

폴리그래프 비교질문검사(comparison question technique: CQT) 기본 가정의 타당성에 대한 증거: 오류긍정 비율*

한 유 화

박 광 배†

충북대학교 심리학과

비교질문검사(comparison question technique: CQT)는 대한민국에서 가장 일반적으로 사용되고 있는 거짓말탐지검사 기법이다. CQT는 거짓말탐지검사에서 거짓말을 하는 피검사와 진실을 말하는 피검사자의 관련질문과 비교질문에 대한 심리생리적 반응이 다를 것이라는 기본적인 가정에 기초한다. 본 연구에서는 합리적인 추론을 바탕으로 하여 CQT의 기본적인 가정이 타당한지를 검증하였다. 연구방법으로는 CQT의 논리적 가정이 타당하지 않다고 주장하는 Lykken(1988, 1998)이 예측하는 CQT의 오류긍정 비율(기대빈도)과 실제 검사자료에서 산출되는 오류긍정 비율(관찰빈도)의 차이를 χ^2 검증하였다. 연구결과 CQT를 사용한 실제 검사자료의 오류긍정 비율은 1%~7%로 매우 낮게 나타난 것에 반해 Lykken(1988, 1998)이 주장하는 CQT의 오류긍정 비율은 39.5%~47%로 매우 높아 과대추정된 것으로 나타났으며 이것은 CQT의 기본적인 가정이 타당하다는 것을 보여주는 하나의 증거로 제시될 수 있을 것이다. 본 연구는 합리적 추론에 근거하여 CQT 기본적 가정의 타당성에 대한 증거를 제시하였다는 것에 의의가 있으며 논의에서는 본 연구와 더불어 CQT 기본적 가정의 타당성을 검증하기 위해 필수적으로 이루어져야 하는 연구들을 제안하였고, 본 연구의 제한점에 대해 논의하였다.

주요어 : 거짓말탐지검사, 비교질문검사(CQT), CQT의 논리적 타당성, 오류긍정 비율

* 이 논문은 2008년도 정부(과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구임
(M10740030003-07N4003-00310)

† 교신저자: 박광배, 충북대학교 사회과학대학 심리학과, 충청북도 청주시 흥덕구 개신동 성봉로 410
Tel: 043-261-2195, E-mail: kwangbai@chungbuk.ac.kr

범죄수사를 위한 거짓말탐지검사(polygraph test)의 검사기법 중 하나인 비교질문검사(comparison question technique: CQT)는 한 가지 중요한 가정에 기초한다. 그 가정은 거짓말을 하는(죄가 있는) 피검사자와 진실을 말하는(죄가 없는) 피검사자는 현안이 되고 있는 사건과 직접적으로 관련된 ‘관련질문(relevant question)’과 현안사건과 직접적인 관련은 없지만 그와 성질이 유사한 과거의 비행 혹은 범죄에 관한 ‘비교질문(comparison question)’에 대하여 서로 다른 태도를 가질 것이고, 서로 다른 심리상태와 생리반응으로 나타난다는 것이다. 구체적으로는, 거짓말을 하는 피검사자는 관련질문을 받았을 때 비교질문을 받았을 때보다 더 긴장할 것이기 때문에 생리반응 역시 관련 질문을 받았을 때 더 크게 나타날 것이고, 진실을 말하는 피검사자는 관련질문보다 비교질문에 대해서 더 긴장할 것이므로 비교질문을 받았을 때 더 큰 생리반응이 나타날 것이라는 가정이다. 본 연구에서는 합리적 추론을 바탕으로 실제 범죄용의자들에 대한 검사자료를 이용하여 CQT의 이러한 가정(논리)이 타당한지를 검증하였다.

CQT의 논리적 기초

CQT는 거짓말탐지를 위하여 소위 ‘관련-무관 검사(Relevant-Irrelevant Technique: RIT)’의 단점을 보완할 목적으로 개발되었다. RIT는 현안사건에 관한 관련질문(relevant question)과 현안사건과 무관한 질문(irrelevant question)을 연속적으로 제시하여 관련질문과 무관질문에 대한 피검사자의 생리반응을 비교하는 방식으로 채점한 후 피검사자의 거짓말 여부를 판정한다. RIT는 거짓말을 하는 피검사자는 관련질

문을 받았을 때 무관질문을 받았을 때보다 생리적 반응이 더 크게 나타날 것이고, 진실을 말하는 피검사자는 두 종류의 질문에 대하여 동일한 생리적 반응을 보일 것이라는 가정에 기초하였다. 그러나 현안사건과 관련이 없고 진실로 무고하지만 그럼에도 불구하고 범죄혐의를 받아서 추궁당하는 와중에 있는 용의자에게도 관련 질문과 무관질문은 심리적 의미가 매우 다르기 때문에 실제로 진실을 말하는 피검사자라고 할지라도 무관질문에 대한 생리반응보다 관련 질문에 대한 생리반응이 크게 나타날 수밖에 없고 따라서 RIT는 필연적으로 오류공정 비율(false positive rate: 진실을 말하는 피검사자를 거짓 판정하는 비율)이 매우 높게 나타난다(Horowitz, Kircher, Honts, & Raskin, 1997; Horvath, 1988).

