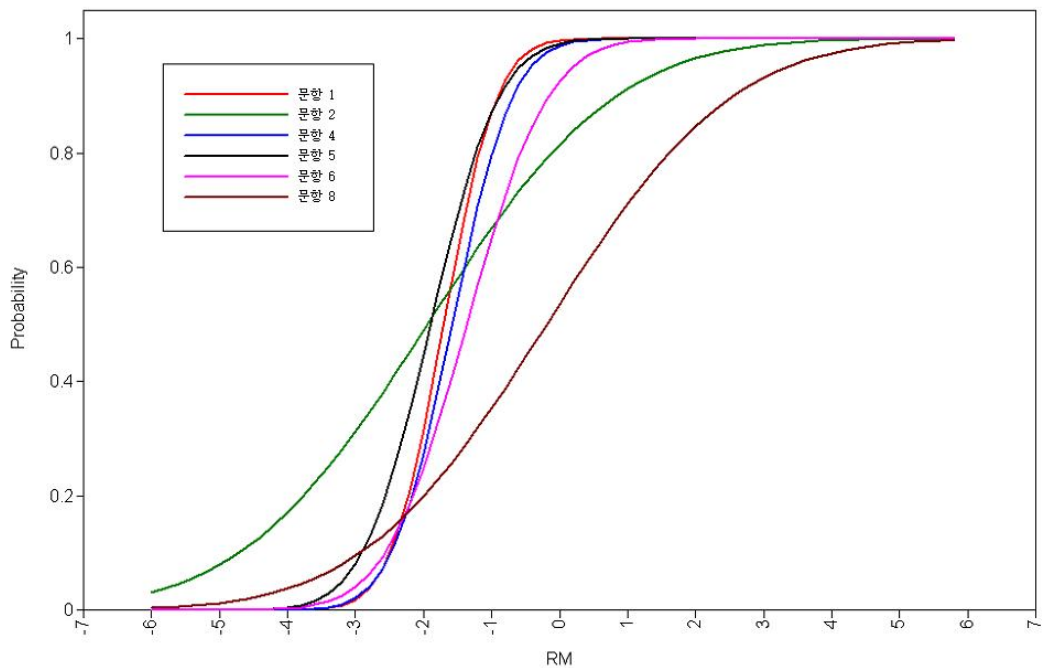


그림 3. RM의 6문항에 대한 ICC

높은 내적 일관성을 보였다. 이 하위 척도 역시 없음을 0점, 약간 있음과 풍부하게 있음을 1점으로 이원 분류하였다. 그러나 RM의 항목 중 문항 8의 인지적 작동은, 상상의 경험은 내부적 처리과정을 거치기 때문에 감각적 정보보다 생각이나 추론이 더 많이 나타나게 되기 때문에 높을수록 그 진술이 허구일 가능성이 높은 항목(김현정, 2010)이므로 없음을 0점을 1로, 약간 있음과 풍부하게 있음을 0으로 역코딩하여 분석하였다. 문항 반응 이론으로 분석한 결과 ‘공간정보(C23)’와 ‘사실적 묘사(C27)’는 모든 사례의 변수값이 1이어서 분석

에 대한 결과를 도출할 수 없어 분석 대상 항목에서 제외하고 나머지 여섯 개의 문항만으로 결과를 산출하였다.

표 12는 RM의 6문항에 대한 변별도와 난이도이다. ‘명료성(C21)’, ‘시간정보(C24)’, ‘재연가능(C25)’, ‘정동(C26)’의 변별도 지수는 모두 1 이상 혹은 1에 가까우며 p 값 역시 네 항목 모두 0.05 이하로 매우 유의미한 것을 볼 수 있다. ‘지각정보(C22)’와 ‘인지적 작동(C28)’의 변별도 지수는 각각 0.363, 0.367이고 변별도 p 값은 0.154, 0.272이다. 난이도 지수를 보면, 유의미한 변별도를 지닌 항목과 동일한 항목들이

표 13. 맥락타당성의 17문항에 대한 변별도 및 난이도

	변별도 지수 (S.E.)	변별도 p	난이도 지수 (S.E.)	난이도 p
아동의 관점 반영 (문항 1)	2.534 (1.438)	0.078	-1.614 (0.409)	0.000
성폭력에 대한 구체적 진술 (문항 2)	1.761 (0.687)	0.010	-1.165 (0.280)	0.000
발생장소 (문항 4)	1.491 (0.981)	0.128	-2.048 (0.807)	0.011
발생시간 (문항 5)	1.172 (0.531)	0.027	-1.215 (0.361)	0.001
가해자의 유인방법 (문항 6)	0.793 (0.247)	0.001	-0.596 (0.318)	0.061
가족이 있었던 곳 (문항 7)	0.231 (0.182)	0.205	1.385 (1.408)	0.325
아동의 옷 (문항 8)	0.971 (0.262)	0.000	-0.447 (0.273)	0.102
가해자의 옷 (문항 9)	0.802 (0.247)	0.001	-0.592 (0.315)	0.060
특정적 사건 존재 (문항10)	0.667 (0.187)	0.000	0.456 (0.371)	0.219
아동의 정서적 상태 진술 (문항11)	1.659 (0.580)	0.004	-1.079 (0.273)	0.000
비밀유지의 강요 (문항12)	0.507 (0.158)	0.001	1.378 (0.685)	0.044
아동의 자발적 보고 (문항13)	1.686 (0.854)	0.048	-1.074 (0.283)	0.000
사건을 들은 사람의 반응 (문항14)	0.425 (0.176)	0.016	0.214 (0.495)	0.666
진술과정 나타난 아동의 감정 (문항16)	1.530 (0.656)	0.020	-1.104 (0.298)	0.000
아동의 성행동 (문항17)	-0.026 (0.223)	0.906	-56.966 (471.326)	0.904
물적 증거 (문항20)	0.013 (0.217)	0.950	111.361 (1808.382)	0.951
가해자의 자백 (문항21)	0.014 (0.215)	0.946	103.582 (1551.066)	0.947

