

그 말은 진실일까?: CBCA의 진실 탐지 효용성에 대한 메타분석적 고찰

김 혜 진 이 상 민 허 태 균 최 승 혁[†]

고려대학교 심리학과

진술의 진실성을 평가하는 절차인 진술타당도분석(Statement Validity Analysis: SVA)은 국내외 범죄 수사 및 법정 상황에서 활용되고 있다. SVA 절차 중 내용분석 단계에서는 준거기반내용분석(Criteria-Based Content Analysis: CBCA) 준거를 사용하여 진술에서 실제 경험에 기반하였을 때 나타나는 특징들이 현출되는지를 평가한다. CBCA 준거의 변별력과 효과크기에 대한 국내 연구는 다양한 패러다임으로 이루어졌지만, 그 연구 결과들의 일관성은 여전히 검증되지 않았다. 이에 본 연구에서는 국내에서 관련 연구가 시작된 2004년부터 2020년까지 수행된 CBCA 준거 관련 연구들(14개의 연구자료)에 대한 메타분석을 실시하였다. 연구 결과, CBCA 총점이 진실과 작화를 성공적으로 변별해내고 있었다. 준거별로는, 네 개의 준거(3번, 4번, 10번, 12번; 모두 인지적 준거)에서 정적(+)인 효과크기가 유의미했다. 그러나 18번 준거(동기적 준거)는 CBCA의 기본 가정과는 반대로, 효과크기가 부적(-)으로 유의미한 것으로 나타났다. 또한, 조절효과분석이 가능한 열한 개의 준거(2번~9번, 12번, 13번, 15번) 각각에 대해 잠재적 조절변인의 영향을 살펴본 결과, 일부 준거들에 대하여 진술인의 성별 및 신분, 연구 유형 및 설계, 사건 유형, 평가자 수, 출판 여부의 조절효과가 유의미한 것으로 나타났다. 따라서 본 연구결과는 진실여부를 타당하게 판단하기 위해서는 CBCA 총점보다는 세부 준거, 특히 인지적 준거에 초점을 맞추는 것이 중요하며, 잠재적 조절변인의 영향 가능성 또한 중요하게 고려해야 함을 시사한다. 관련하여 본 연구의 의의, 제한점 및 추후 연구 방향에 대해 논하였다.

주요어 : CBCA, SVA, 진술분석, 조절변인, 메타분석

[†] 교신저자 : 최승혁, 4단계 BK21 심리학교육연구단 연구교수, 서울 성북구 안암로 145 고려대학교 법학
구관 402호, dawnseye@naver.com

진실성 판단에 대한 문제는 예로부터 많은 관심의 대상이 되어왔다. 정확한 사실 확인부터, 화자의 신용성 판단, 그리고 상업적으로 신뢰할만한 데이터 확보에 이르기까지 진실성 판단은 사회 전반에서 중요한 함의를 가진다. 특히, 범죄 사건에서, 사건 관계인들(피의자, 피해자, 목격자 등)의 진술에 대한 신빙성을 판단하는 것은 객관적인 물적 증거의 확보가 어렵거나, 확보되었더라도 그 증거가 범죄를 입증하는 데 결정적인 역할을 하지 못할 때 더욱 중요해진다.

예를 들어, 성폭행 고소 사건에서, 피해자의 신체에서 피의자의 정액이 검출되었을 때 이는 두 사람 간에 성관계가 발생하였음에 대한 근거는 될 수 있지만, 피의자가 피해자를 성폭행했다는 근거는 될 수 없다. 즉, 피의자가 합의에 의한 성관계였음을 주장한다면, 성관계 여부 자체보다는 그 과정에서 피의자가 강압과 폭력 등을 이용하여 피해자의 자유의사를 억압하였는지가 쟁점이 되기 때문이다(김미영, 2016). 이러한 경우, 피의자와 피해자의 진술 중에서 어떤 것이 더 믿을 만한지가 재판에서 중요한 쟁점이 된다. 현재, 국내의 수사 및 재판 단계에서 진술의 신빙성을 판단하는데 사용되는 대표적인 진술분석 기법은 진술타당도분석(Statement Validity Analysis: SVA)이다.

SVA의 절차와 법적 활용

SVA는 진술내용의 신빙성에 대하여 가설을 세우고 검증하는 일련의 절차이다(Köhnken, 2004). 사회과학의 연구방법론에 따라, SVA는 ① 사건기록을 분석하고, ② 가설을 설정하고, ③ 면담계획(가설검증을 위한 방법과 전략)을

수립하고, ④ 진술인을 면담하여 진술을 확보하고, ⑤ 진실의 증거를 이용하여 진술내용을 분석하고, ⑥ 진술의 일관성과 타당성을 검토하여, ⑦ 최종 결론을 도출하는 과정을 거친다. 이를 통해 도출한 최종 결론이 '진술인의 진술이 진실하다'라면, 이는 당해 진술이 진실의 증거와 부합하며, 여타 외적인 요인(예, 암시, 코칭 등)의 영향 가능성이 작아 타당한 것임을 의미한다(김미영, 2016).

SVA는 국외에서 1950년대부터 법정에서 활용되기 시작하였다.¹⁾ 국내에서는 2006년 검찰과 2007년 경찰에서 SVA를 수사에 활용하기 시작하였고, 2008년 의붓아버지 성폭력 사건²⁾을 기점으로 SVA 결과가 법원에서 증거로 인정되기 시작하였다.

예를 들어, 피고인이 아동인 피해자를 유인하여 간음한 혐의를 받은 사건에서 법원은 피해자의 진술을 분석한 진술분석전문가의 '아동 성폭행 사건 전문가 의견서'를 인용하며 유죄의 근거를 설시하였다(인천지방법원 부천지원 2019.01.11. 선고 2018고합143, 2018전고

1) 학자들은 이때 SVA가 태동하기 시작하였다고 본다. 즉 이 시기에 독일에서 SVA의 모태가 되는 진술현실성분석(Statement Reality Analysis: SRA)과 현실성 증거(reality criteria)를 개발한 심리학자 Udo Undeutsch가 연방대법원의 의뢰를 받아 진술 신빙성 평가를 실시하였는데, 진술의 신빙성이 핵심이 되는 사건의 경우 심리학자와 같은 전문 감정인의 의견을 구하도록 대법원이 판결하면서 관련 전문가 증인이 급속도로 증가하였기 때문이다(Undeutsch, 1967; Undeutsch, 1989).

2) 피해 아동을 대상으로 실시한 인지면담 및 진술 분석 결과통보서를 근거로 피의자를 구속하였고, 대검찰청 소속 진술분석관의 법정 증인을 통해 위 결과가 유죄판결의 증거요지에 실시되었다(김미영, 2016).

11(병합) 판결).³⁾ 마찬가지로, 진술인의 주장이 반복된 경우에도 최초 진술에 대한 진술분석 의견서가 유죄판결의 근거로 인정되기도 하였다(서울고등법원 2020.01.30. 선고 2019노630, 2019전노4(병합), 2019보노19(병합) 판결; 대법원 2020.05.14. 선고 2020도2433 판결).⁴⁾ 또한, 전부는 아니지만 일부 진술분석결과가 법원에서 받아들여진 사례도 있다(부산지방법원 동부지원 2019.12.10. 선고 2019고합141, 2019전고8(병합) 판결).⁵⁾

- 3) 구체적으로, 진술인의 진술에서는 사건 당시 피고인의 행동과 말, 그리고 이에 대한 위 피해자의 반응에 대해서 일관성을 유지하며 논리적으로 묘사하였으며, 사건 당시 느꼈던 감정과 거부 반응에 대해 상당히 구체적으로 묘사하고, 사건 당시 거부와 저항 행동에 대한 피고인의 반응에 대한 구체적인 내용이 나타나 세부정보가 풍부하고 상호작용의 묘사가 나타났다는 점을 피해 관련 진술에 대한 신빙성 판단의 중요 근거로서 들고 있다. 이와 더불어, 법원은 진술 조사 과정에서 피해자의 진술에 영향을 줄 만한 유도, 암시를 비롯하여 피해자로 하여금 과장해서 진술하게 할 만한 주변 요인은 발견되지 않았다는 점에도 주목하였다.
- 4) 성폭력 사건에서 진술분석전문가가 아동의 최초 피해 진술에서 실제로 경험한 내용을 반영하는 독특한 내용이 나타나 신빙성이 높다고 판단 내렸으며, 피해 주장 철회와 관련하여 주변인의 압력 등 타당한 외부적 요인의 가능성이 시사된다는 점을 유죄의 근거로 실시한 바 있다.
- 5) 피고인이 두 명을 대상으로 성범죄를 저지른 혐의를 받은 사건에서 한 피해자의 피해 진술이 신빙성이 있다는 진술분석전문가의 의견은 전반적인 진술의 일관성과 구체성, 타당성 등의 이유로 유죄의 증거 요지로 실시되었다. 한편, 다른 한 피해자의 피해 주장이 신빙성 있다는 진술분석 결과는 인정되지 않았는데, 이는 진술 과정에서 질문 방식의 제한이 있었던 점, 기억의 부재를 일부 호소하는 점, 피고인이 아닌 다

CBCA 준거

그렇다면, 진술분석전문가는 어떠한 기준에 의거해 진술의 신빙성을 판단하는가? SVA의 대표적인 진실 판단 준거 중 하나는 준거기반 내용분석(Criteria-Based Content Analysis: CBCA) 준거이다. CBCA 준거는 '실제 경험에 대한 진술은 허구에 기반한 진술과는 내용과 질적으로 차이가 있다'는 Undeutsch (1967)의 가설에 근거하여, 성폭력 피해 아동들의 진술에서 귀납적으로 도출되어 정립된 19개의 언어적 특징들을 반영하고 있다(Steller, 1989; Steller & Köhnken, 1989). CBCA 준거들은 크게 일반적 특징, 구체적인 내용, 내용의 특수성, 동기 관련 내용, 범죄 특징적 요소로 구분된다(표 1 참고).

Köhnken(1996; 2004)에 의하면, 일반적 특징(1~3번), 구체적인 내용(4~7번), 내용의 특수성(8~13번) 준거들은 인지적 준거로, 동기 관련 내용(14~18번) 준거들은 동기적 준거로 재분류된다. 인지적 준거는 진실된 화자들의 진술에서 나타나는 인지적인 특징들을 반영하고 있고, 동기적 준거는 진실을 이야기하는 것으로 보이고자 원하는 화자의 진술에서 나타나지 않을 인상 관리와 관련된 특징들을 반영하고 있다. 이러한 인지적·동기적 CBCA 준거들은 실제 경험한 것을 진술할 때 나타나는 특징, 즉 진실의 징후이다. 다시 말해, 이러한 준거와 관련된 진술 내용이 더 많거나 확실하게 나타날수록 당해 진술은 실제 경험에 대한 것일 가능성이 크다고 본다(Vrij, 2008).

른 사람에게 피해를 입었을 가능성을 배제할 수 없는 점 등 때문이었다.

표 1. CBCA 준거

일반적 특징(General characteristics)
1. 논리적 일관성(Logical consistency)
2. 정연하지 않은 진술(Unstructured production)
3. 세부정보의 풍부함(Quantity of details)
구체적인 내용(Specific contents)
4. 사건이 발생하게 된 맥락정보(Contextual embedding)
5. 상호작용 묘사(Descriptions of interactions)
6. 대화의 재연(Reproduction of conversation)
7. 사건 도중에 예기치 않은 상황 발생(Unexpected complications during the incident)
내용의 특수성(Peculiarities of content)
8. 일상적이지 않은 독특한 세부정보(Unusual details)
9. 여분의 세부정보(Superfluous details)
10. 이해하지는 못했으나 정확하게 보고한 세부 정보(Accurately reported details not comprehended)
11. 관련된 외적 연합(Related external associations)
12. 주관적인 심리 상태(Accounts of subjective mental state)
13. 가해자의 정신상태 귀인(Attribution of perpetrator's mental state)
동기 관련 내용(Motivation-related contents)
14. 자발적인 수정(Spontaneous corrections)
15. 기억부족의 인정(Admitting lack of memory)
16. 자신의 진술에 대한 의문 제기(Raising doubts about one's own testimony)
17. 자기 비하(Self-deprecation)
18. 가해자 용서(Pardoning the perpetrator)
범죄 특정적 요소(Offence-specific elements)
19. 범행의 전형적인 특징 묘사(Details characteristic of the offense)

주. CBCA 준거의 한국어 명칭은 전문가 또는 연구자마다 조금씩 차이가 있으나, 본 연구에서는 경찰청 (2017)이 발간한 진술분석 매뉴얼을 따른다(pp.86-88).

19번 준거는 그 정의가 모호하고 범행의 전형적인 특징과 진술이 일치하는지 확인할 객관적 방법이 없다는 등의 문제점 때문에 CBCA 준거로 거의 사용되지 않는다(김미영, 2016; Raskin, Esplin, & Horowitz, 1991).