Reid(1947)는 RIT의 이러한 문제점을 극복하기 위해 CQT를 개발하였고, CQT는 현재 대한민국의 범죄 수사 장면에서 가장 일반적으로 사용되고 있는 검사기법이다. CQT는 현안사건에 대해 직접적 지식 또는 경험을 가지고 있는지에 대한 피검사자 진술의 진실성을 평가한다(Raskin, 1986). CQT에는 크게 세 가지 유형의 질문이 포함된다. 첫 번째는 관련질문으로 이 질문은 조사 중인 현안사건에 대한 직접적인 질문(예를 들면, “당신이 00에서 돈을 훔쳤습니까?”)이다. 관련질문에 대하여 피검사자는 현안사건에 가담하거나 연루되지 않았다는 의미로 ‘아니오’라는 대답을 할 것이 기대된다. 두 번째는 비교질문으로 이 질문은 현안사건과 성질이 유사하지만 직접적인 관련성은 없는 일에 대한 질문(예를 들면, “당신은 0000년 이전에 다른 사람의 물건을 훔친 적이 있습니까?”)이다. 비교질문에 대해서도 피검사자는 ‘아니오’라는 대답을 할 것이 기대되는데

사전면담에서 검사관은 이 질문에 대해 부정하는 것이 거짓말탐지검사에서 ‘진실’ 판정을 받는데 유리하다는 것을 피검사자에게 강조함으로써 ‘아니오’라는 대답을 유도한다. 마지막 질문 유형은 무관질문으로 이 질문은 현안사건과 아무 관련이 없는 중성적인 질문 (예를 들면, “당신의 거주지는 서울입니까?”)이고 이 질문에 대해 피검사자는 ‘예’라는 대답을 할 것이 기대된다.

CQT에서는 피검사자의 반응을 평가하기 위하여 관련질문과 비교질문에 대한 생리적 반응을 비교한다. 피검사자가 현안사건과의 관련성에 대해 거짓말을 한다면 사건과 직접적인 관련이 있는 관련질문에 대한 생리적 반응이 비교질문에 대한 생리적 반응보다 클 것이다. 반면에 피검사자가 진실을 말한다면 비교질문에 대한 생리적 반응이 관련질문에 대한 생리적 반응보다 크거나 같을 것이다. CQT의 채점에서 각 관련질문에 대한 반응에는 -3점 ~ +3점까지의 점수를 부여할 수 있다. 관련질문에 대한 생리반응이 비교질문에 대한 생리반응보다 크다면 ‘-’점수, 작다면 ‘+’점수를 부여하는데, 관련질문과 비교질문에 대한 생리반응의 차이가 매우 크면 3점, 작으면 1점, 거의 없으면 0점을 부여한다.

CQT의 비교질문은 현안과 직접적인 관련은 없지만 현안이 되는 사건과 내용이 유사하고 무죄의 용의자일지라도 한번은 경험했을 법한 내용에 대한 질문들이므로 무죄의 용의자들이 관련질문보다 더 관심을 가지게 되고, 따라서 그들의 심리상태에 더 중요하게 작용한다고 생각되는 질문들이다. 다시 말하면 무죄의 용의자들이 비교질문에 대해 더 긴장하고 불안한 마음을 가지도록 조작하여, 만약에 용의자가 진정으로 무고한 사람이라면 관련질문보다

비교질문에 더 강하게 반응하도록 하기 위해 고안된 질문이다.