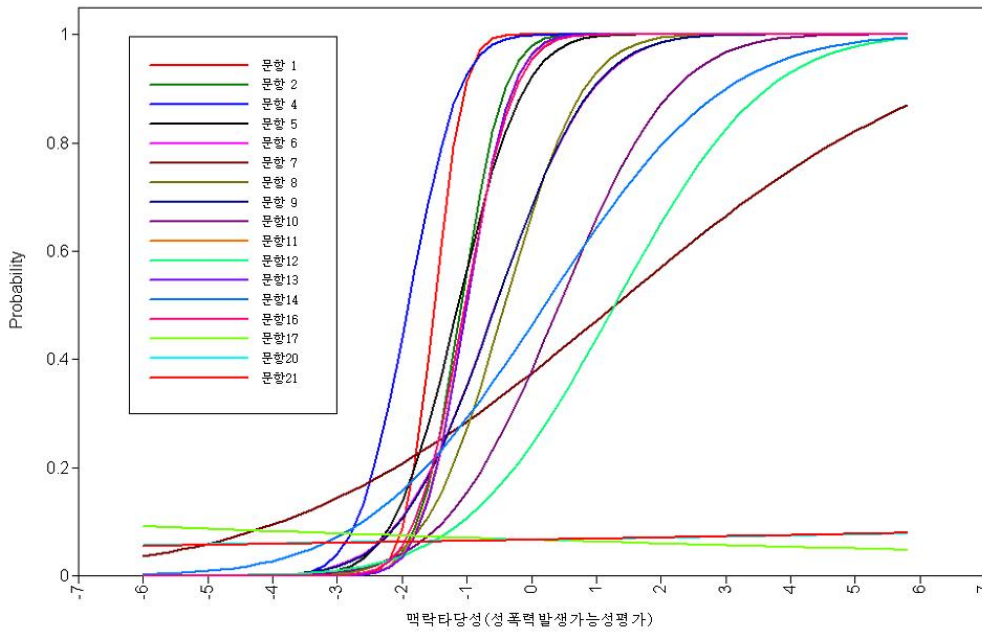


그림 4. 맥락타당성의 17문항에 대한 ICC

마찬가지로 유의미한 난이도를 가졌다. 표 12에 이어서 그림 3은 이 결과를 그래프로 옮긴 ICC이다. 변별도 지수가 유의미한 네 항목의 기울기가 급한 것이 분명하게 보이며 다른 두 항목은 상대적으로 기울기가 완만함을 알 수 있다.

다음으로 맥락타당성, 즉 성폭력발생가능성 평가는 총 25문항에서 문항 분석을 통해 내적 일관성이 높다는 결과를 도출하여 남은 17문항에 관한 자료만을 사용했다. 이 구성 도구는 각 문항에 대해 해당할 경우 1점, 해당하지 않을 경우에는 0점으로 나뉘어 기록한다. 따라서 따로 이원분류 재코딩을 할 필요 없이 자료의 값을 코딩하여 분석에 그대로 사용하였다. 이에 대한 변별도와 난이도, ICC는 표 13과 그림 4로 제시하였다.

맥락타당성의 경우 1 이상의 변별도 지수를

갖는 항목이 7개이며 특히 ‘아동의 관점 반영’은 2.534로 매우 높은 변별도 지수를 가진다. 그러나 ‘아동의 성행동’과 ‘물적 증거’, ‘가해자의 자백’은 매우 낮은 변별도 지수를 가지며 특히 ‘아동의 성행동’은 변별도가 음수이다. 변별도에 관한 p 값은 낮은 변별도 지수의 세 항목과 더불어 ‘가족이 있었던 곳’, ‘발생장소’의 값이 각각 0.906, 0.950, 0.946, 그리고 0.205, 0.128로 유의미하지 않으며 나머지 12개 항목은 대부분 0.05 이하 또는 0.1 이하의 p 값으로 유의미한 결과가 나왔다. 난이도 지수 역시 낮은 변별도를 가진 세 항목에서 매우 큰 결과값과 유의미하지 않은 p 값을 기록하였다. 그 외 난이도 지수에서 유의미하지 않은 p 값은 ‘사건을 들은 사람의 반응’, ‘가족이 있었던 곳’, ‘특정적 사건 존재’의 세 항목에서도 나타났다. 그림 4는 맥락타당성, 즉 성폭력

발생가능성평가의 17항목에 대한 변별도와 난이도를 그래프로 나타낸 것이다. 표에서 확인한 것과 마찬가지로 변별도가 0에 가까운 ‘물적 증거(D242)’와 ‘가해자의 자백(D243)’은 X축의 바로 위에서 살짝 올라가는 모습의 선을 그리고 있으며 두 항목의 변별도 지수가 0.001의 차이로 아주 근소하여 두 선이 거의 겹쳐져있는 것을 볼 수 있다. ‘아동의 성행동(D231)’의 경우 변별도, 즉 기울기가 음수이기 때문에 X축의 바로 위에서 오른쪽으로 갈수록 축에 가까워지고 있다. 그 다음으로 0.231의 변별도 지수를 가진 ‘가족이 있었던 곳(D224)’ 항목이 완전한 대각선의 그래프를 보인다.