진술의 신빙성을 판단하는 기준에 대한 연구들은 1960~1970년대부터 시작되어(Undeutsch, 1967; Trankell, 1972), 1980년대 후반부터는 본격적으로 CBCA 준거에 대한 연구들이 활발히 수행되었다. 국외에서는 현재까지 약 30년 동안 100건 이상의 연구가 CBCA 준거의 변별

력, 즉 실제 경험(진실)과 작화를 구분할 수 있는지를 탐구한 것으로 알려져 있다(Sporer, Manzanero, & Masip, 2021).

초기 연구들에 대한 Vrij(2005)의 질적 고찰에 의하면, 37개의 연구 중, 35개의 연구에서 진실한 진술에서 작화 진술보다 유의미하게 높은 CBCA 총점수가 나타났다. 다만, 연구마다 초점을 맞춘 준거와 그에 대한 결과가 다소 상이하게 나타났다. 예를 들어, Esplin, Boychuk과 Raskin(1988)은 19개의 CBCA 준거 모두를 적용하여, 10번, 16번, 17번 준거를 제외한 나머지 준거들에서 진실과 작화집단 간 차이가 나타남을 보고하였다. 반면, Sporer(1997)는 CBCA 준거 1~13번에만 초점을 맞추었고 그중 1번과 4번 준거에서만 진실과 작화집단 간 차이가 나타남을 보고하고 있다.

나아가, 다양한 방법으로 수행된 CBCA 준거 관련 연구들을 양적으로 고찰하는 메타분석 연구 또한 다수 수행되었다(예, Amado, Arce, & Fariña, 2015; Amado, Arce, Fariña, & Vilariño, 2016; Hauch, Sporer, Masip, & Blandón-Gitlin, 2017; Oberlader, 2019; Oberlader et al., 2016). 그리고 이러한 메타분석 연구들을 종합적으로 볼 때 CBCA 준거가 각기 다른 연구 설계 및 집단에서 대체로 CBCA의 기본 가정(Undeutsch 가설)과 부합한다고 볼 수 있다(Oberlader et al., 2021).

그러나 세부적으로 들여다보면 조금씩 다른 결과가 발견된다. 예를 들어, CBCA 준거가 실제 경험에 대한 진술과 작화 진술을 제대로 변별하는지 메타분석을 실시한 결과, 아동과 청소년을 대상으로 한 연구들을 분석하였을 때는 CBCA 총점을 비롯하여 모든 준거가 CBCA의 가정과 부합하는 방향으로 유의미한 효과크기를 보였지만(Amado et al., 2015), 성인

을 대상으로 한 연구들을 분석하였을 때는 17번과 18번 준거에서 유의미한 결과가 산출되지 않았다(Amado et al., 2016). 또한, CBCA 준거에 대한 평가자 간 신뢰도에 초점을 맞춘 메타분석(Hauch et al., 2017)에서는 2번과 9번 준거를 제외한 나머지 CBCA 준거들에서만 높은 신뢰도를 보였다.

국내에서 2000년대 중반부터 CBCA의 변별력을 확인하기 위해 수행된 일련의 연구들을 살펴봤을 때에도 모든 준거들이 의미 있게 진실집단과 작화집단을 구분하는 것은 아니었다. 예를 들어, 조은경(2004)은 13세 미만 아동 성폭력 사건에 대하여 피해자 진술 외 증거를 토대로 피해 '확실', '애매', '의문' 집단으로 분류한 후 피해 진술을 분석한 결과, 확실 집단의 CBCA 총점이 애매 집단과는 차이가 없었으나 의문 집단보다는 높은 것으로 나타났다. 준거별로는 3번, 4번, 5번, 8번, 12번 등의 준거들만이 피해 확실 집단과 의문 집단을 구분하고 있었다. 또한, 이미선(2018)은 3~6세 아동을 대상으로 비밀 은폐 여부에 따른 아동 진술의 특성을 분석하였다. 그 결과, 비밀을 숨기지 않고 발설한 집단에서 은폐 집단보다 CBCA 총점이 더 높았다. 준거별로는 2번, 3번, 6번, 9번 등의 준거만이 두 집단을 의미 있게 구분하는 것으로 나타났다.

한편, 이수정(2010)은 CBCA 준거들 중 11개 준거(1번, 3~10번, 12번, 13번)만을 사용하여 실제 아동청소년 대상 성폭력 사건의 피해 진술과 실험적으로 피해를 가정하게 한 후의 진술을 비교하였다. 그 결과, CBCA 총점이 그리고 7번, 9번, 10번을 제외한 나머지 준거들만이 실제 피해 진술과 작화 진술을 의미 있게 구분하는 것으로 나타났다.

이와 같이, 실제로는 모든 준거가 변별력을

보이지 않는 결과는 오히려 자연스러운 것이라고 할 수 있다. 실제 경험에 대한 진술을 하고 있다 할지라도 반드시 모든 증거와 관련된 진술이 포함되리라는 보장은 없기 때문이다. Köhnken(1999) 또한 사건의 유형에 따라 적용이 가능한 증거가 상이할 수 있으므로 확보된 진술에 CBCA의 모든 증거를 기계적으로 대입하는 것은 삼가야 한다고 제안하고 있다.

그러나 가장 큰 문제는 CBCA의 가정과 반대되는 결과가 산출될 때에 있다. 이는 실제 경험한 것을 진술할 때에, 그렇지 않을 때보다, CBCA의 증거와 관련된 내용이 더 많이 현출될 것이라는 기본 가정이 틀릴 수 있다는 것을 의미하기 때문이다. 연구들을 살펴보면, CBCA 증거들 중 일부에서 그 가정과 반대되는 방향으로 진실 진술과 작화 진술 간 차이가 나타났다.

예를 들어, Köhnken, Schimossek, Aschermann과 Höfer(1995)는 영상물 시청(목격) 경험에 대하여 본 그대로 진술 또는 작화하게 한 결과, 전자에서 더 높은 CBCA 총점을 확인하였다. 그리고 준거별로는 2번, 3번, 6번, 8번, 9번, 12번, 14번, 15번, 16번 등의 준거들이 진실의 진술에서 더 많이 나타났다. 그러나 준거 13번(가해자의 정신상태 귀인)은 작화진술에서 더 많이 나타났다.

이렇듯 CBCA의 기본 가정과 상충되는 결과는 국내 연구에서도 나타났다. 김현정(2010)은 실제 아동 성폭력 피해 진술과 실험적으로 작화하게 한 피해 진술을 분석한 결과, 전자에서 더 높은 CBCA 총점이 나타났다. 그리고 준거별로는 1번, 3번, 6번, 8번, 11번, 12번, 14번, 15번 등의 준거들에서 가정과 일치하는 결과가 산출되었다. 그러나 16번 준거(자신의 진술에 대한 의문 제기)는 CBCA 가정과 반대로

작화 집단에서 더 많이 현출되는 것으로 나타났다.

윤대중(2019) 또한 CBCA 가정과 반대되는 결과를 일부 보고하고 있다. 즉, 이 연구에서는 범칙면으로 확보한 실제 범죄사건의 성인 피해자, 용의자, 피의자의 진술을 분석하였는데, 그 결과, 세 집단 간 CBCA 총점의 유의미한 차이가 확인되었으며 특히 피해자 진술에서 CBCA 총점이 높게 나타났다. 준거별로는 3번, 4번, 5번, 8번, 9번, 12번, 13번 등의 준거들에서 가정과 일치하는 방향으로 집단 간 차이가 나타났다. 그러나 14번 준거(자발적인 수정)는 CBCA의 가정과는 반대로 피해자 집단보다 (작화 가능성이 높다고 가정된) 용의자 집단에서 더 많이 현출되는 것으로 나타났다.

잠재적 조절변인

일부 연구 결과에서 변별력을 갖는 CBCA 준거가 다소 상이하게 나타난다는 점, 특히, CBCA의 가정과 반대되는 방향으로 일부 준거들이 현출된다는 점은 연구의 대상, 방법 및 특징들이 서로 다르다는 것에 일정 부분 기인할 수 있다. 이에 메타분석 연구들과 관련 논의들을 바탕으로 조절변인들을 유추해 보면 다음과 같다.

먼저, 연구 대상자의 연령이나 성별 등의 진술인 특성이 조절변인이 될 수 있다. 많은 연구들(김미영, 2016; Akehurst, Köhnken & Höfer, 2001; Vrij, Akehurst, Soukara, & Bull, 2002)은 진술인의 연령이 증가할수록 진술능력이 발달하고 진위여부와 상관없이 CBCA 점수가 높아져 변별력이 떨어질 수 있음에 유의해야 한다고 제안한다. 그러나 최근 메타분석

에서는 연령에 따른 조절효과가 나타나지 않았다(Oberlader et al., 2021).

또한, 일반적으로 언어적 묘사 능력은 여성이 더 뛰어난 것으로 보고되고 있다(Hyde & Linn, 1988). 그러나 진술자의 성별에 따른 조절효과가 나타나지 않는 경우도 있고(Oberlader et al., 2016), 진술인 집단에서 여성의 비율이 높을수록 효과크기가 상승한다는 결과도 보고된다(Oberlader et al., 2021).

진술자의 신분 또한, 조절변인으로 고려될 수 있다. 일반적으로, 사람들은 사건의 피의자가 진실되지 않을 것이라는 고정관념을 가지고 있다(정의철, 이창호, 2017). 그러나 법원의 최종판단이 내려지기까지 사건의 피의자가 거짓을 이야기하는 범인이라고 단정 지을 수는 없다. 따라서 진술자의 신분 변인은 검증이 필요한 중요한 변인이다.

관련하여, Porter와 Yuille(1996)는 범죄를 저지른 피의자의 진술(작화 조건)에서보다 무고한 피의자의 진술(진실 조건)에서 CBCA 준거가 더 많이 현출되었음을 보고하고 있다. 또한, 메타분석을 포함한 많은 연구들(Amado et al., 2016; Vrij et al., 2002; 2004)은 진술인의 신분(피의자, 피해자)과 상관없이 CBCA 총점이 진실한 진술에서 더 높게 나타남을 보고하고 있다.

연구 유형(현장, 실험) 및 설계(집단간, 집단내)도 조절변인이 될 수 있다. 메타분석 연구 결과를 살펴보면, 실험 연구보다 실제 범죄 사건의 진술을 분석한 현장 연구인 경우에 더 큰 효과크기가 산출된다(Amado et al., 2015; Amado et al., 2016; Hauch et al., 2017; Oberlader et al., 2021).

또한, 집단간 설계가 집단내 설계보다 약간은 더 큰 효과크기를 보이는 것을 알 수

있다(Oberlader, 2019). 그러나 이와는 별개로, Vrij(2019)는 진술 내용 외에도 진술인의 성격(예, 말이 많음), 진술 준비 정도, 사건 특성 등이 동일할 때 정확한 비교를 할 수 있다면 서 집단내 설계 연구를 독려하고 있다.

사건 유형도 조절변인으로 고려될 수 있다. CBCA는 본래 아동 성학대 피해자의 진술의 진위여부를 평가하기 위해 개발되었지만, 경제적, 정서적, 신체적인 피해를 입은 경험에 대하여 진술하도록 한 경우에도 CBCA 결과가 진실 진술과 작화 진술에서 차이를 보였다는 보고가 있다(Merckelbach, 2004). 이것은 실제 사건에 대한 경험은 상상하는 것보다 더 강렬한 기억이 수반되며(Schacter & Addis, 2007) 이러한 기억이 진술내용에 자연스레 반영될 수 있기 때문이다.

이러한 이유로 CBCA는 아동 성폭력 사건 뿐만 아니라 성인 성폭력과 기타 범죄 사건에도 활용된다(Ruby & Brigham, 1997). 그러나 Amado 등(2016)의 메타분석에 의하면, 실험 연구보다는 현장 연구에서, 그중에서도 일반사건에서보다 성범죄와 가정폭력 사건에서 큰 효과크기가 산출되는 것으로 나타났다.

CBCA 연구들 간 나타나는 차이점으로 평가자 유형이 있다. 일반적으로 CBCA 연구에서는 평가의 객관성을 담보하기 위해 저자보다는 제3자가, 1인보다는 다수의 평가자가 준거들을 평정하고 평가자간 신뢰도를 보고한다(Boychuk, 1991; Wells & Loftus, 1991). CBCA 준거의 평가자간 신뢰도 결과를 메타분석한 Hauch 등(2017)에 따르면, 준거 2번과 9번을 제외한 대부분의 준거들에서 높은 평가자간 신뢰도가 나타났다.