무고한 용의자가 비교질문에서 강하게 반응하도록 하는 조건화의 효과는 사전면담(pretest interview) 과정에서 극대화된다. 사전면담은 피검사자가 검사를 받는 이유에 대한 설명, 검사 동의서 작성, 심리 및 생리변화와 폴리그래프에 대한 설명, 사건내용에 대한 확인, 검사에 사용될 질문에 대한 검토를 하는 순서로 진행된다. 이러한 기본적인 면담과 더불어 이 과정의 가장 중요한 역할은 피검사자의 심리상태를 피검사자의 유/무죄 상태에 따라서 조건화하는 효과를 극대화 하는 것이다. 사전면담에서 실제로 유죄이고 거짓말하는 피검사자는 관련질문에 대해 더 걱정하고 불안해하도록, 실제로 무죄이고 진실을 말하는 피검사자는 비교질문에 대해 더 걱정하고 불안해하도록 하는 조건화의 효과가 극대화 되도록 한다. 사전면담에서 모든 피검사자는 비교질문을 받았을 때 그러한 과거의 행동이나 경험이 없다는 의미의 대답(아니오)을 하는 것이 CQT에서 최종적으로 ‘거짓’ 판정을 받지 않는데 유리하다는 설명을 듣는다. 따라서 피검사자가 무죄이고, 진실을 말한다면 피검사자는 ‘거짓’ 판정을 받지 않기 위해 최대한의 노력을 기울여야 하므로 비교질문에 더 집중하게 된다. 그러나 피검사자가 유죄이고, 거짓말탐지검사에서 거짓말을 해야 한다면 피검사자는 사건과 직접적으로 관련된 관련질문에서 거짓말을 함으로써 자신의 행위를 부정해야 하므로 관련질문에 더 집중하게 된다. Offe와 Offe(2007)는 사전면담과정에서 비교질문에 대한 설명을 하는 것이 거짓말탐지검사의 정확성을 높인다는 것을 검증하였는데 이러한 결과는 사전면담에서 피검사자의 심리상태를 조건화하는 것

이 효과가 있다는 것을 과학적으로 증명한 것이다.

CQT 논리의 타당성

CQT는 피검사자가 거짓말하는 사람인지, 진실을 말하는 사람인지에 따라서 가장 위협적인 질문이 각기 다르다는 가정에 기초하고 있다. Ben-Shakhar와 Furedy(1990)는 이러한 CQT의 논리가 타당하지 않다고 주장하였다. 이들은 실제로 거짓말탐지검사를 받는 모든 용의자들에게 관련질문은 매우 중요하기 때문에 이러한 CQT의 논리가 성립되지 않고, 조사 중인 범죄에 초점을 맞춘 관련질문이 보다 일반적인 비교질문과 동일한 수준의 관심과 걱정의 대상이 될 것이라는 기대는 할 수 없으며, 이것은 무죄의 용의자나 유죄의 용의자에게 모두 마찬가지로 주장하였다.

Horowitz 등(1997)은 CQT의 논리가 타당한지를 검증하였다. 이들은 모의범죄 실험에 참여한 참가자들의 유/무죄(거짓/진실)를 판단한 후, 참가자들에게 검사에 사용되었던 각 질문들이 검사 결과에 대해서 갖는 중요성 정도를 평가하도록 하였다. 그 결과 유죄인 피검사자와 무죄인 피검사자 모두 관련질문이 비교질문보다 더 중요하다고 판단하였고, 비교질문이 무관질문보다 더 중요하다고 판단하였다. 그러나 거짓말탐지검사가 진행되는 동안에 각 질문에 대해 자신이 정서적으로 얼마나 더 강하게 반응했었는지를 물었을 때는 무죄인 피검사자는 비교질문, 유죄인 피검사자는 관련질문에 대해서 강하게 반응했다고 보고하였다. 이것은 CQT의 논리가 타당하다는 것을 의미하며 실제 생리적 반응도 이와 일치하는 결과를 나타냈다. 또한 Horvath(1988)는 모의범죄

연구를 실시하여 각 피검사자들이 각 질문에 대해 걱정되고 불안한 정도를 평가하도록 하였다. 그 결과 무죄의 피검사자들은 관련질문보다 비교질문에 대해 더 불안해하는 반면, 유죄의 피검사자들은 비교질문보다 관련질문에 대해 더 불안해하고 걱정하는 것으로 나타났다. 더 나아가서, 각 질문에 대해 불안해하고 걱정하는 정도가 실제 생리적 반응과 정적인 상관관계를 보인다고 보고하였다. 이와 유사하게 Offe와 Offe(2007)가 피검사자들에게 거짓말탐지 검사를 실시한 후 각 질문에 대한 스트레스 정도를 평가하도록 한 연구에서 피검사자가 거짓말을 한 경우에는 관련질문에서, 피검사자가 진실을 말한 경우에는 비교질문에서 스트레스 정도가 더 높다는 것을 확인하였다.