다음은 성폭력 피해 아동에게서 또래 아이

들보다 더 조숙한 성적 행동이 나타나는지를 판단하는 성폭력 피해 아동의 행동변화 평가 항목에 관한 IRT 분석 결과이다. 행동변화평가는 총 27문항이나 모든 사례에서 0점을 기록한 여덟 문항을 제외한 19개 문항으로 문항 분석을 실시했으며 내적 일관성이 없어 문항과 총점 간 상관이 작거나 부적인 문항 6개를 제외하고 13문항에 대한 분석을 실시하였다. 이 척도는 아동의 보호자가 아동의 최근 행동을 회상하여 직접 기입하는 방식으로 행동변화가 없으면 0점, 행동변화가 한 달에 한 번이면 1점, 일주일에 한 번이면 2점을 부여한다. 각 문항에 대해 0, 1, 2점으로 나뉘어 값을 매겼기 때문에 M-PLUS에 사용하기 위해

표 14. 행동변화의 13문항에 대한 변별도 및 난이도

	변별도 지수 (S.E.)	변별도 <i>p</i>	난이도 지수 (S.E.)	난이도 <i>p</i>
어머니 또는 다른 여성의 가슴을 만지려고 함 (문항 1)	1.694 (0.950)	0.075	1.743 (0.428)	0.000
자신의 몸을 사람이나 물체에 비벼댐 (문항 2)	1.844 (1.196)	0.123	2.286 (0.228)	0.000
자신의 입을 어머니나 다른 여성의 가슴에 갖다 댐 (문항 3)	3.411 (1.856)	0.066	1.773 (0.392)	0.000
어른에게 생식기를 보여줌 (문항 4)	1.778 (0.884)	0.044	2.306 (0.292)	0.000
여자(남자)아이가 남자(여자)의 옷을 입으려고 함 (문항 6)	1.790 (1.578)	0.257	2.302 (0.171)	0.000
손으로 자신의 생식기를 만짐 (자위행위) (문항 8)	1.434 (0.963)	0.136	2.074 (0.502)	0.000
집에 있을 때 자신의 생식기를 만짐 (문항 9)	0.639 (0.318)	0.044	2.787 (0.825)	0.001
어른이 뽀뽀하거나 안아줄 때 화들짝 놀람 (문항11)	0.989 (0.473)	0.037	1.736 (0.544)	0.001
성행위 소리를 냄 (신음소리, 거친 숨소리) (문항13)	1.212 (0.436)	0.005	2.206 (0.556)	0.000
유혹적으로 말함 (문항14)	1.778 (0.884)	0.044	2.306 (0.292)	0.000
다른 아이의 생식기를 만짐 (문항15)	1.174 (0.643)	0.068	1.770 (0.565)	0.002
어른의 옷을 벗기려고 함 (바지나 셔츠 등) (문항22)	2.109 (1.071)	0.049	1.883 (0.436)	0.000
벗은 몸의 사람을 뚫어지게 쳐다봄 (문항25)	2.302 (1.169)	0.049	1.637 (0.439)	0.000

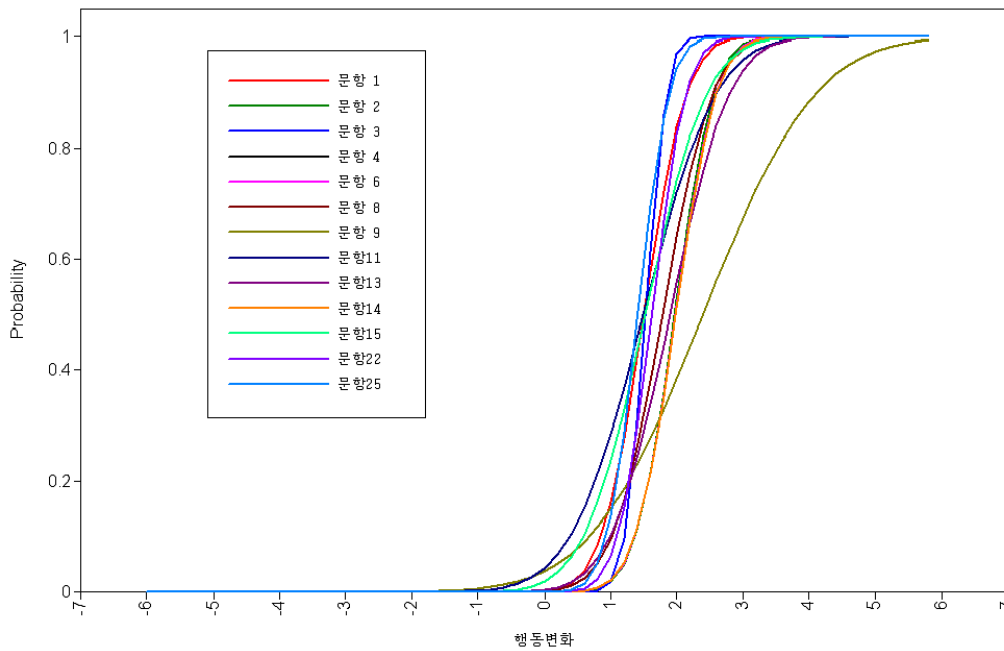


그림 5. 행동변화의 13문항에 대한 ICC

없음을 0점, 있음을 1점으로 이원 분류하여 재코딩 하였다. 이에 대한 변별도와 난이도, ICC는 표 14와 그림 5에서 제시하였다.