또한, 어떠한 CBCA 준거를 사용했는지도 중요할 수 있다. 각 준거별로 산출된 효과크

기를 단순 합산하여 비교하였을 때에는 일반적으로 인지적 준거(1번~13번)가 동기적 준거(14번~18번)보다 큰 효과크기를 산출한다(Amado et al., 2015). 그러나 Undeutsch 가설에 부합하는 준거(1번~14번)만을 사용한 연구들과 19개 준거를 모두 사용한 연구들 간에는 효과크기의 차이가 거의 없었다(Amado et al., 2015; Oberlader, 2019; Oberlader et al., 2016). 한편, Undeutsch 가설 버전(1번~14번)과는 별개로, 준거 일부만 따로 떼어 사용하는 것보다는 전부를 사용하는 것이 효과크기가 더 크게 나타남을 보고하는 메타분석 연구도 있다(Oberlader et al., 2016; Oberlader et al., 2021).

연구의 출판 유형 또한 잠재적 조절변인이 될 수 있다. 일반적으로, 통계적으로 유의미하지 않은 연구들은 출판되지 않는 경향이 있기 때문이다(홍세희, 2013; 황성동, 2014). 메타분석 연구들(Oberlader, 2019; Oberlader et al., 2016)에서도, 통계적으로 유의미하지는 않지만, 출판된 연구가 출판되지 않은 연구보다 효과크기가 더 크게 나타났다.

이와 같이, 여러 메타분석 연구들과 관련 논의들을 바탕으로 CBCA 결과에 영향을 미칠 수 있는 잠재적 조절변인들을 유추해 보면, 크게 진술자의 연령 및 성별, 진술자의 신분, 연구 유형 및 설계, 사건 유형, 평가자 요인, 준거 요인, 출판 유형 등을 도출할 수 있다.

국내에 SVA가 도입되어 CBCA가 일선 수사 현장에서 다루어진 지 20년 가까이 되었고, 법정에서 진술의 신빙성을 다투는 중요한 근거로 활용되는 사례 또한 점차 증가하고 있다. 그러나 국내의 CBCA 준거 관련 연구들을 통합하여 CBCA 총점뿐 아니라 각 준거들의 타당성과 효과성을 살펴본 메타분석 연구는 찾아볼 수 없는 것이 현실이다. 이에 본 연구에

서는 그간 국내에서 수행되었던 CBCA 연구들을 종합하여 CBCA의 준거들이 그 기본 가정에 부합하게 진술의 진실 여부를 성공적으로 변별할 수 있는지, 이에 영향을 미칠 수 있는 조절변인은 무엇인지 살펴보고자 한다.

방 법

문헌검색

메타분석을 위한 자료 수집은 국내에서 진술분석과 관련된 연구논문이 처음 발표된 2004년부터 2020년까지 약 17년간 국내에서 발간된 학술지 논문과 학위논문을 대상으로 하였다. 이를 위해 국내 학술지들을 전문적으로 검색할 수 있는 공인된 문헌 검색 사이트인 RISS, KISS, DBpia를 이용하였다.

메타분석 연구에서는 가능한 많은 검색 자료를 찾아내는 것이 중요하기에 먼저 검색어로 '진술분석'을 선택하여 위의 3개 사이트에 대해 본 연구의 1저자와 2저자가 독립적으로 검색을 진행하였다. 그 결과 검색된 자료에는 실질적인 진술분석에 관한 SVA(Statement Validity Analysis: 진술타당도분석), CBCA(Criteria-Based Content Analysis: 준거기반 내용분석), RM (Reality Monitoring: 현실성 검토), SCAN(Scientific Contents Analysis: 과학적 내용분석)과 더불어 본 연구와 전혀 무관한 사항(예, 역사 문헌의 진술서 관련 내용), 유아교육(예, 유아 교육내용 진술방식) 또는 언어학 등에 관한 자료가 상당히 많이 포함되어 나타나는 문제가 발생하였다(예, DBpia에서 '진술분석'으로 검색하면 1천여 건 이상의 문헌이 식별).

이러한 문제점을 해결하기 위해 메타분석에서 관련 문헌에서 표현하는 용어 2가지를 조합하여 검색하는 전략을 구사한다는 신우중(2015)의 제안에 따라서 본 연구에서 포함시키고자 하는 CBCA에 관한 양적 분석 데이터를 찾기 위해 검색어로 'CBCA, 진술분석'과 '진술타당도분석도구'를 최종 선정하였다.

이와 더불어 관련된 공개자료를 추가적으로 찾기 위해 실제 진술분석을 수행하는 전문가에 문의한 결과 국내 검찰 및 경찰 등 공공기관에서 진술분석에 관한 연구용역 등을 진행한다는 사실을 확인하고 관련된 정부 용역 보고서 등을 추가하기 위해 PRISM 사이트를 이용하였다. 검색어는 앞서 검색한 것과 같이 '진술분석', 'CBCA, 진술분석'과 '진술타당도분석도구'를 사용하였다.

자료의 선정 및 제외 기준

자료의 선정은 체계적 고찰과 메타분석 연구 지침(Preferred Reporting Items for Systematic Reviews and Meta-Analyses: PRISMA)을 따랐다(황성동, 2014 참고). PRISMA에 따른 자료 수집의 흐름도는 그림 1과 같다.

구체적인 자료 선정 및 제외 기준은 다음과 같다. 첫째, 키워드 검색을 통해 발견한 자료 중, 중복되는 연구는 제외하였다(예, 학술지에 게재된 연구와 동일한 데이터를 사용한 학위 논문 등). 둘째, 제목, 키워드, 초록 분석 수준에서 해당되지 않는 논문들은 제외하였으며, 온라인상 원문을 확인할 수 없는 논문과 비공개 연구용역자료 등은 제외하였다. 셋째, CBCA 총점과 준거별 점수의 평균, 표준편차, p 값, F 값 등이 명시되어 효과크기를 계산할 수 있는 연구만을 선정하였고, 질적 연구 등은

제외하였다(예, 진술분석 절차에 관한 연구, 진술분석 문항의 내적합치도에 관한 연구 등). 넷째, 하나의 연구에서 나온 여러 유형의 데이터는 각각 분리하여 메타분석에 포함하였다. 최종적으로 본 메타분석에는 14개의 연구자료가 포함되었다.

자료 코딩

선정 기준에 맞는 연구자료의 코딩 방식은 다음과 같다⁶⁾: 진실/작화 집단, 진술자의 연령(성인/비성인/혼합), 진술자의 성별(여성/혼합), 진술자의 신분(피해자/가해자/혼합), 연구유형(현장/실험), 연구설계(집단간/집단내), 사건유형(성폭력/일반/혼합), 준거 평가자의 수(1인/2인 이상), 출판유형(학술지/비학술지)기.

진실집단은 실제 어떤 사실을 경험했거나 피해의 발생 가능성이 높다고 연구에서 설정한 집단이었고, 작화 집단은 어떤 사실을 상

-
- 6) 당초 조절변인으로 고려했던 준거 요인은 11개와 19개의 준거를 사용한 연구자료가 각각 1개와 2개이고 나머지는 모두 18개의 준거를 사용하고 있었다. 따라서 하위집단의 비율 상 편향된 결과가 발생할 것이기 때문에 준거 요인에 대해서는 분석하지 않았다.
 - 7) 일반적으로 메타분석 연구에서는 출판편향의 가능성을 확인하고 효과크기의 평균값을 교정하는 방법으로 Egger의 회귀분석과 Trim과 Fill 기법 등을 사용한다. 그러나 이 방법들은 출판된 연구와 비출판된 연구들의 효과크기를 직접적으로 비교 분석해 주지는 못한다. 본 연구는 CBCA 준거들의 효과크기에 영향을 미치는 잠재적 조절변인을 찾는 것이 주목적이다. 따라서, 메타 ANOVA의 전제조건을 충족한 준거들에 대해서는 출판유형의 조절효과를 분석하였고, 조건을 충족하지 못한 CBCA 총점에 대해서는 출판편향의 가능성을 분석하였다.

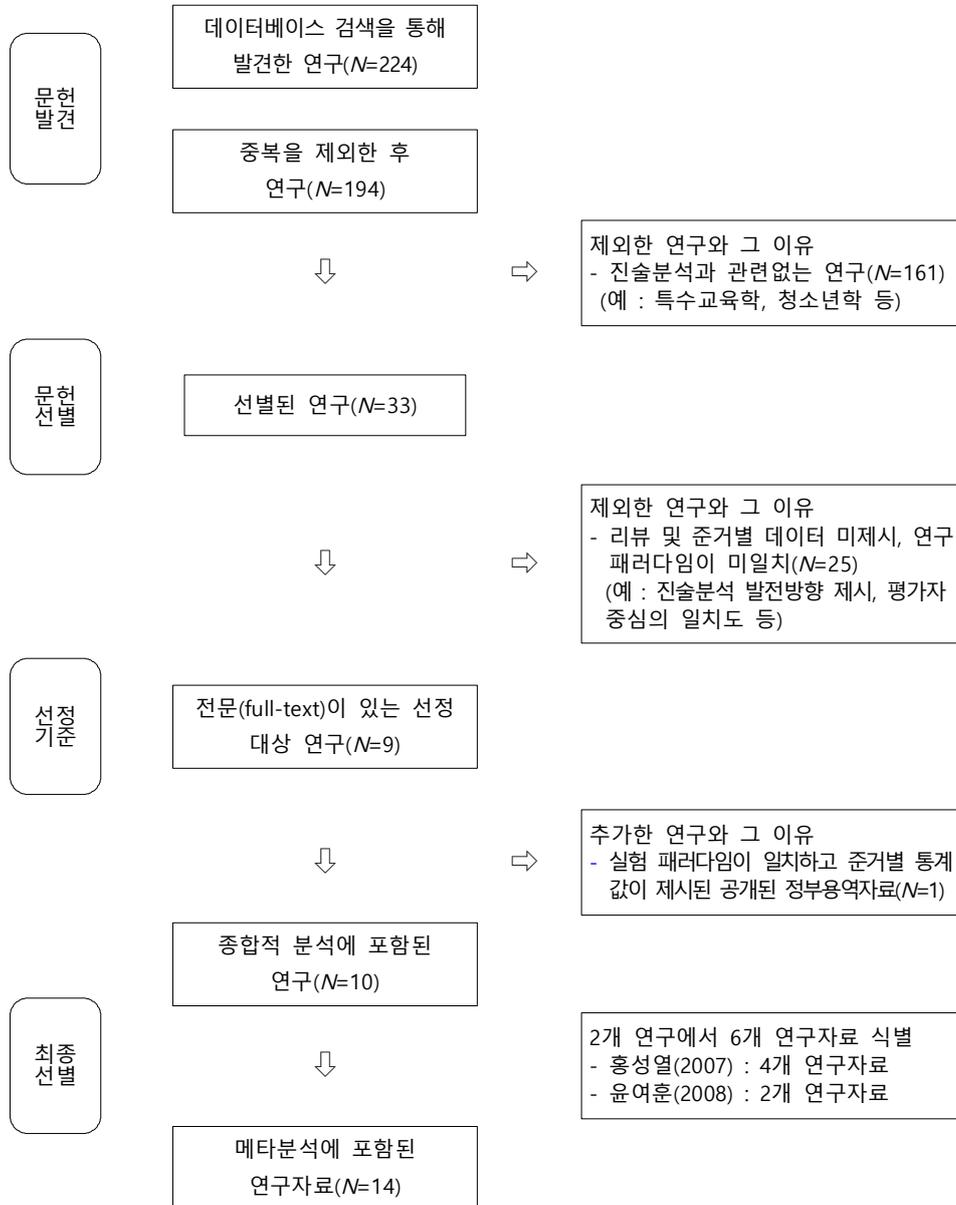


그림 1. PRISMA 흐름도

상하거나 피해의 발생 가능성이 낮다고 연구에서 설정한 집단이었다. 사건유형에서 일반 사건은 성폭력을 제외한 폭행, 절도, 및 기타 사건을 포함하였다. 출판형태에서 비학술지에

는 학위논문(석사/박사)과 보고서, 연구총서와 같이 학술지에 게재되지 않은 것을 포함하였다. CBCA 준거의 통계수치가 정상적으로 보고되지 않았고 저자와의 연락이 닿지 않은 경우

(예, 점수의 범위가 0~2사이인데, 표준편차가 50으로 보고), 관련 수치는 입력하지 않았다.

각 자료의 코딩은 연구과정에서 훈련을 받은 본 연구의 2저자가 담당하였다. 1저자와 교신저자는 코딩 결과물을 검토하고 수정하는 역할을 담당하였다. 본 메타분석에 포함된 연구자료들의 특성은 부록 1에 제시되었다.

자료 분석

CBCA 준거가 진실과 작화 집단을 구분할 수 있는지 확인하기 위해, CBCA 총점과 준거별 점수의 효과크기, 비중복 백분위 지수(U_3)⁸⁾, 그리고 이에 영향을 미칠 수 있는 잠재적 조절변인의 효과를 분석하였다.