Elaad(2003)는 CQT의 논리가 타당한지를 검증하기 위해 두 가지 실험을 설계하였다. 실험 1에서는 수사관들에게 가상적인 절도사건에 대한 유죄인 용의자와 무죄인 용의자의 시나리오를 각각 들려주고, 용의자들이 거짓말탐지검사를 받게 된다면 검사에 사용되는 각 질문들에 대해 어느 정도의 걱정을 하고 관심을 가질지를 용의자의 입장에서 평가하도록 하였다. 그 결과 유죄인 용의자의 시나리오를 들은 피험자들은 모든 다른 유형의 질문들보다 관련질문에 대해 더 많이 걱정하고, 관심을 가질 것이라고 평가하였고 무죄인 용의자의 시나리오를 들은 피험자들은 비교질문에 대해 가장 많이 걱정하고, 관심을 가질 것이라고 평가하였다.

Elaad(2003)의 실험 2는 CQT의 논리가 타당하지 않다면 (1) 적어도 많은 무죄의 피검사자들이 비교질문에서보다 관련질문에서 더 큰 생리적 반응을 보이고, (2) Lykken(1988)이 주장

하듯이 오류긍정 비율이 우연 수준인 50%에 가까울 것이라는 추론에 근거하여 실시되었다. Lykken(1988)의 연구에 따르면 CQT에서 무죄의 용의자가 거짓 판정을 받을 비율이 47%에 이르며 유죄의 용의자가 거짓판정을 받을 비율은 평균 83%이다. 또 Lykken(1998)은 4개의 현장 연구들을 요약하여 CQT에서 유죄인 용의자가 거짓판정을 받을 비율은 평균 85.5%이지만 무죄인 용의자가 거짓판정을 받을 비율도 39.5%로 매우 높다고 보고한 바 있다. Elaad(2003)의 실험 2에서는 Lykken(1988, 1998)의 연구에서 산출된 CQT의 정확성과 오류긍정 비율을 사용하여 기대빈도를 산출하고, 무선적으로 선택한 실제 검사자료를 전문 검사관들에게 다시 채점 의뢰하여 얻은 각 판정의 관찰빈도가 Lykken(1988, 1998)의 연구결과에 의해 추정된 기대빈도와 부합하는지를 검증하였다. 연구 결과 Lykken(1988, 1998)이 추정한 CQT의 오류긍정 비율이 과대추정 되었으며, CQT의 실제 오류긍정 비율은 50% 보다 현저히 낮은 것으로 나타났다. Elaad(2003)는 이러한 결과를 바탕으로 CQT의 논리적 가정이 타당하다고 주장하였다.

Elaad(2003)가 사용한 방법은 실제 거짓말하는 피검사자와 진실을 말하는 피검사자의 특정한 비율을 가정하고 Lykken(1988, 1998)이 주장하는 CQT의 정확성과 오류긍정 비율을 사용하여 관련된 기대빈도들을 산출한 후 실제 검사자료의 판정결과로 산출된 관찰빈도들이 기대빈도들과 유의미한 차이가 나타나는지를 χ^2 분석을 사용하여 검증하는 것이었다. 구체적으로 Elaad(2003)는 실제 거짓말하는 피검사자의 비율이 최소한 30%는 될 것이라고 가정하고 Lykken(1988)이 예측하는 기대빈도를 산출하였다. 그 결과 거짓판정에 대한 기대빈도

는 31.4, 관찰빈도는 19, 진실판정에 대한 기대빈도는 22.6, 관찰빈도는 35였으며 이때의 χ^2 값은 4.83으로 나타나 유의미하였다. 그러나 Elaad(2003)의 연구 결과가 Lykken(1988, 1998)이 주장하는 CQT의 오류긍정 비율이 과대추정 되었다는 주장을 반증한다고 하기에는 연구 결과의 해석과 관련하여 모호한 점이 있다.

Elaad(2003)는 실제 거짓말하는 피검사자와 진실을 말하는 피검사자의 특정한 비율(30% v. 70%)을 가정하고 기대빈도를 산출하였다. 이 경우 만약 관찰빈도와 기대빈도가 유의미한 차이를 보인다면 그것은 두 가지 이유에 의해 가능할 수 있다. 하나는 실제 거짓말하는 피검사자와 진실을 말하는 피검사자의 비율이 Elaad(2003)가 가정한 것과 다르기 때문일 수 있고, 다른 하나는 CQT의 오류긍정 비율이 Lykken(1988, 1998)이 주장한 것과 다르기 때문일 수 있다. 따라서 Elaad(2003)가 Lykken(1988, 1998)의 기대빈도와 실제 자료가 부합하지 않는다는 결과(유의미한 χ^2 값)를 얻었더라도 그것은 실제 거짓말하는 피검사자와 진실을 말하는 피검사자의 비율이 가정한 것과 다르기 때문일 가능성과 Lykken(1988, 1998)이 주장하는 CQT의 오류긍정 비율이 과대추정 되었기 때문일 가능성이 모두 존재한다. 그럼에도 불구하고 Elaad(2003)는 첫 번째 가능성을 무시하고 그의 연구 결과가 두 번째 가능성 때문에 산출된 것으로 해석하고 있다.