행동변화에 대한 변별도 지수는 대부분이 1 점 이상이며 가장 낮은 변별도 지수는 ‘집에 있을 때 자신의 생식기를 만짐(E28)’으로 0.639 점을 기록하였으며 ‘어른이 뽀뽀하거나 안아 줄 때 화들짝 놀람(E211)’이 0.989점의 변별도 지수를 보였다. 변별도 지수 결과를 보면 2점 이상(벗은 몸의 사람을 뚫어지게 쳐다봄, E225; 어른의 옷을 벗기려고 함(바지나 셔츠 등), E222)과 3점 이상(자신의 입을 어머니나 다른 여성의 가슴에 갖다 댐, E23)의 항목도 있다. 변별도에 대한 각 항목의 p 값은 ‘여자(남자)아이가 남자(여자)옷을 입으려고 함(E26)’의 0.257, ‘손으로 자신의 생식기를 만짐(자위 행위)(E28)’의 0.136, ‘자신의 몸을 사람이나 물

체에 비벼댐(E22)’의 0.123을 제외하고는 모두 0.1 이하의 p 값으로 유의미한 결과를 갖는다. 난이도 지수는 모두 1점 혹은 2점 이상이며 난이도에 대한 p 값은 대부분 0.000, 혹은 0.001, 0.002로 유의미하다. 표 14에 나타난 사실들은 그림 5에서 뚜렷이 확인할 수 있다. 그림에서 항목들은 거의 유사한 수준으로 높은 변별도와 난이도를 가지고 있음을 보인다. 그 중 0.639의 변별도 지수를 가진 ‘집에 있을 때 자신의 생식기를 만짐(E29)’ 항목이 다른 항목들에 비해서 완만한 기울기를 보여주고 있다.

마지막으로 타당성 평가 척도에 관한 IRT 결과이다. 타당성 평가는 총 26문항이나 모든 사례에서 0점을 기록한 11개 문항을 제외한 15개 문항으로 문항 분석을 실시했으며 내적 일관성이 없어 문항과 총점 간 상관관이 부적인

문항 하나를 제외한 14개의 문항에 대한 IRT 분석을 실시하였다.

이 척도는 아동 진술의 타당성을 평가하기 위한 것으로 부모의 영향, 면담자의 암시 등을 검토하거나 아동의 진술을 분석하여 아동 진술을 종합적으로 고려하는 것이다. 각 항목은 0점, 1점, 2점의 3점 척도로 구성되어 있는데 총점이 낮을수록 아동의 진술이 외부의 영향을 받지 않아 아동진술의 타당성이 높다고 해석할 수 있다(아동피해자조사보고서, 2009). 분석을 위하여 본 연구에서는 모든 항목을 없

음을 1점, 있음 및 풍부하게 있음을 0점으로 이원 분류한 후 역코딩 하였다. 이에 대한 변별도와 난이도, ICC는 표 15와 그림 6에서 제시하였다.

분석 결과 ‘부모의 생각과 느낌을 강요(F2111)’ 문항에서 잔여 공분산 행렬(Residual Covariance Matrix)이 양적 확장을 이루지 못하여, 즉 알고자 하는 척도와 데이터 간 관계가 전혀 없는 것으로 나타나 분석이 되지 않았다. 이 항목에 관하여서는 결론 및 논의 부분에서 설명하였다. 그 외의 13항목에 대해 아래의

표 15. 타당성평가의 13문항에 대한 변별도 및 난이도

	변별도 지수 (S.E.)	변별도 <i>p</i>	난이도 지수 (S.E.)	난이도 <i>p</i>
가족구성원의 생각과 느낌을 강요 (문항 2)	0.681(0.391)	0.082	-2.668(0.886)	0.003
아동이 부모의 눈치를 보며 진술 (문항 3)	1.051(0.502)	0.036	-1.162(0.431)	0.007
관련 없는 신체적 증상을 성폭력으로 오해 (문항 4)	1.035(0.676)	0.126	-2.366(0.555)	0.000
부모가 아동에게 특정한 내용진술을 의식적으로 강요 (문항 9)	1.312(0.532)	0.014	-1.887(0.486)	0.000
의도는 나타나지 않으나 부모가 특정한 내용 진술 요구 (문항10)	2.307(1.402)	0.100	-2.191(0.662)	0.001
아동이 부모로부터 진술에 대한 훈련/연습 흔적 (문항11)	2.307(1.402)	0.100	-2.191(0.662)	0.001
조사자가 피해아동에게 특정 내용 진술을 유도 (문항12)	0.605(0.319)	0.058	-1.337(0.663)	0.044
면담자가 피해아동에게 특정 내용 진술을 유도 (문항13)	0.727(0.416)	0.080	-2.292(0.901)	0.011
피해아동이 조사자의 눈치를 보며 진술 (문항14)	0.818(0.350)	0.020	-1.928(0.539)	0.000
아동의 과거/현재 정신 병력 (문항18)	2.307(1.402)	0.100	-2.191(0.662)	0.001
아동이 학대경험으로 자신의 경험을 지나치게 부풀림 (문항19)	2.183(3.859)	0.572	-2.211(0.454)	0.000
소통불능 (문항21)	0.417(0.307)	0.174	-2.888(1.858)	0.120
비교의 거짓말 (문항24)	0.880(0.459)	0.055	-2.272(0.613)	0.000