모든 효과 크기는 표본 크기의 편파를 교정한 지수인 Hedges' g 를 사용하였다(Hedges & Olkin, 1985). 이 값은 .2는 小, .5는 中, .8은 大로 해석된다.

메타분석에서 평균 효과크기를 계산하는 방식에는 고정효과모형과 무선평과모형이 있다. 고정효과모형은 모든 연구의 모집단 효과 크기가 동일하다고 가정하고, 연구들 간 관찰된 효과크기의 차이는 표집오차에 의한 것이라고 간주한다. 반면, 무선평과모형은 연구의 방법 및 특성 등의 차이로 인해 모집단의 효과크기가 이질적이라고 가정하고, 연구들 간 관찰된 효과크기의 차이는 표집오차와 연구 간 분산에 의한 것이라고 간주한다(황성동, 2014).

본 메타분석에 포함된 연구들은 서로 다른 연구자에 의해 각기 다른 연구방법으로 수행되었다. 일부 같은 연구자에 의해 수행된 연구들도 있지만, 이 연구들도 연구 대상의 연령이나 성별의 비율 및 진술자의 신분 등이 서로 달랐다. 따라서 본 분석에서는 연구들 간 이질성을 가정하고 무선평과모형을 적용하였다.

아울러, 연구들 간 이질성에 영향을 미칠 수 있는 잠재적 원인을 탐색하기 위해 조절효과 분석을 실시하였다. 조절효과 분석(메타 ANOVA)은 CBCA 총점 및 준거별로 최소 10개의 연구자료가 있을 때, 동질성 검증치(Q)와 이질성 검증치(I^2)를 확인하여 수행되었다. 즉, 연구들 간 효과크기가 동질적이지 않다는 것을 의미하는 초과분산($Q-df > 0$)이 존재하면서 Q 값의 유의확률이 .10보다 작은 경우에, 그리고 이질성의 크기를 나타내는 I^2 값이 50% 이상인 경우에 한해 조절효과 분석이 실시되었다. I^2 값은 전체 관찰된 분산 중 실제 연구 간 분산이 차지하는 비율로, 25%이면 小, 50%이면 中, 75%이면 大로 간주된다(황성동, 2014). 모든 분석은 Comprehensive Meta-Analysis(CMA), Version 3.0(Borenstein, Hedges, Higgins, & Rothstein, 2009)을 사용하여 수행되었다.

결 과

준거별 효과크기

CBCA 총점 및 준거 별 효과크기를 각각 분석하였는데, 연구자료들 중, 총점 또는 준거의 통계값이 제시되지 않은 것들이 있어 각 분석에 포함된 연구는 상이하다(표 2 참고).

8) 비중복 백분위 지수는 통제 집단 평균(50%ile)을 기준으로 실험 집단 평균의 위치를 표시한 것으로, 효과크기를 z 점수와 동일하다고 가정하고 표준정규분포표에서 z 점수까지의 누적 면적에 해당되는 값을 찾은 뒤 그 값을 백분위(%ile)로 나타낸 것이다.

표 2. CBCA 준거의 효과크기

준거	k	N	효과크기 (SE)	95% CI		Z(p)	U ₃	효과크기 이질성			연구일련번호
				하한값	상한값			Q(df)	p	I ²	
1	7	681	.047(.121)	-.191	.286	.391(.696)	51.87	34.311(6)	<.001	82.513	2,5,6,7,9,11,12
2	10	940	.062(.074)	-.082	.206	.843(.399)	52.47	24.443(9)	.004	63.179	3,4,5,6,7,9,10,12,13,14
3	10	830	.358(.142)	.079	.638	2.516(.012)	63.98	70.861(9)	<.001	87.299	2,3,5,6,7,8,9,10,12,14
4	11	848	.312(.132)	.054	.570	2.369(.018)	62.25	61.537(10)	<.001	83.750	2,3,4,5,6,7,8,9,10,12,14
5	11	848	.213(.118)	-.019	.444	1.801(.072)	58.43	46.691(10)	<.001	78.583	2,3,4,5,6,7,8,9,10,12,14
6	13	1,195	.116(.087)	-.054	.286	1.340(.180)	54.62	53.906(12)	<.001	77.739	2,3,4,5,6,7,8,9,10,11,12,13,14
7	10	863	.083(.083)	-.080	.247	1.001(.317)	53.31	29.897(9)	<.001	69.897	2,3,4,5,6,8,9,12,13,14
8	12	1,123	.059(.070)	-.078	.196	.849(.396)	52.35	32.206(11)	.001	65.845	2,3,4,5,6,7,8,9,11,12,13,14
9	12	998	.117(.092)	-.063	.297	1.271(.204)	54.66	43.177(11)	<.001	74.523	2,3,4,5,6,7,8,9,10,11,12,13
10	7	516	.132(.050)	.034	.230	2.636(.008)	55.25	.939(6)	.988	<.001	2,3,4,8,9,12,14
11	6	422	.259(.170)	-.073	.592	1.529(.126)	60.22	9.022(5)	.108	44.583	2,3,7,9,10,12
12	10	750	.483(.173)	.145	.822	2.801(.005)	68.55	77.679(9)	<.001	88.414	2,3,4,6,7,8,9,10,12,14
13	11	1,052	-.052(.080)	-.209	.105	-.646(.518)	47.93	39.893(10)	<.001	74.933	2,3,4,5,6,8,9,11,12,13,14
14	10	1,083	-.074(.044)	-.161	.013	-1.673(.094)	47.05	13.366(9)	.147	32.664	2,3,5,6,7,9,11,12,13,14
15	11	984	-.079(.073)	-.221	.064	-1.086(.278)	46.85	25.205(10)	.005	60.326	2,3,4,5,6,7,9,10,12,13,14
16	2	62	-1.132(1.310)	-3.699	1.435	-.865(.387)	12.88	16.946(1)	<.001	94.099	2,4
17	5	702	-.026(.060)	-.143	.091	-.439(.661)	48.96	9.710(4)	.046	58.807	5,6,11,12,13
18	6	801	-.106(.037)	-.178	-.033	-2.862(.004)	45.78	1.976(5)	.852	<.001	2,3,11,12,13,14
전체	8	335	1.519(.216)	1.095	1.942	7.029(<.001)	93.56	16.655(7)	.020	57.972	1,2,3,4,7,8,9,10

주. k는 포함된 연구 수, N은 전체 표본 수, 효과크기는 Hedges' g 값(random), SE는 표준오차, U₃는 비중복 백분위 지수

먼저, CBCA 총점의 평균 효과크기를 분석한 결과, 그 크기는 1.519($p < .001$)로 매우 큰 수준이었다. U₃ 지수를 통해 효과크기를 해석해 보면, CBCA 총점의 백분위는 93.56%ile로, 작화 집단의 평균 백분위를 50%ile로 보았을 때, 진실 집단이 43.56%ile만큼 더 평균 점수가 높다는 것을 의미한다. 그러나 연구들 간 효과 크기의 이질성이 $Q=16.655(df=7, p=.020)$, $I^2=57.972\%$ 로 상당한 수준인 것으로 나타났다(표 2 참고). 표 3에는 CBCA 총점을 보고한 연구들의 효과크기가 제시되어있다.

이에 이질성의 원인을 탐색하기 위한 메타

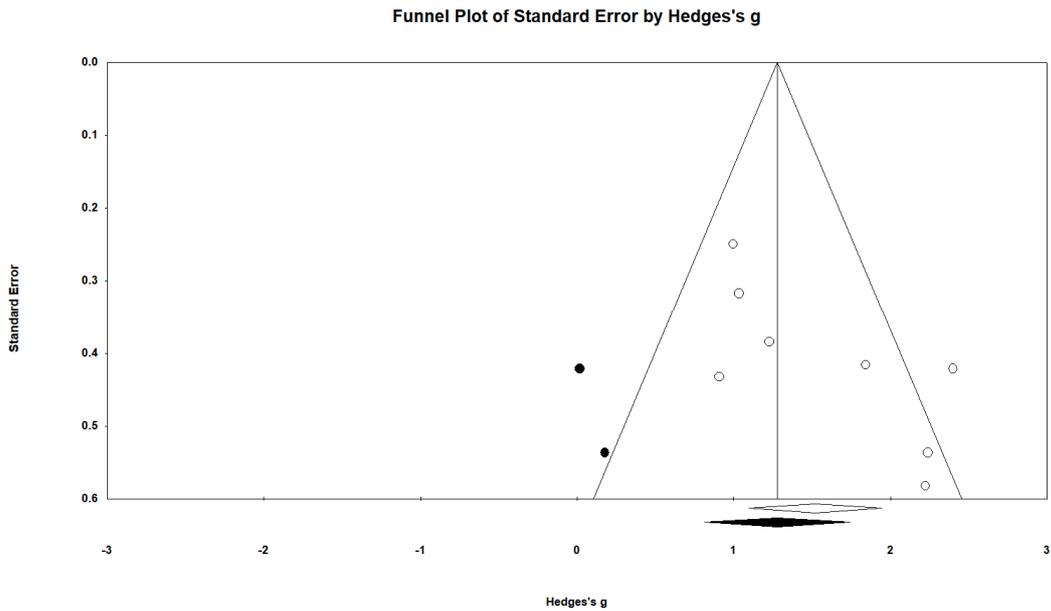
ANOVA를 실시하고자 했으나, CBCA 총점을 보고한 연구의 수가 8개로, 메타 ANOVA를 수행할 수 있는 기준인 10개의 연구 수에 미달하여(황성동, 2014), 대안으로 출판 편향의 가능성을 분석하였다.

먼저, funnel plot을 살펴보면(그림 2 참고), 연구의 표본 수가 적어 표준오차가 커질수록 효과크기가 크게 나타나는 경향이 있었다. 이를 검증하기 위해 Egger의 회귀분석을 시행한 결과(표 4 참고), 초기값이 4.082($p=.038$)로 출판편향 가능성이 있는 것으로 나타났다. 이에 Trim과 Fill 기법을 이용하여 출판편향을 교정

표 3. CBCA 총점을 보고한 연구들의 효과크기

연구 일련번호	N	효과크기 (SE)	95% CI		Z(p)	U ₃	Hedges' g 와 95% CI
			하한값	상한값			
9	22	.909(.432)	.061	1.757	2.101(.036)	81.83	
7	71	.995(.249)	.507	1.484	3.994(<.001)	84.01	
3	47	1.034(.317)	.413	1.655	3.263(.001)	84.94	
10	72	1.228(.385)	.473	1.982	3.190(.001)	89.03	
2	44	1.842(.416)	1.026	2.658	4.426(<.001)	96.73	
4	18	2.225(.582)	1.083	3.366	3.821(<.001)	98.70	
8	22	2.239(.537)	1.187	3.291	4.172(<.001)	98.74	
1	39	2.400(.422)	1.573	3.226	5.691(<.001)	99.18	
Fixed		1.392(.134)	1.130	1.654	10.417(<.001)	91.80	
Random		1.519(.216)	.047	1.095	7.029(<.001)	93.56	

주. 효과크기는 Hedges' g 값(random), SE는 표준오차, U₃는 비중복 백분위 지수



● : 삽입된 효과크기, ◇ : 평균 효과크기, ◆ : 교정된 평균 효과크기

그림 2. CBCA 총점 효과크기의 funnel plot

하였다(표 4 참고). 그 결과, 2개의 연구를 추가하여 교정한 이후 효과크기는 1.277로 나타났다. 즉, 출판편향 가능성을 고려해 보면 CBCA 총점의 효과크기가 약간 작아질 수 있

표 4. CBCA 총점 효과크기의 출판편향 검증

Egger의 회귀분석					
초기값	SE	95% CI		t	p
		하한값	상한값		
4.082	1.535	.325	7.839	2.659	.038

Trim과 Fill 기법					
추가된 연구	ES(g)	95% CI		Q	
		하한값	상한값		
관측값	1.519	1.095	1.942	16.655	
교정값	2	1.277	.814	1.740	30.205

음을 알 수 있다.

다음으로, CBCA 준거별로 평균 효과크기를 분석한 결과, 다섯 개의 준거(3번, 4번, 10번, 12번, 18번)에서 유의미한 효과크기가 산출되었다(표 2 참고). 구체적으로, 준거 3번의 평균 효과크기는 .358($p=.012$, $U_3=63.98$), 4번은 .312($p=.018$, $U_3=62.25$), 10번은 .132($p=.008$, $U_3=55.25$), 12번은 .483($p=.005$, $U_3=68.55$), 18번은 -.106($p=.004$, $U_3=45.78$)로 산출되었다. 이들 중, 3번, 4번, 12번 준거에서는 연구들 간 효과크기가 이질적인 것으로 나타나 해석에 주의를 요한다(이질성에 영향을 미치는 요인에 대해서는 다음 '준거별 조절효과분석' 부분에서 다루었다). 특히, 준거 18번은 연구들 간 효과크기가 이질적이지는 않았지만, 그 방향이 부적(-)인 것으로 나타났다. 즉, 효과크기가 -.106으로 그 크기는 작았지만, CBCA의 기본 가정과는 반대로 진실 집단보다 작화 집단에서 준거 관련 내용이 더 많이 현출된 것으로 나타났다.