Elaad(2003)는 거짓말탐지검사의 실제 검사자료를 분석하는 현장연구에서는 피검사자들의 실제 거짓말여부(ground truth)를 확인할 수 없기 때문에 실제 거짓말하는 피검사자와 실제로 진실을 말하는 피검사자의 비율이 특정하다고 가정하고 분석을 실시하였다. Elaad(2003)는 가장 합리적인 비율이라는 이유로 이러한

비율(최소 30%의 피검사자가 거짓판정)을 가정 하였으나 그 비율이 사실을 반영하거나 실제로도 그렇다는 근거가 없었다. 따라서 본 연구에서는 실제 자료의 실제 거짓말여부를 추정하여 Elaad(2003)의 연구 결과를 해석하기 위해 있을 수 있는 두 가지 가능성 중 첫 번째 가능성을 제거한 후, Lykken(1988, 1998)의 오류긍정 비율이 과대추정 되었다는 것을 검증하고자 한다. 각 피검사자들의 실제 거짓말여부는 한유화와 박광배(2008, p.108)가 제시한 잠재계층분석(Latent Class Analysis) 결과를 사용하여 추정하고자 한다. 한유화와 박광배(2008)는 2006년 실시된 거짓말탐지검사의 실제 자료 중 182건의 사례를 사용하여 실제로 거짓말하는 계층과 진실을 말하는 계층이 존재한다고 가정하는 2계층 모형을 검증 하였다. 본 연구에서 사용한 한유화와 박광배(2008)의 잠재계층분석 결과는 기소여부와 거짓말탐지검사의 최종판정 패턴이 ‘기소-거짓’, ‘불기소-거짓’, ‘기소중지-거짓’, ‘기소-진실’, ‘기소-판단불능’, ‘기소중지-판단불능’인 경우에는 피검사자의 실제 거짓말여부를 ‘거짓’으로 추정하고, ‘불기소-진실’, ‘기소중지-진실’, ‘불기소-판단불능’인 경우에는 ‘진실’로 추정한다는 것이다. 본 연구에서는 Lykken(1988, 1998)의 논리(CQT의 기본가정이 타당하지 않다면 오류긍정 비율은 우연 수준인 50%에 가까울 것이다)를 적용하여 Elaad(2003)의 연구 결과를 해석을 모호하게 하는 가능성들을 제거하고 Elaad(2003)의 연구 결과가 CQT의 논리가 타당하다는 것을 지지한다는 것을 확인하고, 동일한 연구 방법을 사용하여 CQT의 논리적 타당성을 검증하고자 한다. 다시 말하면, Lykken(1988, 1998)의 오류긍정 비율이 과대추정 되었다는 것을 검증하는 것은 CQT를 사용했을

경우의 오류긍정 비율이 우연수준인 50%보다 낮다는 것을 검증하는 것이고, CQT를 사용하여 거짓말탐지검사를 실시했을 경우 오류긍정 비율이 낮게 나타나기 위해서는 CQT의 기본적인 논리가 타당해야 하므로 본 연구에서는 이러한 추론 과정을 통해 CQT의 논리적 타당성을 검증하고자 한다.

방 법

분석자료

2006년에 대검찰청과 4개의 지방검찰청에서 형사사건에 대해 실시된 거짓말탐지검사 자료 중 거짓말탐지검사의 판정결과(거짓/진실/판단불능)와 검찰의 기소여부(기소/불기소/기소중지)가 알려져 있는 208건의 사례를 선택하여 사용하였다. 이 사례들은 흥부호흡, 복부호흡, 혈압맥박, 피부전도수준을 측정하는 거짓말탐지검사 장비를 사용하여 전문 검사관에 의해 검사된 사례들이었다. 각 사례들의 거짓말탐지검사에 사용된 검사기법은 CQT 중 하나로 Backster(1963)에 의해 개발된 Backster zone comparison test였다. 208건 사례들에서 피검사자의 평균연령은 44.5세(표준편차: 11.92세)였고 17세~82세까지 다양한 연령의 피검사자들이 포함되어 있었다. 피검사자들 중 남성은 147명(70.7%), 여성은 61명(29.3%)이었고, 피검사자의 신분은 피의자 138명(66.3%), 고소인 42명(20.2%), 피해자 14명(6.7%) 등이었다. 피검사자들의 거짓말탐지검사 당시 최종판정은 ‘거짓’인 사례가 95건(45.7%), ‘진실’인 사례가 69건(33.2%), ‘판단불능’인 사례가 44건(21.2%) 포함되어 있었다. 또한 피의자가 기소된 사례

가 98건(47.1%), 불기소된 사례가 97건(46.6%), 기소중지된 사례가 13건(6.3%)이었다. 피검사자들의 거짓말탐지검사 총점은 -23점~10점 사이에 분포하였고, 총점평균은 -8.50점(표준편차: 7.64)이었다.