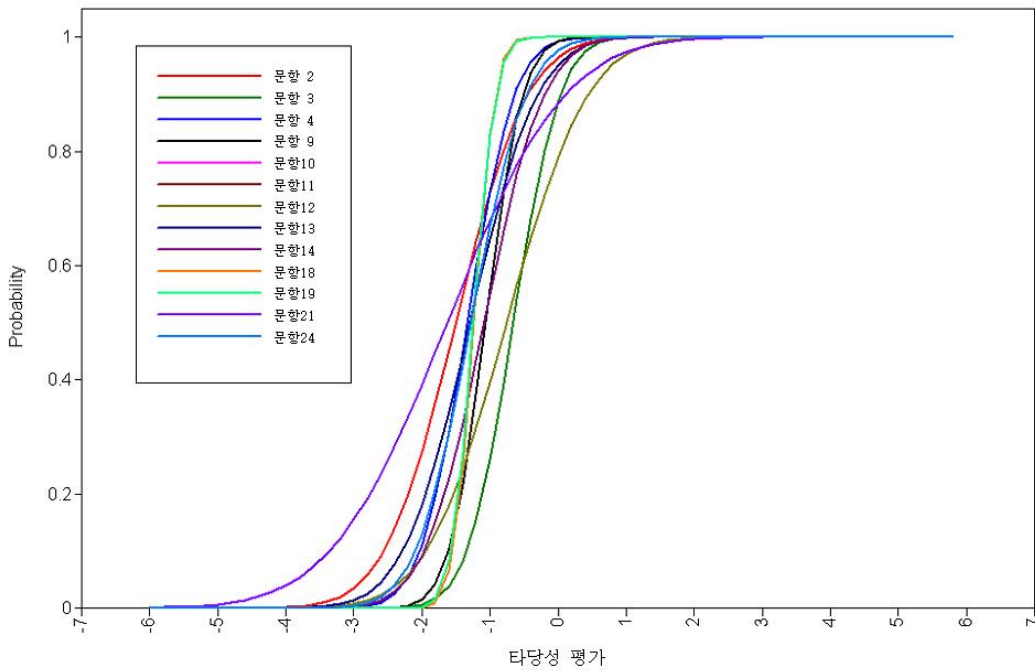


그림 6. 타당성 평가의 14문항에 대한 ICC

표 15에서 변별도 지수를 보면 2점 이상의 변별도를 보이는 항목이 네 개, 1점 이상이 세 항목, 나머지 여섯 항목에서 0점에서 1점 사이의 변별도 지수를 보인다. 가장 높은 변별도 지수는 2.307점으로 세 항목(의도는 나타나지 않으나 부모가 특정한 내용 진술 요구, F2142; 아동이 부모로부터 진술에 대한 훈련/연습 흔적, F2143; 아동의 과거/현재 정신 병력, F2221)이 동일한 변별도 지수를 보이고 있다. 한편 가장 낮은 변별도 지수는 0.417로 ‘소통불능(F224)’ 항목이며 그 다음으로는 ‘조사자가 피해아동에게 특정 내용 진술을 유도(F2151)’ 항목이 0.605점을 기록했다. 변별도의 p 값은 ‘아동이 학대경험으로 자신의 경험을 지나치게 부풀림(F222)’이 0.572로 유의하지 않으며 더불어 ‘소통불능(F224)’과 ‘관련 없

는 신체적 증상을 성폭력으로 오해(F2121)’ 항목을 제외한 나머지 항목은 모두 0.100 이하의 p 값을 보여 유의하다고 할 수 있다. 난이도 지수는 모든 항목이 음수를 기록하였으며 ‘소통불능(F224)’의 항목이 가장 높은 0.120의 p 값을 보여 이 항목을 제외한 나머지 항목은 모두 유의하다. 이 분석 결과에 관한 ICC를 보여주는 그림 6에서 보면 모든 항목이 좁은 범위 내의 난이도를 가지고 있기 때문에 구별하기가 힘든 점이 있지만 ‘소통불능(F224)’ 항목의 그래프가 X축의 가장 왼쪽에서부터 완만하게 올라가는 것을 확인 할 수 있다. 또한 이 F224 항목의 그래프와 위쪽에서 만나 거의 겹치는 곡선은 ‘조사자가 피해아동에게 특정 내용 진술을 유도(F2151)’ 항목으로 다른 곡선의 기울기에 비해 완만한 것을

확인할 수 있다.