준거별 조절효과분석

조절효과분석에 앞서, CBCA의 18개 준거 중 조절분석이 가능한 10개 이상의 연구가 포함된 12개 준거(2~9번, 12~15번) 각각에 대해 연구들 간 효과크기의 이질성 정도를 확인하였다(표 2 참고). 그 결과, 준거 2번에서는 $Q=24.443(df=9, p=.004)$, $I^2=63.179\%$, 준거 7번에서는 $Q=29.897(df=9, p<.001)$, $I^2=69.897\%$, 준거 8번에서는 $Q=32.206(df=11, p=.001)$, $I^2=65.845\%$, 준거 9번에서는 $Q=43.177(df=11, p<.001)$, $I^2=74.523\%$, 준거 13번에서는 $Q=39.893(df=10, p<.001)$, $I^2=74.933\%$, 준거 15번에서는 $Q=25.205(df=10, p=.005)$, $I^2=60.326\%$ 로 나타나 연구들간 효과 크기의 이질성 수준이 상당한 것으로 확인되었다.

그리고 3번 준거에서는 $Q=70.861(df=9, p<.001)$, $I^2=87.299\%$, 4번 준거에서는 $Q=61.537(df=10, p<.001)$, $I^2=83.750\%$, 5번 준거에서는 $Q=46.691(df=10, p<.001)$, $I^2=78.583\%$, 6번 준거에서는 $Q=53.906(df=12, p<.001)$, $I^2=77.739\%$, 12번 준거에서는 $Q=77.679(df=9,$

$p < .001$), $f^2 = 88.414\%$ 로 나타나 연구들간 효과 크기의 이질성이 큰 수준인 것으로 확인되었다. 그러나 준거 14번에서는 $Q = 13.366(df = 9, p = .147)$, $f^2 = 32.664\%$ 로 연구들 간 효과 크기의 이질성이 유의미하지 않은 것으로 나타났다.

이에, 연구들 간 효과크기가 이질적이라고 확인된 11개의 준거들(2~9번, 12번, 13번, 15번)에 대하여 조절변인(진술자 연령, 성별과

신분, 연구유형 및 설계, 사건유형, 평가자 수, 출판 유형)의 영향력을 각각 분석하였다. 조절변인에 따른 효과크기의 차이는 각 집단 내에서는 모집단 효과크기가 서로 다르다는 무선모형을 가정하였고, 집단 간에는 고정모형을 가정하고 계산한 값으로 분석하였다(황성동, 2014).

표 5부터 표 8까지, 준거별 조절변인들에

표 5. 조절효과 분석(인지적 준거 - 일반적 특징)

조절 변인	준거 2 정연하지 않은 진술				R^2	준거 3 세부정보의 풍부함				
	하위범주 (연구수)	효과크기(ρ)	효과크기 이질성 $Q(df)$ p			하위범주 (연구수)	효과크기(ρ)	효과크기 이질성 $Q(df)$ p		R^2
연령	성인(2)	-.056(.787)	.515(2)	.773	-.84	성인(1)	-.054(.933)	1.193(2)	.551	-1.956
	비성인(5)	.107(.428)				비성인(7)	.630(.018)			
	혼합(3)	.113(.460)				혼합(2)	.270(.583)			
성별	여성(1)	.602(.179)	1.497(1)	.221	.040	여성(2)	.842(.032)	1.839(1)	.175	.109
	혼합(9)	.046(.527)				혼합(8)	.274(.060)			
진술자 신분	피해(5)	.007(.944)	.488(2)	.784	-1.286	피해(7)	.586(.014)	2.497(2)	.287	-1.570
	가해(2)	.061(.573)				가해(1)	-.204(.712)			
	혼합(2)	-.062(.658)				혼합(1)	-.054(.922)			
연구 유형	실험(7)	.081(.307)	.463(1)	.496	-.080	실험(8)	.255(.080)	1.943(1)	.163	.161
	현장(3)	-.085(.714)				현장(2)	.770(.024)			
연구 설계	집단간(5)	.331(.054)	3.018(1)	.082	.160	집단간(6)	1.014(<.001)	44.647(1)	<.001	.876
	집단내(5)	.005(.945)				집단내(4)	-.167(.033)			
사건 유형	성폭력(3)	.070(.759)	.019(2)	.990	-.160	성폭력(5)	1.040(<.001)	22.638(1)	<.001	.635
	일상(6)	.067(.421)				일상(5)	-.061(.593)			
	혼합(1)	.000(1.000)				혼합(0)	-			
평가자	1인(2)	.234(.400)	.407(1)	.524	-.040	1인(2)	.791(.026)	1.920(1)	.166	.146
	2인이상(8)	.050(.520)				2인이상(8)	.260(.073)			
출판 유형	학술지(2)	.908(<.001)	12.760(1)	<.001	.808	학술지(4)	1.178(<.001)	24.750(1)	<.001	.672
	비학술지(8)	.004(.941)				비학술지(6)	-.039(.714)			

주. 효과크기는 Hedges' $g_{\text{fixed}}(\text{random})$, $R^2 = \left(\frac{T_{\text{between}}^2}{T} \right)$

표 6. 조절효과 분석(인지적 준거 - 구체적 내용)

준거 4 사건이 발생하게된 맥락 정보						준거 5 상호작용 묘사				
조절 변인	하위범주 (연구수)	효과크기(ρ)	효과크기 이질성		R^2	하위범주 (연구수)	효과크기(ρ)	효과크기 이질성		R^2
			$Q(df)$	p				$Q(df)$	p	
연령	성인(2)	.550(.238)				성인(2)	.465(.238)			
	비성인(7)	.472(.057)	.141(2)	.932	-1.904	비성인(7)	.383(.078)	.606(2)	.738	-1.732
	혼합(2)	.316(.485)				혼합(2)	.080(.833)			
성별	여성(2)	1.344(<.001)	9.086(1)	.003	.333	여성(2)	.814(.025)	3.268(1)	.071	.159
	혼합(9)	.138(.247)				혼합(9)	.125(.284)			
진술자 신분	피해(7)	.497(.023)				피해(7)	.325(.076)			
	가해(1)	-.340(.492)	2.476(2)	.290	-1.167	가해(1)	-.305(.445)	2.310(2)	.315	-.961
	혼합(2)	.487(.228)				혼합(2)	.402(.229)			
연구 유형	실험(8)	.108(.390)	7.038(1)	.008	.325	실험(8)	.026(.810)	8.356(1)	.004	.378
	현장(3)	.910(.001)				현장(3)	.826(.001)			
연구 설계	집단간(7)	.917(<.001)	40.117(1)	<.001	.877	집단간(7)	.768(<.001)	23.777(1)	<.001	.744
	집단내(4)	-.179(.015)				집단내(4)	-.134(.114)			
사건 유형	성폭력(5)	.874(<.001)	32.664(2)	<.001	.789	성폭력(5)	.666(<.001)	19.953(2)	<.001	.659
	일상(5)	-.141(.108)				일상(5)	-.107(.253)			
	혼합(1)	1.632(.003)				혼합(1)	1.371(.010)			
평가자	1인(3)	1.057(<.001)	9.500(1)	.002	.351	1인(3)	.960(<.001)	11.214(1)	.001	.451
	2인이상(8)	.093(.451)				2인이상(8)	.010(.925)			
출판 유형	학술지(4)	.984(<.001)	12.646(1)	<.001	.474	학술지(4)	.742(.004)	6.122(1)	.013	.293
	비학술지(7)	.027(.817)				비학술지(7)	.051(.662)			
준거 6 대화의 재연						준거 7 사건 도중에 예기치 않은 상황 발생				
조절 변인	하위범주 (연구수)	효과크기(ρ)	효과크기 이질성		R^2	하위범주 (연구수)	효과크기(ρ)	효과크기 이질성		R^2
			$Q(df)$	p				$Q(df)$	p	
연령	성인(2)	.230(.400)				성인(2)	.296(.197)			
	비성인(8)	.239(.087)	.999(2)	.607	-.717	비성인(5)	.130(.386)	1.628(2)	.443	-.686
	혼합(3)	-.000(.999)				혼합(3)	-.057(.737)			
성별	여성(2)	1.018(.003)	7.461(1)	.006	.226	여성(2)	-.129(.693)	.455(1)	.500	-.057
	혼합(11)	.041(.614)				혼합(8)	.100(.255)			
진술자 신분	피해(7)	.254(.020)				피해(6)	.180(.008)			
	가해(2)	-.276(.042)	9.384(2)	.009	.364	가해(2)	-.172(.021)	15.227(2)	<.001	.914
	혼합(3)	-.001(.991)				혼합(2)	.207(.028)			

표 6. 조절효과 분석(인지적 준거 - 구체적 내용) (계속)

준거 6 대화의 재연						준거 7 사건 도중에 예기치 않은 상황 발생				
조절 변인	하위범주 (연구수)	효과크기(ρ)	효과크기 이질성		R^2	하위범주 (연구수)	효과크기(ρ)	효과크기 이질성		R^2
			$Q(df)$	p				$Q(df)$	p	
연구 유형	실험(10)	.069(.440)	1.725(1)	.189	.075	실험(8)	.039(.639)	3.329(1)	.068	.114
	현장(3)	.418(.094)				현장(2)	.582(.042)			
연구 설계	집단간(7)	.711(<.001)	21.878(1)	<.001	.604	집단간(5)	.294(.125)	1.488(1)	.223	.029
	집단내(6)	-.090(.187)				집단내(5)	.035(.704)			
사건 유형	성폭력(5)	.605(.001)	14.236(2)	.001	.415	성폭력(4)	.211(.308)	2.576(2)	.276	.029
	일상(7)	-.050(.520)				일상(5)	.035(.704)			
	혼합(1)	1.060(.039)				혼합(1)	.779(.121)			
평가자	1인(3)	.794(.002)	8.247(1)	.004	.264	1인(3)	.523(.031)	3.751(1)	.053	.143
	2인이상(10)	.017(.835)				2인이상(7)	.028(.740)			
출판 유형	학술지(4)	.991(<.001)	20.811(1)	<.001	.585	학술지(3)	.068(.791)	.005(1)	.946	-.057
	비학술지(9)	-.048(0.468)				비학술지(7)	.087(.337)			

주. 효과크기는 Hedges' $g_{adj}(random)$, $R^2 = (\frac{T^2_{between}}{T})$

표 7. 조절효과 분석(인지적 준거 - 내용의 특수성)

준거 8 일상적이지 않은 독특한 세부정보						준거 9 여분의 세부정보				
조절 변인	하위범주 (연구수)	효과크기(ρ)	효과크기 이질성		R^2	하위범주 (연구수)	효과크기(ρ)	효과크기 이질성		R^2
			$Q(df)$	p				$Q(df)$	p	
연령	성인(2)	.064(.761)	2.278(2)	.320	-.704	성인(2)	.312(.222)	.517(2)	.772	-.453
	비성인(7)	.212(.071)				비성인(8)	.108(.406)			
	혼합(3)	-.076(.617)				혼합(2)	.176(.486)			
성별	여성(2)	.496(.127)	1.929(1)	.165	.074	여성(2)	.901(.008)	6.132(1)	.013	.264
	혼합(10)	.034(.624)				혼합(10)	.036(.673)			
진술자 신분	피해(6)	.243(.035)	4.887(2)	.087	-.192	피해(6)	.330(.052)	3.334(2)	.189	-.609
	가해(2)	-.145(.297)				가해(2)	-.148(.467)			
	혼합(3)	.001(.994)				혼합(3)	.071(.706)			
연구 유형	실험(10)	-.003(.965)	7.166(1)	.007	.370	실험(9)	.021(.812)	5.025(1)	.025	.245
	현장(2)	.752(.006)				현장(3)	.601(.013)			
연구 설계	집단간(6)	.737(<.001)	20.939(1)	<.001	.963	집단간(7)	.651(<.001)	22.806(1)	<.001	.792
	집단내(6)	-.056(.127)				집단내(5)	-.113(.067)			

표 7. 조절효과 분석(인지적 준거 - 내용의 특수성)

(계속)