본 연구에서는 208건의 사례를 사용하여 두 번의 분석을 실시하였다. 첫 번째는 현재 대한민국의 검찰에서 사용하고 있는 판정기준 점수(-12점)를 사용하여 각 피검사자들에 대한 거짓말탐지검사의 판정을 내린 후 각 판정에 대한 관찰빈도를 산출하여 Lykken(1988, 1998)이 예측하는 기대빈도와의 차이를 검증하는 것이었고, 두 번째는 한유화와 박광배(2008)가 제시한 판정기준 점수(-8점)를 사용하여 각 피검사자들에 대한 거짓말탐지검사의 판정을 내린 후 각 판정에 대한 관찰빈도를 산출하여 Lykken(1988, 1998)이 예측하는 기대빈도와의 차이를 검증하는 것이었다.

분석

본 연구에서는 208건의 사례에 대해 각 사례의 실제 거짓말여부를 추정된 후, 추정된 실제 거짓말여부와 거짓말탐지검사 결과의 이원교차분할표에서 각 셀에 해당하는 관찰빈도와 기대빈도를 산출하였다. 첫 번째 분석에서는 현재 대한민국의 검찰에서 사용하고 있는 판정기준 점수(-12점)를 사용하여 각 피검사자들에 대한 거짓말탐지검사의 판정을 내린 후 각 판정에 대한 관찰빈도를 산출하였다. 두 번째 분석에서는 한유화와 박광배(2008)가 제시한 판정기준 점수(-8점)를 사용하여 각 피검사자들에 대한 거짓말탐지검사의 판정을 내린 후 각 판정에 대한 관찰빈도를 산출하였다. 현재 대한민국의 검찰에서 사용하는 판정기준

의 과학적인 설정을 위해 이루어진 한유화와 박광배(2008)의 연구에 따르면 -12점을 사용하여 거짓말탐지검사의 최종판정을 내리는 것보다 -8점을 사용하여 최종판정을 내리는 것이 거짓말탐지검사의 정확성을 높이는 것으로 나타났다. -8점을 사용하여 최종판정을 내리게 되면 -12점을 사용하여 최종판정을 내리는 것보다 필연적으로 오류긍정 비율이 높게 나타난다. 따라서 본 연구에서는 -8점을 최종판정을 위한 판정기준 점수로 사용하여 각 판정에 대한 관찰빈도를 산출하였을 경우에도 Lykken(1988, 1998)이 예측하는 것보다 낮은 오류긍정 비율이 나타나는지 확인하기 위하여 두 번의 분석을 실시하였다. 기대빈도는 이원교차분할표의 외곽비율들에 의해 산출될 수 있는데, 본 연구에서는 잠재계층분석을 통해 추정된 실제 거짓말여부에 따른 2개의 외곽비율과 Lykken(1988, 1998)이 제시한 거짓말탐지검사의 정확성(84%, 85.5%)과 오류긍정 비율(47%, 39.5%)을 사용하여 산출하였다. 이렇게 산출한 기대빈도는 Lykken(1988, 1998)이 예측하는 실제 거짓말여부와 거짓말탐지검사 결과의 이원교차분할표에 포함된 각 셀의 기대빈도가 된다. 실제 거짓말여부와 거짓말탐지검사 결과의 이원교차분할표의 예시를 표 1에 제시하였다.

표 1. 실제 거짓말여부와 거짓말탐지검사 결과의 이원교차분할표의 예시

		거짓말탐지검사 결과(i)		전체
		거짓(1)	진실(2)	
실제 거짓말 여부(i)	거짓(1)	n_{11}	n_{12}	$n_{1.}$
	진실(2)	n_{21}	n_{22}	$n_{2.}$
전체		$n_{.1}$	$n_{.2}$	n_{ij}