결론 및 논의

클 전반에 걸쳐 연구한 아동 성폭력 피해자 진술의 신빙성 분석 도구는 어떤 유형의 가해자로부터 어느 정도의 기간 동안 피해를 당한 경우 어떠한 증상이 반드시 나타나야 한다는 식의 정답이 있는 도구가 아니며 성폭행 피해에 있어서 그러한 정답은 있을 수 없다. 즉, 이러한 요소가 없기 때문에 진술이 틀린 것이 아니라 해당하는 요소가 있기 때문에 진술에 더 신빙성이 있다고 해석할 수 있는 것이다. 서론에서 언급했듯이 본 연구에서 이 도구와 사례를 분석하여 얻고자 하는 것은 이 도구가 피해 사실을 얼마나 잘 분별해내는가에 관한 것이다. 따라서 IRT 분석 결과에서 난이도와 변별도 중 변별도를 더 중시하였다. 또한 내적 일관성 분석 결과를 보면 척도별로 총점 간 상관이 낮거나 부적인 문항이 여러 개 있거나 IRT 분석에서 모든 사례의 해당 항목이 모두 0점이어서 배제한 항목이 종종 나타났다. 이에 대하여 항목이 상관이 없거나 변별하지 못하여 불필요하다고 해석하기 보다는 분석에 사용한 사례 수가 적기 때문에 항목의 내적 일관성과 변별도 및 난이도를 충분히 나타내지 못한 것으로 볼 수 있다.

하위 척도 다섯 개에 관한 IRT 분석 결과에서 성행동 변화 척도는 난이도와 변별도가 눈에 띄게 좋지 않은 결과를 보여주었는데 그 이유로는 이 분석에 쓰인 데이터에서 피해자의 대부분이 1회성 혹은 단기 피해를 받았기 때문에 피해 기간에 있어서 성행동에 변화가 발생할 정도의 영향을 주지 못하였다는 점을

들 수 있다.

CBCA의 IRT 분석 결과는 변별도 지수가 매우 넓은 범위에 걸쳐서 분포해있다. 그 중 가장 높은 변별도를 나타내는 것은 ‘대화 재현’으로 이 준거가 나타났을 경우 실제 성폭력 발생 사실의 신빙성을 높일 수 있는 문항이다. 그 이유로는 성폭력을 경험하지 않은 아동은 무엇이 일어났는지를 이야기할 때 축어적으로 재현할 가능성이 적기 때문이다(Raskin & Esplin, 1991). 그 다음으로 높은 변별도 지수는 ‘주관적 경험’으로 사건 당시 피해자 자신의 주관적인 인지 상태 또는 정서 상태에 대해 진술을 하는(아동 피해자 조사 매뉴얼, 2009) 경우 피해 사실을 뒷받침해줄 수 있다. ‘주관적 경험’과 유사한 정도의 높은 변별도 지수를 갖는 문항은 ‘상호작용’이다. 가해자와 피해 아동 사이의 상호작용이 진술에 나타날 경우 신빙성을 높이는 결과는 아동 대상 성범죄에서 아동을 유인할 때 아동에게 비폭력적으로 접근하는 애착형이 가장 빈번하게 사용되기 때문으로 볼 수 있다. 실제로 형사정책연구원에서 청소년을 대상으로 한 성폭력 범죄 400건을 분석한 결과 애착형이 282건으로 전체의 70.5%를 차지하였다(김지선, 박지선, 최수형, 2009). 그러나 이 세 항목 중 p 값은 ‘주관적 경험’만이 0.079로 유의미하다. 반대로 낮은 변별도 지수를 보이는 문항으로는 ‘가해자 정신상태 추정’, ‘세부 정보 과잉’, ‘이해하지 못한 정보’, 그리고 ‘일상적이지 않은 정보’의 문항이 있다. 특히 ‘가해자 정신상태 추정’ 문항이 낮은 변별도를 보이고 있는데 이는 문항의 특성상 상대적으로 순진한 아동이나 성인의 심리를 잘 이해하지 못하는 아동에 있어서는 나타나기가 더욱 어렵기(Raskin & Esplin, 1991) 때문이라고 볼 수 있다.

사실성 평가의 항목들에 대한 분석 결과 ‘명료성’이 1.272로 가장 높은 변별도 지수를 보여주었으며 0.000의 유의미한 p 값을 가지고 있다. 이는 진술에 있어 허위로 꾸민 것은 담고 있기 힘든 세세하고 분명한 묘사가 있을수록 진술이 믿을 만하다는 것을 보여준다. 더불어 진술 과정에서 구체적인 설명을 요구하는 질문에 대한 대답으로도 평가할 수 있다. 그 다음으로 ‘시간정보’ 항목이 유사하게 높은 변별도 지수와 유의미한 p 값을 보여주는데 이는 성폭력 발생 당시의 상황을 설명함에 있어서 아동의 일과와 관련하여 기억을 되짚는 과정에서 쉽게 나타날 수 있기 때문이다. 모든 값이 1이기 때문에 분석에서 제외한 ‘공간정보’ 역시 유사한 이유로 항상 진술에 나타났으며 아동이 진술할 때 가해자의 위치나 주변 상황을 설명하기 위해 꼭 나타나기 쉽기 때문이다. 반면 ‘시간정보’는 0인 값인 있는데 이는 성폭력이 발생한 후 오랜 기간이 지난 피해자의 경우 기억나지 않을 수 있기 때문이다. 사실성 평가에서 낮은 변별도 지수와 유의미하지 않은 p 값을 가진 항목으로는 ‘지각정보’와 ‘인지적 작동’이 있다.