준거 8 일상적이지 않은 독특한 세부정보						준거 9 여분의 세부정보				
조절 변인	하위범주 (연구수)	효과크기(p)	효과크기 이질성		R ²	하위범주 (연구수)	효과크기(p)	효과크기 이질성		R ²
			Q(df)	p				Q(df)	p	
사건 유형	성폭력(4)	.720(<.001)	18.316(2)	<.001	.815	성폭력(5)	.595(.001)	14.975(2)	.001	.566
	일상(7)	-.052(.241)				일상(6)	-.076(.324)			
	혼합(1)	.814(.086)				혼합(1)	.952(.057)			
평가자	1인(3)	.871(<.001)	14.749(1)	<.001	.704	1인(3)	.664(.008)	5.985(1)	.014	.283
	2인이상(9)	-.035(.477)				2인이상(9)	.016(.857)			
출판 유형	학술지(4)	.728(.002)	9.833(1)	.002	.481	학술지(4)	.701(.001)	10.665(1)	.001	.453
	비학술지(8)	-.020(.728)				비학술지(8)	-.032(.686)			
준거 12 주관적인 심리 상태						준거 13 가해자의 정신 상태 귀인				
조절 변인	하위범주 (연구수)	효과크기(p)	효과크기 이질성		R ²	하위범주 (연구수)	효과크기(p)	효과크기 이질성		R ²
			Q(df)	p				Q(df)	p	
연령	성인(2)	.276(.689)	1.229(2)	.541	-3.288	성인(2)	.257(.198)	2.908(2)	.234	-.051
	비성인(6)	.925(.023)				비성인(6)	-.098(.388)			
	혼합(2)	.179(.793)				혼합(3)	-.139(.344)			
성별	여성(2)	1.251(.004)	3.840(1)	.050	.182	여성(2)	-.252(.483)	.329(1)	.566	-.026
	혼합(8)	.329(.057)				혼합(9)	-.041(.626)			
진술자 신분	피해(7)	.786(.003)	1.050(1)	.306	-.831	피해(6)	-.102(.496)	.614(2)	.736	-.718
	가해(0)	-				가해(2)	-.082(.674)			
	혼합(2)	.219(.651)				혼합(3)	.074(.682)			
연구 유형	실험(7)	.333(.070)	1.579(1)	.209	.197	실험(9)	-.093(.248)	3.503(1)	.061	.103
	현장(3)	.795(.012)				현장(2)	.474(.104)			
연구 설계	집단간(7)	.943(<.001)	19.619(1)	<.001	.601	집단간(5)	.107(.594)	.744(1)	.388	.000
	집단내(3)	-.175(.301)				집단내(6)	-.082(.351)			
사건 유형	성폭력(5)	1.217(<.001)	36.230(2)	<.001	.783	성폭력(4)	-.057(.791)	4.780(2)	.092	.103
	일상(4)	-.181(.142)				일상(6)	-.082(.329)			
	혼합(1)	.868(.092)				혼합(1)	1.065(.040)			
평가자	1인(3)	1.127(.001)	5.916(1)	.015	.303	1인(3)	.623(.011)	8.369(1)	.004	.256
	2인이상(7)	.232(.176)				2인이상(8)	-.119(.114)			
출판 유형	학술지(4)	1.121(<.001)	8.421(1)	.004	.328	학술지(3)	-.207(.447)	.359(1)	.549	-.026
	비학술지(6)	.151(.392)				비학술지(8)	-.036(.671)			

주. 효과크기는 Hedges' $g_{adj}(\text{random})$, $R^2 = \left(\frac{T^2_{between}}{T} \right)$

다른 각 하위 범주의 효과크기 및 하위 범주 들 간 효과크기의 이질성을 확인할 수 있다. 조절효과 분석에 포함된 연구수가 적을 경우 해석에 유의할 필요가 있고, 부적(-)인 방향으

표 8. 조절효과 분석(동기적 준거 - 동기 관련 내용)

준거 15 기억부족의 인정					
조절 변인	하위범주 (연구수)	효과크기(p)	효과크기 이질성		R ²
			Q(df)	p	
연령	성인(2)	-.181(.390)			
	비성인(6)	-.011(.935)	.483(2)	.785	-.920
	혼합(3)	-.081(.602)			
성별	여성(1)	.225(.610)	.485(1)	.486	-.040
	혼합(10)	-.087(.244)			
진술자 신분	피해(6)	.022(.841)			
	가해(2)	-.220(.104)	2.182(2)	.336	-.217
	혼합(2)	-.176(.309)			
연구 유형	실험(8)	-.042(.566)	2.955(1)	.086	.120
	현장(3)	-.454(.046)			
연구 설계	집단간(6)	.039(.816)	.613(1)	.434	-.040
	집단내(5)	-.107(.194)			
사건 유형	성폭력(4)	-.005(.979)			
	일상(6)	-.083(.320)	.292(2)	.864	-.160
	혼합(1)	-.269(.576)			
평가자	1인(2)	-.533(.054)	2.850(1)	.091	.120
	2인이상(9)	-.050(.485)			
출판 유형	학술지(3)	.575(.013)	8.883(1)	.003	.560
	비학술지(8)	-.133(.023)			

주. 효과크기는 Hedges' *g*(random).

$$R^2 = \left(\frac{T_{between}^2}{T} \right)$$

로 효과크기가 나타난 것에 대해서는 주의 깊게 살펴볼 필요가 있다.

먼저, 인지적 준거 중 일반적 특징 요소에 해당하는 준거들(2, 3번)에 대해 살펴보았다. 연구설계의 조절효과는 준거 3번에서만 나타났다($p < .001$), 연구설계의 집단 구분에 의해 설명되는 분산(R^2)은 87.6%이었다. 구체적으로,

집단간 연구설계에서는 정적방향으로 큰 효과크기가 나타난 반면, 집단내 설계에서는 작지만 부적인 효과크기가 나타났다. 사건유형의 조절효과도 준거 3번에서만 나타났는데 ($p < .001$, $R^2 = .635$), 성폭력 사건에서만 유의미하게 큰 효과크기를 보였다.

출판유형에 따른 조절효과는 준거 2번 ($p < .001$, $R^2 = .808$)과 3번($p < .001$, $R^2 = .672$)에서 나타났다. 모두 학술지에서만 크고 통계적으로 유의미한 효과크기를 보였다.

다음으로, 인지적 준거 중 구체적인 내용 요소에 해당하는 준거들(4, 5, 6, 7번)을 살펴보았다. 진술자의 성별에 따른 조절효과는 준거 4번($p = .003$, $R^2 = .333$)과 6번($p = .006$, $R^2 = .226$)에서 나타났다. 모두 진술자가 여성일 때에만 유의미하고 큰 효과크기를 보였다.

진술자의 신분에 따른 조절효과는 준거 6번 ($p = .009$, $R^2 = .364$)과 7번($p < .001$, $R^2 = .914$)에서 나타났다. 구체적으로, 두 준거에서 모두 피해자는 정적으로, 가해자는 부적방향으로 작은 효과크기가 나타났다. 혼합조건의 효과크기는 준거 7번에서만 정적방향으로 작게 나타났다.

연구유형에 따른 조절효과는 준거 4번 ($p = .008$, $R^2 = .325$)과 5번($p = .004$, $R^2 = .378$)에서 나타났다. 두 준거에서 모두 현장연구일 때에만 유의미하고 큰 효과크기를 보였다.

연구설계에 따른 조절효과는 준거 4번 ($p < .001$, $R^2 = .877$), 5번($p < .001$, $R^2 = .744$), 6번 ($p < .001$, $R^2 = .604$)에서 나타났다. 구체적으로, 준거 4에서는 집단간 연구설계에서 큰 효과크기가 나타났다. 그러나 집단내 설계에서는 작지만 부적으로 유의한 효과크기가 나타났다. 준거 5번과 6번에서는 집단간 연구설계에서만 정적으로 상당히 큰 효과크기가 나타났다.

사건유형에 따른 조절효과는 준거 4번

($p < .001$, $R^2 = .789$), 5번($p < .001$, $R^2 = .659$), 6번($p = .001$, $R^2 = .415$)에서 나타났다. 모두 혼합사건인 경우에 큰 효과크기를 보였고, 성폭력 사건에서는 상당히 큰 효과크기를 보였다. 일반사건에서는 세 준거에서 모두 유의미한 효과크기를 산출하지 못했다.

평가자 유형에 따른 조절효과는 준거 4번($p = .002$, $R^2 = .351$), 5번($p = .001$, $R^2 = .451$), 6번($p = .004$, $R^2 = .264$)에서 나타났다. 모든 준거에서 평가자가 1인일 때에는 큰 효과크기를 보인 반면, 2인 이상일 때에는 유의미한 효과크기를 산출하지 못했다.

출판유형에 따른 조절효과는 준거 4번($p < .001$, $R^2 = .474$), 5번($p = .013$, $R^2 = .293$), 6번($p < .001$, $R^2 = .585$)에서 나타났다. 모두 학술지인 경우에만 큰 효과크기를 산출했고, 비학술지인 경우에는 유의미한 효과크기가 나타나지 않았다.

다음으로, 인지적 준거 중 내용의 특수성 요소에 해당하는 준거들(8, 9, 12, 13번)을 살펴보았다. 진술자의 성별에 따른 조절효과는 준거 9번에서만 나타났고($p = .013$, $R^2 = .264$), 여성이 진술자일 때만 유의미하고 큰 효과크기가 산출되었다.

연구유형에 따른 조절효과는 준거 8번($p = .007$, $R^2 = .370$)과 9번($p = .025$, $R^2 = .245$)에서 나타났다. 두 준거에서 모두 현장연구일 때에만 유의미하고 다소 큰 효과크기를 산출했다.

연구설계에 따른 조절효과는 준거 8번($p < .001$, $R^2 = .963$), 9번($p < .001$, $R^2 = .792$), 12번($p < .001$, $R^2 = .601$)에서 나타났다. 모든 준거에서, 집단간 설계일 때에만 유의미하고 상당히 큰 효과크기를 산출했다.

사건유형에 따른 조절효과는 준거 8번($p < .001$, $R^2 = .815$), 9번($p = .001$, $R^2 = .566$), 12번

($p < .001$, $R^2 = .783$)에서 나타났다. 모든 준거에서 성폭력 사건인 경우에 상당히 큰 효과크기를 보였고, 혼합사건인 경우에는 준거 13번에서만 유의미하고 큰 효과크기를 나타냈다. 일반사건일 때에는 모든 준거에서 유의미한 효과크기를 산출하지 못했다.

평가자에 따른 조절효과는 8번($p < .001$, $R^2 = .704$), 9번($p = .014$, $R^2 = .283$), 12번($p = .015$, $R^2 = .303$), 13번($p = .004$, $R^2 = .256$) 준거에서 모두 나타났다. 모든 준거에서 1인이 평가할 때에만 통계적으로 유의하고 상당히 큰 효과크기를 산출했다.

출판유형에 따른 조절효과는 준거 8번($p = .002$, $R^2 = .481$), 9번($p = .001$, $R^2 = .453$), 12번($p = .004$, $R^2 = .328$)에서 나타났다. 모두 학술지인 경우에만 유의미하고 상당히 큰 효과크기를 산출했다.

마지막으로, 동기적 준거 중 동기관련내용 요소에 해당하는 준거 15번을 살펴보았다. 이 준거에서는 출판유형에 따른 조절효과만 나타났다($p = .003$, $R^2 = .560$). 구체적으로, 학술지일 때에는 중간크기의 정적인 효과크기가 나타났다. 그러나 비학술지인 경우에는 작지만 부적으로 유의미한 효과크기가 나타났다.

논 의

본 연구에서는 국내외 수사현장 및 법정 단계에서 활용되고 있는 SVA에서 사용하는 CBCA 준거에 관한 국내 연구들을 종합적으로 살펴보는 메타분석을 수행하였다.

본 연구의 주요 결과는 다음과 같다. 먼저, CBCA의 총점은 출판편향을 고려하여 교정한 이후에도 진술의 진실성을 성공적으로 변별하

는 것으로 나타났고, 그 평균 효과크기 또한 큰 수준이었다($g=1.277$). 개별 준거에서는 진실과 작화를 통계적으로 유의하게 구분 짓는 준거가 다섯 개였다. 이 중, 준거 3번, 4번, 10번, 12번은 인지적 준거에 해당하고, 18번은 동기적 준거에 해당하는 것이다.

구체적으로, 그 평균 효과크기는 3번 세부정보의 풍부함(.358), 4번 사건이 발생하게 된 맥락정보(.312), 10번 이해하지는 못했으나 정확하게 보고한 세부 정보(.132), 12번 주관적인 심리 상태(.483), 18번 가해자 용서(-.106)로 나타났다. 그런데, 이 중에서 3번, 4번, 10번 준거의 효과크기는 연구들 간 이질성이 컸기 때문에 조절변인의 영향을 고려하여 신중하게 해석하여야 한다. 특히, 18번 준거의 평균 효과크기가 부정(-)으로 나타났다는 것은 매우 중요한 문제이다. 이는 CBCA의 가정과는 반대로 진실 집단보다 작화 집단에서 관련 준거의 내용이 더 많이 나타났다는 것을 의미하기 때문이다.