표 1에서 각 셀의 관찰빈도와 Lykken(1988, 1998)이 예측하는 기대빈도는 다음과 같이 산출할 수 있다. 각 셀의 관찰빈도를 n , 기대빈도를 μ 라고 하면, n_1 과 n_2 는 잠재계층분석 결과로 추정된 실제 거짓말여부가 거짓인 빈도와 진실인 빈도이다. 또한 n_1 과 n_2 는 각 분석 자료에 대해 특정한 판정기준 점수를 적용하여 판단한 거짓말탐지검사의 거짓, 진실 판정 빈도이다. 관찰빈도 n_{11} 은 실제 거짓말여부가 거짓이고, 거짓말탐지검사 결과 거짓 판정받은 피검사자의 수, n_{12} 는 실제 거짓말여부가 거짓이고, 거짓말탐지검사 결과 진실 판정받은 피검사자의 수, n_{21} 은 실제 거짓말여부가 진실이고, 거짓말탐지검사 결과 거짓 판정받은 피검사자의 수, n_{22} 는 실제 거짓말여부가 진실이고, 거짓말탐지검사 결과 진실 판정받은 피검사자의 수로 각각 정의된다. 본 연구에서는 기대빈도를 산출하기 위해 필요한 실제로 거짓말하는 사람이 거짓말탐지검사에서 거짓 판정받을 확률(정확성)과 실제로 진실을 말하는 사람이 거짓 판정받을 확률(오류긍정 비율)이 Lykken(1988, 1998)에 의해 알려져 있으므로 두 가지 비율과 실제 거짓말한 사람의 빈도(n_1)와 실제 진실을 말한 사람의 빈도(n_2)를 사용하여 각 셀의 기대빈도를 산출할 수 있다. CQT를 사용한 거짓말탐지검사의 정확성을 p_{11} , 오류긍정 비율을 p_{21} 이라고 하면 각 셀의 기대빈도는 다음과 같이 정의된다.

$$\mu_{11} = n_1 \times p_{11} \quad \text{수식(1)}$$

$$\mu_{12} = n_1 \times (1 - p_{11}) \quad \text{수식(2)}$$

$$\mu_{21} = n_2 \times p_{21} \quad \text{수식(3)}$$

$$\mu_{22} = n_2 \times (1 - p_{21}) \quad \text{수식(4)}$$

본 연구에서는 위와 같은 수식에 의해 산출된 기대빈도와 관찰빈도의 합치여부를 판단하기 위해 χ^2 검증을 실시하였다.

결 과

분석 자료의 실제 거짓말여부 추정

208건의 전체 분석자료에 포함되어 있는 검찰의 기소여부와 거짓말탐지검사의 최종판정을 표 2에 제시한 관찰변인의 범주패턴에 따라 분류하여 각 피검사자의 실제 거짓말여부를 추정하였다. 그 결과 실제 거짓말여부는 208건의 사례 중 127건(61.1%)의 사례가 거짓, 81건(38.9%)의 사례가 진실로 추정되었다. 실제 거짓말여부가 거짓으로 추정된 사례들 중 기소여부와 거짓말탐지검사 최종판정의 패턴이 ‘기소-거짓’인 경우는 68건(53.5%), ‘불기소-거짓’인 경우는 19건(15.0%), ‘기소중지-거짓’인 경우는 8건(6.3%), ‘기소-진실’인 경우는 12건(9.4%), ‘기소-판단불능’인 경우는 18건(14.2%), ‘기소중지-판단불능’인 경우는 2건(1.6%)로 나타났다. 또한 실제 거짓말 여부가 진실로 추정된 사례들 중 기소여부와 거짓말탐지검사 최종판정의 패턴이 ‘불기소-진실’인 경우가 54건(66.7%), ‘기소중지-진실’인 경우가 3건(3.7%), ‘불기소-판단불능’인 경우가 24건(29.6%)으로 나타났다.

χ^2 검증

첫 번째 분석에서 실제 거짓말여부와 거짓

표 2. 관찰변인의 범주패턴에 따른 사후확률과 잠재계층할당

관찰변인의 범주패턴		n	사후확률		잠재계층 할당
기소여부	거짓말탐지검사 최종판정		거짓	진실	
기소	거짓	60	1.00	0.00	거짓
불기소	거짓	11	0.98	0.02	거짓
기소중지	거짓	8	1.00	0.00	거짓
기소	진실	11	0.98	0.02	거짓
불기소	진실	50	0.04	0.96	진실
기소중지	진실	3	0.48	0.52	진실
기소	판단불능	18	1.00	0.00	거짓
불기소	판단불능	19	0.21	0.79	진실
기소중지	판단불능	2	0.86	0.14	거짓

한유화, 박광배(2008). 범죄수사를 위한 거짓말탐지 검사(polygraph test)의 판정기준과 정확성. 한국심리학회지: 사회문제, 14(4), p.108.에서 인용