맥락타당성 하위 척도 분석에서 ‘아동의 관점 반영’ 문항이 매우 높은 변별도 점수를 보이며 p 값 역시 유의미한데 이는 아동이 경험한 것을 아동의 관점에서 서술하는 진술이기 때문에 높은 변별도가 당연하다고 할 수 있다. 그러나 ‘아동의 성행동’과 ‘물적 증거’, ‘가해자의 자백’은 변별도 지수가 매우 낮거나 혹은 부적인데 그 이유로는 아동 대상 성폭력의 특징에 기인한 것을 들 수 있다. 아동을 대상으로 한 성폭력의 경우 물적 증거가 거의 발생하지 않는 특성을 가지고 있다. 또한 아동 성폭력 사건에서는 주로 아동이 유일한 증인

이다(Myers, 1998). 그러한 연유로 가해자가 순순히 자백하지 않는 것을 고려해야 할 것이다. 또한 ‘아동의 성행동’ 항목이 부적 변별도 지수를 기록한 것에 대하여, 본 연구에서 사용한 피해 사례의 유형 분석 결과 피해 기간이 6개월 이상의 장기 피해에 비해 1회성이 세 배에 달하였으며 1회성의 경우 발생 직후 진술이 이루어진 경우가 많았다. 그로 인해 ‘아동의 성행동’에 있어서 변화가 일어나기에는 부족하였다고 해석할 수 있다.

행동변화 척도의 항목들에 대한 변별도는 전반적으로 매우 높은 변별도 지수와 유의미한 p 값을 보였다. ‘집에 있을 때 자신의 생식기를 만짐’ 항목이 가장 낮은 0.639에 머물렀다. 그러나 이 항목의 변별도 지수는 척도 내 다른 항목들에 비하여 상대적으로 낮은 편일 뿐 절대적으로 보았을 때는 결코 낮은 수준이라고 할 수 없다. 이 척도에서, 아동의 성행동이 변화한다는 것은 매우 장기적이거나 강한 피해 정도에 노출되었기 때문일 것이다. 그러므로 성행동의 변화는 피해 사실에 관한 진술의 신빙성을 매우 뒷받침해준다고 해석할 수 있다.

타당성평가의 분석에서 ‘부모의 생각과 느낌을 강요’ 문항의 경우 이 문항에 대한 대답이 피해 사실에 관한 진술의 신빙성과 전혀 관련이 없는 것으로 나타났다. 이는 부모가 아동에게 부모의 생각과 느낌을 강요하여 진술토록 하는 상황이 실제 아동이 피해를 당했는지와 관련없이 발생하는 것으로 해석할 수 있다. 이는 맥락타당성에서 낮은 혹은 부적의 변별도를 보이는 항목의 결과와 전혀 다른 형태의 결과이다. 낮은 변별도는 해당 항목에 대한 값이 어떠한지 진술의 신빙성에 도움이 거의 되지 않는다는 것으로 해석할 수 있

으며 부적인 변별도는 그 값이 낮은 경우 오히려 진술 신빙성에 도움이 된다고 해석할 수 있다. 그러나 이 ‘부모의 생각과 느낌을 강요’ 항목은 그 값과 진술 신빙성이 상관을 가지고 있지 않아 분석이 불가능하며 더불어 해석 또한 할 수 없다. 그 외 타당성평가의 나머지 항목에 대한 변별도 분석 결과에서는 해당 항목들이 대체로 높은 변별도 지수와 유의미한 p 값을 가지고 있다. 이 척도의 항목들은 다른 척도들과 달리 해당하지 않을수록 진술의 신빙성이 높으나 내적 타당도 분석과 IRT 분석을 시행하기 위하여 역코딩 하였기 때문에 결과 해석에 있어서는 다른 척도들과 동일하다. 항목 중 ‘소통불능’이 변별도 지수 0.417로 낮은 변별력을 보여준다. 이 항목은 아동이 타인과 의사소통하는 데 어려움을 겪는지에 관한 것으로, 진술을 하는 데 있어 설명능력이 부족한 것이 성폭력 발생 사실을 부정하는 것은 아니며 단지 그 사실을 전달하는 것이 곤란한 것이기 때문에 ‘소통불능’의 해당 사항이 진술의 신빙성을 낮추는 것은 아니다.

분석 결과 맥락타당성 척도에서 낮은 수준의 변별도를 가진 세 항목과 동일값을 가져 분석이 불가능한 항목들을 제외하고 분석 결과를 얻을 수 있었던 항목들은 대부분 유의미하다고 해석가능한 수준의 변별도를 가지고 있는 것을 확인하였다. 그러나 본 연구에서는 적은 사례수의 한계점을 가지고 있다. 그로 인해 난이도와 변별도에서 유의미하다고 해석할 수 있는 p 값의 기준을 통상 사용하는 0.05에서 0.10 내외로 변경하여 완화하게 된 점도 연구의 한계로 볼 수 있다. 또한 내적 합치도 분석으로 한 번 구별한 후 남은 항목으로 IRT 분석을 실시했으나 모든 사례에서 항목이 동일한 값을 가져 제외한 항목들에 대한 분석은

시행할 수 없었다. IRT의 특징은 표본 수에 관계없이 문항을 분석할 수 있다는 특징을 가지고 있으나 사례 수가 많아질수록 수치가 더욱 안정적이고 동일한 값으로 인해 분석에서 제외되는 항목이 감소할 것이다. 추후 연구에서는 더 많은 사례를 추가해 다시 분석을 실시하여 제외한 항목들에 대한 변별도와 난이도를 구하여야 할 것이며, 더불어 전반적으로 상승한 내적 타당도를 기대할 수 있을 것이다.