이러한 결과는 Amado 등(2015)이 수행한 메타연구에서 모든 CBCA 준거가 유의미한 양적 효과크기를 보였다는 것과는 다소 상이한 결과이다. 하지만 이 연구에서도 인지적 준거들(1번~13번)의 효과크기($d_s=.33\sim.87$)에 비해 동기적 준거들(14번~18번)의 효과크기가 작은 수준($d_s=.17\sim.25$)에 머물렀다는 것은 인지적 준거에 비해 동기적 준거의 진위 변별 타당성이 비교적 떨어지는 것이라고 볼 수 있다. 이를 넓게 해석하면 본 연구결과와 일면 맥을 같이 한다고 볼 수 있다.

심지어, Raskin 등(1991)은 실제 경험한 사건에 대한 기억과 직접적으로 관련된 1번~14번 준거(Underschied의 가설에 부합하는 준거)만 사용하고 나머지 준거들은 진술의 진위 판단을

위해 사용하지 말 것을 제안하기도 했다. 이를 볼 때, 본 연구결과는 추후 실무에서 사건 당사자들 진술의 진실성을 판단할 때에는 동기적 준거보다는 인지적 준거에 좀 더 초점을 맞추어야 함을 시사하는 것이다.

또한, 본 연구에서는 CBCA 준거별로 총 8개(진술자의 연령, 성별, 및 신분, 연구의 유형 및 설계, 사건유형, 평가자의 수, 출판유형)의 잠재적 조절변인을 선정하여 그 영향력을 확인했다. 그 결과, 첫째, 진술자의 연령은 분석 가능한 모든 준거들에서 아무런 영향력도 발휘하지 못했다. 이는 최근의 메타연구(Oberlander, 2019)에서 연령에 따른 조절효과가 나타나지 않은 것과 일치하는 것이다.

그러나, Oberlander(2019)의 연구는 진술자를 18세 미만과 이상 집단으로 나누었고, 본 연구에서는 비성인, 성인, 혼합(아동, 청소년, 성인 포함) 집단으로 나누어 분석하였기 때문에 아동들과 성인들 간 차이가 없다고 주장하기에는 제한점이 크다. 많은 연구자들(김미영, 2016; Santtila, Roppola, Runtti, & Niemi, 2000; Vrij et al., 2004; Wells & Loftus, 1991)이 연령이 증가할수록 진술능력 및 인상관리 능력이 발달하여 CBCA 점수의 변별력이 떨어질 수 있다는 것을 지적하고 있다. 또한, Piaget의 인지발달이론에 의하면 아동청소년 시기 또한 세부 연령대별로 전조작기와 구체적 조작기, 나아가 형식적 조작기 간에 인지/추론 및 언어 능력에 있어 큰 차이가 존재한다(Huitt & Hummel, 2003). 따라서 아동, 청소년, 성인 간 진술 능력이나 특성 차이에 따른 영향이 있을 것으로 예상되므로 발달단계나 연령대를 보다 세분화하여 심도 있는 연구를 수행해야 할 필요가 있다.

둘째, 성별에서의 조절효과는 여성만을 대

상으로 할 때 남녀가 혼합되어 있는 조건보다 통계적으로 유의한 차이가 4개 준거(4번, 6번, 9번, 12번)에서 나타났다. 이는 일반적으로 여성이 더 뛰어난 언어적 묘사 능력을 보일 수 있다는 점(Hyde & Linn, 1988)과 맥락이 유사하다고 볼 수 있다. 그러나 진술 신빙성의 평가 맥락에서 성차가 통계적으로 유의미하게 나타나지 않았다는 이전 연구결과(예, Roma, San Martini, Sabatello, Tatarelli, & Ferracuti, 2011; Oberlader, 2019)와 본 연구에서 여성만을 대상으로 한 연구자료가 2개밖에 포함되지 않았다는 점에서는 해석에 유의할 필요가 있다.

셋째, 실험연구와 현장연구 간 차이를 살펴본 결과, 4개 준거(4번, 5번, 8번, 9번)에서 조절효과가 나타났다. 모두 현장연구가 실험연구보다 더 큰 효과크기를 보였는데, 이는 대체로 실험연구보다는 현장연구에서 진실한 진술과 작화된 진술간의 차이가 더 크게 나타나는 경향이 있다는 연구들(Amado et al., 2015; 2016; Oberlader et al., 2021; Vrij, 2005)과 일치하는 결과이다.

넷째, 연구설계에서는 7개의 준거(3번, 4번, 5번, 6번, 8번, 9번, 12번)에서 조절효과가 나타났다. 모두 집단간 설계가 집단내 연구설계보다 큰 효과크기를 보였다. 이는 Oberlader (2019)의 메타분석 연구와 일치하는 결과이다. 심지어, 집단내 설계에서는 3번과 4번 준거에서 작지만 유의미한 부적 효과크기가 산출되었다. 즉, CBCA의 기본 가정과는 반대로, 작화 집단에서 진실집단보다 준거 관련 내용이 더 많이 산출되었다는 것이다. 따라서 집단내 설계로 수행된 연구결과를 해석할 때에는 좀 더 주의를 기울여야 할 필요가 있다.

다섯째, 사건유형에 따른 조절효과가 7개 준거(3번, 4번, 5번, 6번, 8번, 9번, 12번)에서

나타났는데, 모두 성폭력 사건이 일반사건보다 일관되게 더 큰 효과크기를 보였다. 이는 Amado 등(2016)의 메타분석에서 성범죄 사건에서 일반 사건에서보다 더 큰 효과크기를 산출했다는 결과와 일치한다.

여섯째, 진술자 신분에 따라서는 6번과 7번 준거에서 조절효과가 나타났다. 즉, 모두 피해자의 진술에서는 정적인 효과가 가해자의 진술에서는 부적인 효과가 나타났다. 그러나 진술인의 신분이라는 조절변인의 영향 가능성은 그 조작적 정의에 따라 상이할 수 있다. 연구설계에 따라 피해자가 진실집단, 피의자가 작화집단으로 분류되기도 하고(윤대중, 2019) 피해자와 가해자가 진실과 작화집단 모두에 포함(윤여훈, 2008; 홍성열, 2007)되기도 하기 때문이다.

진술인의 신분에 상관없이 CBCA 총점이 진실한 진술에서 더 높게 나타났다는 연구 결과(Amado et al., 2016; Vrij et al., 2002; 2004)를 고려하면, 신분을 막론하고 진실한 사람의 진술은 CBCA 준거를 더 많이 포함할 가능성이 높다는 결론을 내릴 수 있다. 그러므로 본 연구 결과는 CBCA 준거 관련 연구에 있어 최대한 동일한 조작적 정의를 사용하는 것이 중요하다는 점을 역설적으로 보여준다고 해석할 수 있다.

그리고 특히 수사자료를 활용하는 현장연구의 경우, 진술자 신분의 영향을 연구하기에 제한점이 더욱 클 수 있다. CBCA의 평가 대상이 되는 진술자료는 자유진술을 중심으로 하는 면담과정을 통해 확보되어야 하지만 일반적으로 수사기관에서의 피의자 조사 방식은 그렇지 못할 가능성이 있다(Vrij, 2005)는 점을 유념해야 하기 때문이다. 이에 진술인 신분의 조절효과에 대한 해석에는 각별한 주의를 기

울일 필요가 있다.

일곱째, 평가자가 2인 이상일 때 보다 1인이 평가할 때 7개 준거(4번, 5번, 6번, 8번, 9번, 12번, 13번)에서 더 큰 효과크기가 나타났다. 이러한 결과는 CBCA의 객관성을 담보하기 위해 다수의 평가자가 준거들을 평정하고 평가자간 신뢰도를 보고하는 국외연구들(Boychuk, 1991; Hauch et al., 2017; Wells & Loftus, 1991)의 방법론을 국내 연구에서도 따라야 할 필요성을 반증하는 것이라 할 수 있다.

여덟째, 출판유형에 따른 조절효과가 9개의 준거(2번, 3번, 4번, 5번, 6번, 8번, 9번, 12번, 15번)에서 발견되었다. 즉, 이들 준거에서 모두 학술지인 경우에 더 큰 효과크기가 나타났다. 이는 일반적으로 효과크기가 크거나 통계적으로 유의한 연구들이 출판되고 그렇지 않은 연구들은 결측될 가능성이 높다는 일반적인 연구 상식과 일치한다(홍세희, 2013; 황성동, 2014). 심지어 15번 준거(기억부족의 인정)에서는 비학술지인 경우에 부적인 효과크기가 산출되었다. 따라서 동기적 준거 중 하나인 '기억부족의 인정' 준거를 실무에 적용하고 해석할 때에는 각별한 주의를 기울일 필요가 있다.

국외에서는 그간 다양한 연구방법으로 수행된 CBCA 연구들에 대한 메타분석이 이루어져 왔다. 그러나 국내에 SVA가 도입되어 CBCA 준거가 실무와 학계에서 논의된 지 20여 년 가까이 되고 법원에서의 활용도도 점차 증가하고 있음에도 불구하고, 국내의 CBCA 연구들에 대한 메타분석은 수행되지 못하였다. 이러한 상황에, 본 연구가 CBCA 준거에 관한 국내 연구들을 체계적으로 통합한 메타분석을 수행하여 준거별로 효과크기를 비교 분석하고 이

에 영향을 미치는 조절변인들을 밝혔다는 것은 본 연구의 가장 큰 의의이다.

또한, 본 메타분석은 국내에서 수행된 연구들의 방법론적 한계점을 추론할 수 있게 한다. 첫째, 진실집단과 작화집단 간 동등성 가정의 문제이다. 즉, 일부 연구들에서 진실집단에 비해 작화집단의 수가 너무 적거나 두 집단 간 연령에서의 불균형이 발견된다. 또한, 일부 연구들에서는 진실집단으로 수사기관에 피해 사실을 신고한 사람들을 설정하고 이에 대응되는 작화집단은 실험적으로 피해를 상상하게 한 사람들을 설정하고 있다. 이러한 두 집단의 진실은 그 이익이나 위험(stake)이 확연히 달라 직접적으로 비교할 수 없는 것이다.

둘째, 진실 확보 방법과 관련된 문제이다. 일부 연구들에서는 진실집단과 작화집단 간 진실을 확보한 방법(진술서, 진술녹화물, 진술조서, 또는 직접 면담 실시), 면담기법(NICHD 프로토콜, 표준 경찰 면담, 또는 인지면담 등)이 상이하거나, 어떤 방법으로 진실을 확보하였는지 명확히 밝히지 않고 있다. 진실확보 방식과 면담기법에 따라 진실의 내용이나 질이 달라질 수 있다는 점을 고려했을 때(김미영, 2016), 이는 심각한 문제이다.

셋째, 진실의 평가와 관련된 문제이다. 많은 연구들이 평가자의 전문성 관련 정보나 평가자의 훈련 방식 등에 대해 구체적으로 밝히고 있지 않다. 또한, 평가자가 1인인 경우(대부분이 연구의 저자)가 많고 2인 이상이 평가했다 하더라도 평가자간 신뢰도를 보고한 경우가 극히 드물다. 이는 진실의 평가 결과에 대한 신뢰성과 타당성을 확신할 수 없게 만드는 중요한 문제이다.

이러한 의의에도 불구하고, 본 연구는 몇 가지 한계가 있다. 첫째, CBCA 연구가 수행되

기 시작한 2004년부터 2020년까지의 관련 연구들을 총망라했음에도 불구하고 본 메타분석에 포함된 연구자료는 14개에 불과하다는 것이 가장 큰 한계이다. 특히, CBCA 총점을 보고한 연구자료는 8개에 불과하여 조절효과 분석을 수행하지 못했다.

둘째, 국외에서 수행된 메타분석 연구들은 CBCA 총점만을 바탕으로 조절효과를 분석하고 있기 때문에, 준거별로 분석한 본 연구의 결과와 직접적으로 비교할 수 없다는 한계가 있다. 따라서 본 연구의 결과를 현장에 적용할 때에는 세심한 주의가 필요할 것이다.

그러나 한편, 그간 국외에서는 CBCA 총점을 활용한 연구가 광범위하게 이루어졌던 것과는 달리, 실무적으로는 진술의 진실성을 판단하는데 총점은 활용도가 낮고 권고되지도 않는다(Hauch et al., 2017; Oberlander et al., 2021). 따라서 국외 연구와의 직접 비교를 위해 CBCA 준거를 모두 활용하는 연구도 필요하겠지만, 보다 중요하게는 CBCA의 세부 준거 및 조절변인에 대한 논의에 초점을 맞추는 연구가 보다 활발히 수행될 필요가 있다.

참고문헌⁹⁾

경찰청 (2017). 진술분석전문가 의견 작성 매뉴얼. 서울: 경찰청.

*고은영, 채규만 (2011). 성폭력 피해 아동의 진술에 대한 준거기반 내용분석의 활용 가능성 연구. 한국범죄심리연구, 7, 5-22.

김미영 (2016). 진술분석 절차에 관한 연구. 경기대학교대학원 박사학위논문.

*김현정 (2009). 성폭력 피해 아동의 진술 신빙성 평가도구의 상대적 유용성 비교. 경기대학교대학원 박사학위논문.

*김현정 (2010). CBCA와 RM을 이용한 성폭력 피해 아동의 진술 신빙성 평가. 한국심리학회지: 여성, 15(3), 355-379.

신우중 (2015). 체계적 고찰과 메타분석의 개요. Hanyang Med Review, 35(1), 9-17.

*윤대중 (2019). 통합적 심리증거 수사기법의 효율성에 대한 사례연구. 충남대학교대학원 박사학위논문.

*윤여훈 (2008). 아동들의 가해·피해 진술에 대한 준거기반 내용분석의 타당성 연구. 강원대학교대학원 석사학위논문.

*이미선 (2018). 비밀을 이야기 하지 않는 아동의 진술특성: 준거기반내용분석 (CBCA)을 중심으로. 피해자학연구, 26(3), 71-95.

*이수정 (2010). 아동 성폭력 피해 진술에 대한 신빙성 분석도구들의 타당도 연구. 한국심리학회지: 사회 및 성격, 24(2), 105-116.

*이수정, 유진옥 (2010). 지적장애인 성폭력 피해자의 진술 특성에 대한 실증연구. 한국경찰연구, 9(2), 161-184.

정의철, 이창호 (2017). 아동학대 사건에 대한 뉴스 프레임 분석과 대안 모색: 평택 아동학대 사건을 중심으로. 한국방송학보, 31(2), 77-108.

*조은경 (2004). 성폭력 피해 아동의 진술 타당도 분석 및 활용방안에 관한 연구. 한국형사정책연구원 연구총서.

*홍성열 (2007). 진술분석기법의 타당성연구. 대검찰청용역연구과제.

홍세희 (2013). 메타분석의 이론과 적용. 서울: 에스앤엠 리서치 그룹.

9) * 표가 있는 논문은 본 연구의 메타분석에 포함된 논문임.

- 황성동 (2014). 알기 쉬운 메타분석의 이해. 서울: 학지사.
- Amado, B. G., Arce, R., & Fariña, F. (2015). Undeutsch hypothesis and criteria-based content analysis: A meta-analytic review. *The European Journal of Psychology Applied to Legal Context*, 7(1), 3-12.
- Amado, B. G., Arce, R., Fariña, F., & Vilariño, M. (2016). Criteria-Based content analysis (CBCA) reality criteria in adults: A meta-analytic review. *International Journal of Clinical and Health Psychology*, 16(2), 201-210.
- Akehurst, L., Köhnken, G., & Höfer, E. (2001). Content credibility of accounts derived from live and video presentations. *Legal and Criminological Psychology*, 6, 65-83.
- Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. (2009). *An introduction to meta-analysis*. New York: John Wiley & Sons.
- Boyчук, T. (1991). *Criteria-Based Content Analysis of children's statements about sexual abuse: A field-based validation study*. Unpublished doctoral dissertation, Arizona State University.
- Esplin, P. W., Boyчук, T., & Raskin, D. C. (1988). A field validity study of Criteria-Based Content Analysis of children's statements in sexual abuse cases. Paper presented at the NATO Advanced Study Institute on Credibility Assessment, Maratea, Italy.
- Hauch, V., Sporer, S. L., Masip, J., & Blandón-Gitlin, I. (2017). Can credibility criteria be assessed reliably? A meta-analysis of Criteria-based Content Analysis. *Psychological Assessment*, 29(6), 819-834.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.
- Huitt, W., & Hummel, J. (2003). Piaget's theory of cognitive development. *Educational psychology interactive*, 3(2), 1-5.
- Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin*, 104, 53-69.
- Köhnken, G. (1996). Social psychology and the law. In G. R. Semin and K. Fiedler (eds.), *Applied social psychology* (pp. 257-82). London: Sage.
- Köhnken, G. (1999). Statement validity assessment. *preconference programme of applied courses "Assessing credibility", organised by the European Association of Psychology and Law, Dublin, Ireland*.
- Köhnken, G. (2004). Statement Validity Analysis and the 'detection of the truth'. In P. A. Granhag & L. A. Strömwall (Eds.), *Deception detection in forensic contexts* (pp. 41-63). Cambridge, England: Cambridge University Press.
- Köhnken, G., Schimossek, E., Aschermann, E., & Höfer, E. (1995). The cognitive interview and the assessment of the credibility of adults' statements. *Journal of Applied Psychology*, 80(6), 671.
- Merckelbach, H. (2004). Telling a good story: Fantasy proneness and the quality of fabricated memories. *Personality and Individual Differences*, 37, 1371-1382.
- Oberlader, V. A. (2019). *Meta-analyses on the validity of verbal tools for credibility assessment* (doctoral dissertation). Bonn, Germany: University of Bonn.
- Oberlader, V. A., Naefgen, C., Koppehele-Gossel,

- J., Quinten, L., Banse, R., & Schmidt, A. F. (2016). Validity of content-based techniques to distinguish true and fabricated statements: A meta-analysis. *Law and Human Behavior, 40*, 440-457.
- Oberlander, V. A., Quinten, L., Banse, R., Volbert, R., Schmidt, A. F., & Schönbrodt, F. D. (2021). Validity of content-based techniques for credibility assessment: How telling is an extended meta-analysis taking research bias into account?. *Applied Cognitive Psychology, 35*(2), 393-410.
- Porter, S., & Yuille, J. C. (1996). The language of deceit: An investigation of the verbal clues to deception in the interrogation context. *Law and Human Behavior, 20*, 443-459.
- Raskin, D. C., Esplin, F. W., & Horowitz, S. (1991). *Investigative interviews and assessment of children in sexual abuse cases* (Unpublished manuscript). University of Utah, Salt Lake City, UT.(as cited in Oberlander, 2005).
- Roma, P., San Martini, P., Sabatello, U., Tatarelli, R., & Ferracuti, S. (2011). Validity of criteria-based content analysis (CBCA) at trial in free-narrative interviews. *Child Abuse & Neglect, 35*, 613-620.
- Ruby, C. L., & Brigham, J. C. (1997). The usefulness of the Criteria-Based Content Analysis technique in distinguishing between truthful and fabricated allegations. *Psychology, Public Policy, and Law, 3*, 705-737.
- Santtila, P., Roppola, H., Runtti, M., & Niemi, P. (2000). Assessment of child witness statements using Criteria-Based Content Analysis (CBCA): The effects of age, verbal ability, and interviewer's emotional style. *Psychology, Crime and Law, 6*(3), 159-179.
- Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*(1481), 773-786.
- 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.
- Sporer, S. L., Manzanero, A. L., & Masip, J. (2021). Optimizing CBCA and RM research: recommendations for analyzing and reporting data on content cues to deception. *Psychology, Crime & Law, 27*(1), 1-39.
- Steller, M. (1989). Recent developments in statement analysis. In *Credibility assessment* (pp. 135-154). Springer, Dordrecht.
- Steller, M., & Kohnken, G. (1989). Statement analysis: Credibility assessment of children's testimonies in sexual abuse cases. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 217-245). New York: Springer.
- Trankell, A. (1972). *Reliability of evidence*. Stockholm, Sweden: Beckmans.
- Undeutsch, U. (1967). Beurteilung der glaubhaftigkeit von aussagen. *Handbuch der psychologie, 11*, 26-181.
- Undeutsch, U. (1989). The development of statement reality analysis. In *Credibility assessment* (pp. 101-119). Springer, Dordrecht.

- Vrij, A. (2005). Criteria-Based Content Analysis: A Qualitative Review of the First 37 Studies. *Psychology, Public Policy, and Law, 11*(1), 3.
- Vrij, A. (2008). *Detecting lies and deceit: Pitfall and opportunities*. John Wiley & Sons, LTD.
- Vrij, A. (2019). Deception and truth detection when analyzing nonverbal and verbal cues. *Applied Cognitive Psychology, 33*(2), 160-167.
- 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*(3), 261-283.
- Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2004). Detecting deceit via analyses of verbal and nonverbal behavior in children and adults. *Human Communication Research, 30*, 8-mentary: Is this child fabricating? Reactions to a new assessment technique. In J. Doris (Ed.), *The suggestibility of children's recollections: Implications for eyewitness testimony* (pp. 168-171). Washington, DC: American Psychological Association.
- 부산지방법원 동부지원 2019.12.10. 선고 2019
고합141, 2019전고8(병합) 판결.
- 인천지방법원 부천지원 2019.01.11. 선고 2018
고합143, 2018전고11(병합) 판결.
- 서울고등법원 2020.01.30. 선고 2019노630,
2019전노44(병합), 2019보노19(병합) 판결.
- 대법원 2020.05.14. 선고 2020도2433 판결.
- 1 차원고접수 : 2021. 06. 08.
심사통과접수 : 2021. 06. 23.
최종원고접수 : 2021. 07. 14.

Is it true?: A Meta-analysis on the Efficacy of CBCA in Detecting Truths

Hye Jin Kim Sangmin Lee Taekyun Hur Seung-Hyuk Choi

School of Psychology, Korea University

Statement Validity Analysis (SVA) is utilized in criminal investigations and the court to assess the credibility of given statements. During this procedure, the criteria for Criteria-Based Content Analysis (CBCA) are used to evaluate whether statements include the characteristics reflecting actual experiences about the event in question. Various studies had been conducted on the efficacy (classification rates) of CBCA criteria, yet the consistency of the findings was not investigated. In the current study, a meta-analysis was conducted with Korean CBCA studies reported from 2004 to 2020 (a total of fourteen studies). As a result, the total score of CBCA was found to successfully discriminate truth and fabrication. A significant positive (+) effect size was found with four criteria (3, 4, 10, and 12), all of which are classified as cognitive criteria. However, contrary to the underlying assumption for CBCA, criterion 18, classified as one of the motivational criteria, showed a significant negative (-) effect size. Meanwhile, moderator analyses were possible for eleven criteria (2~9, 12, 13, 15) and the results showed the significant effects of potential moderator variables such as the gender and status of the participants, study types and designs, number of raters, and publication status. The current results suggests that more careful attention is required to each criterion—especially the cognitive criteria—rather than the total CBCA score as well as the possible moderator effects in order to assess truthfulness of the statements. The implication, limitations, and suggestions for future studies were discussed.

Key words : CBCA, SVA, Statement Analysis, Moderators, Meta-Analysis

부록 1. 연구자료의 특성

일련 번호	연구	N	연령	성별	진술자 신분	연구 유형	연구 설계	사건유형	평가자	준거 수	출판유형
1	고은영, 채규만(2011)	39	13세 미만	혼합 (남:4, 여:35)	피해자	실험	집단간	성폭력	2인이상	18	학술지
2	김현정(2010)	44	만 8~13세	혼합 (남:2, 여:42)	피해자	실험	집단간	성폭력	2인이상	19	학술지
3	김현정(2009)	47	만 2~13세 미만	혼합 (-)	피해자	현장	집단간	성폭력	1인	19	비학술지 (박사학위)
4	윤대중(2019)	18	19~51세	혼합 (남:9, 여:9)	혼합 (가해자9, 피해자9)	현장	집단간	혼합	1인	18	비학술지 (박사학위)
5	윤여훈(2008)	98	초등생 (4~6학년)	혼합 (남:44, 여:54)	가해자	실험	집단내	일반사건	2인이상	18	비학술지 (석사학위)
6	윤여훈(2008)	91	초등생 (4~6학년)	혼합 (남:33, 여:58)	피해자	실험	집단내	일반사건	2인이상	18	비학술지 (석사학위)
7	이미선(2018)	71	만 3~6세	혼합 (남:37, 여:34)	-	실험	집단간	일반사건	2인이상	18	학술지
8	이수정(2010)	22	만 11~15세	여	피해자	실험	집단간	성폭력	1인	11	학술지
9	이수정, 유진옥(2010)	22	만 13~35세	여	피해자	실험	집단간	성폭력	2인이상	18	학술지
10	조은경(2004)	72	만 3~13세	혼합 (-)	피해자	현장	집단간	성폭력	2인이상	18	비학술지 (연구총서)
11	홍성열(2007)	189	초등생 (4~6학년)	혼합 (-)	혼합 (가해자98, 피해자91)	실험	집단내	일반사건	2인이상	18	비학술지 (연구보고서)
12	홍성열(2007)	166	대학생	혼합 (-)	혼합 (가해자60, 피해자106)	실험	집단내	일반사건	2인이상	18	비학술지 (연구보고서)
13	홍성열(2007)	158	초등생, (4~6학년), 대학생	혼합 (-)	가해자	실험	집단내	일반사건	2인이상	18	비학술지 (연구보고서)
14	홍성열(2007)	197	초등생, (4~6학년) 대학생	혼합 (-)	피해자	실험	집단내	일반사건	2인이상	18	비학술지 (연구보고서)

주. '-': 명확히 확인할 수 없음.