참고문헌

- 김지선, 박지선, 최수형 (2009). 청소년대상 성범죄자 유형 분류와 프로파일링. 보건복지부 의뢰 미발표연구, 한국형사정책연구원.
- 김현정 (2010). 성폭력피해아동의 진술분석 도구의 상대적 유용성 비교연구. 경기대학교 박사학위 청구논문.
- 민유숙 (2009). 아동피해자진술의 평가에 관한 법원의 관점. 범죄심리사자격관리위원회 공동심포지움 발표논문집, 49-60.
- 아동피해자조사매뉴얼 (2009). 2009년도 진술 분석전문가 교육 자료집. 경기대학교 범죄심리학과.
- 이수정 (2009). 아동 대상 성범죄 조사절차의 미비점 및 개선방안 연구: 진술타당성분석(Statement Validity Analysis)의 형사사법적 활용가능성, 한국경찰연구, 8(3)호, 141-172.
- Coy v. Iowa, 56 U.S.L.W. 191 (1988).
- Faller, K. C. (2007). Interviewing Children About Sexual Abuse, Oxford: Oxford University Press.
- Gabora, N. J., Spanos, N. P., & Joab, A. (1993). The effects of complainant age and expert psychological testimony in a simulated child

- sexual abuse trial, *Law and Human Behavior*, 17, 103-119.
- Gibbons, R. D., Immekus, J. C., & Bock, R. D. (2007). The Added Value of Multidimensional IRT Models, *Didactic Workbook*.
- Hagman, B. T., Kuerbis, A. N., Morgenstern, J., Bux, D. A., Parsons, J. T., & Heidinger, B. E. (2009). An Item Response Theory (IRT) analysis of the short Inventory of Problems-Alcohol and Drugs (SIP-AD) among non-treatment seeking men-who-have-sex-with-men: Evidence for a shortened 10-item SIP-AD, *Addictive Behaviors*, 34, 948-954.
- Heger, A. H., Emans, S. J., & Muram, D. (1999). Evaluation of the Sexually Abused Child, 2nd, *Oxford: Oxford University Press*.
- Kovera, M. B., Gresham, A. W., Borgida, E., & Gray, E. (1997). Does Expert Psychological Testimony Inform of Influence Juror Decision Making? A Social Cognitive Analysis. *Journal of Applied Psychology*, 87, 178-191.
- Malloy, L. C., Lyon, T. D., & Quas, J. A. (2007). Filial Dependency and Recantation of Child Sexual Abuse Allegations. *American Academy of Child and Adolescent Psychiatry*, 46, 162-170.
- McGouth, L. (1995). For the record: Videotaping investigative interviews. *Psychology, Public Policy, and Law*, 1, 370-386.
- Muthén, B. O., & Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: multiple-group and growth modeling in Mplus, *Mplus web notes: No. 4, Version 5*.
- Muthén, B. O., & Muthén, L. K. (2007). Mplus User's Guide, 5th, *Los Angeles, CA: Muthén & Muthén*.
- Myers, J. E. B. (1998). Forensic Impact of the Child Sexual Abuse Medical Examination. *Child abuse & neglect*, 22, 653-656
- Osburn, H. G. (2000). Coefficient Alpha and Related Internal Consistency Reliability Coefficients. *Psychological Methods*, 5, 343-355.
- Pipe, M. E., Lamb, M. E., Orbach, Y., & Esplin, P. W. (2004). Recent research on children's testimony about experienced and witnessed events. *Developmental Review*, 24, 440-468.
- Quas, J., DeCicco, V., Bulkley, J., & Goodman, G. (1996, Spring/Summer). District attorney's views of legal innovations for child witnesses. *APLS News*, pp.7-12.
- 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.
- Reise, S. P., Ainsworth, A. T., & Haviland, M. G. (2005). Item Response Theory: Fundamentals, Applications, and Promise in Psychological Research. *Current Directions in Psychological Science*, 14, 95-101.
- Ross, D. F., Lindsay, R. C. L., & Marsil, D. F. (1999). The Impact of Hearsay Testimony on Conviction Rates in Trials of Child Sexual Abuse: Toward Balancing the Rights of Defendants and Child Witnesses. *Psychology, Public Policy, and Law*, 5, 439-455.

1 차원고접수 : 2009. 11. 14.

심사통과접수 : 2009. 12. 8.

최종원고접수 : 2009. 12. 28.

Characteristics of Item Responses of Statement Validity Analysis of Sexually Molested Children

Choi Hye Rim

Lee Soo Jung

Kyonggi University

This study analyzed fifty child molestation cases reported in five one-stop centers from March to August in 2009. Among statement validity tests for child victims, scales including Criterion Based Content Analysis, Reality Monitoring, Contextual Evaluation, Behavioral Change, Validity Evaluation were analyzed. For item analysis, Item Response Theory was applied to produce difficulty and differentiation indices after consideration internal consistency among items for each scale. Results presented all items except three items to have fair differentiation function. However, power was not enough because of a small sample size. This problem should be solved on further researches.

Key words : *Statement Validity Test, Internal Consistency, Item Response Theory, Item Characteristic Curve*