말탐지검사 결과(-12점을 판정기준 점수로 사용)의 이원교차분할표에 포함되는 각 셀의 관찰빈도를 산출한 결과 실제 거짓말여부가 거짓인 피검사자는 127명(61.1%), 실제 거짓말여부가 진실인 피검사자는 81명(38.9%)으로 나타났다. 또한 208명의 피검사자 중 실제 거짓말여부가 거짓이면서 거짓말탐지검사 결과 거짓 판정을 받은 피검사자는 90명(43.3%), 진실 판정을 받은 피검사자는 37명(17.8%), 실제 거짓말여부가 진실이면서 거짓말탐지검사 결과 거짓 판정을 받은 피검사자는 1명(0.4%), 진실 판정을 받은 피검사자는 80명(38.5%)으로 나타났다. 수식(1) ~ 수식(4)를 사용하여 Lykken (1988)이 예측하는 각 셀의 기대빈도를 산출한 결과 거짓판정을 받을 피검사자는 144.75명(69.6%), 진실 판정을 받을 피검사자는 63.25명(30.4%)으로 나타났다. 또한 실제 거짓말여부가 거짓이면서 거짓말탐지검사 결과 거짓 판

정을 받을 피검사자는 106.68명(51.3%), 진실 판정을 받을 피검사자는 20.32명(9.8%), 실제 거짓말여부가 진실이면서 거짓말탐지검사 결과 거짓 판정을 받을 피검사자는 38.07명(18.3%), 진실 판정을 받을 피검사자는 42.93명(20.6%)으로 나타났다. Lykken(1998)이 예측하는 각 셀의 기대빈도는 거짓판정을 받을 피검사자가 140.58명(67.6%), 진실판정을 받을 피검사자가 67.42명(32.4%)으로 나타났다. 또한 실제 거짓말여부가 거짓이면서 거짓말탐지검사 결과 거짓 판정을 받을 피검사자는 108.59명(52.1%), 진실 판정을 받을 피검사자는 18.41명(8.9%), 실제 거짓말여부가 진실이면서 거짓말탐지검사 결과 거짓 판정을 받을 피검사자는 31.99명(15.4%), 진실 판정을 받을 피검사자는 49.01명(23.6%)으로 나타났다. 이렇게 산출된 관찰빈도와 기대빈도가 합치하는지 알아보기 위해 χ^2 검증을 실시한 결과를 표 3에 제시하

표 3 추정된 유/무죄 상태와 거짓말탐지검사의 판정에 따른 각 셀의 기대빈도, 관찰빈도(-12점을 판정기준 점수로 사용)와 χ^2

피검사자의 유/무죄상태		Lykken(1988)		χ^2	Lykken(1998)		χ^2
		거짓	진실		거짓	진실	
유죄 (거짓)	기대빈도	106.68	20.32	84.41***	108.59	18.41	71.57***
	관찰빈도 ¹⁾	90	37		90	37	
무죄 (진실)	기대빈도	38.07	42.93		31.99	49.01	
	관찰빈도 ¹⁾	1	80		1	80	

1) 현재 대한민국의 검찰에서 사용하고 있는 판정기준(-12점)에 의한 판정 빈도

*** $p < .001$

였다.

표 3에서 보는 바와 같이 -12점을 판정기준 점수로 사용하였을 경우 Lykken(1988, 1998)에 의해 예측되는 거짓말탐지검사의 각 판정에 대한 기대빈도는 대한민국의 검찰에서 이루어진 거짓말탐지검사의 각 판정에 대한 관찰빈도와 유의미한 차이가 있는 것으로 나타났다 ($\chi^2=84.41$, $df=1$, $p < .001$; $\chi^2=71.57$ $df=1$, $p < .001$).

두 번째 분석에서 실제 거짓말여부와 거짓말탐지검사 결과(-8점을 판정기준 점수로 사

용)의 이원교차분할표에 포함되는 각 셀의 관찰빈도를 산출한 결과 실제 거짓말여부가 거짓인 피검사자는 127명(61.1%), 실제 거짓말여부가 진실인 피검사자는 81명(38.9%)으로 나타났다. 또한 208명의 피검사자 중 실제 거짓말여부가 거짓이면서 거짓말탐지검사 결과 거짓 판정을 받은 피검사자는 98명(47.1%), 진실 판정을 받은 피검사자는 29명(14.0%), 실제 거짓말여부가 진실이면서 거짓말탐지검사 결과 거짓 판정을 받은 피검사자는 5명(2.4%), 진실 판정을 받은 피검사자는 76명(36.5%)으로 나타

표 4 추정된 유/무죄 상태와 거짓말탐지검사의 판정에 따른 각 셀의 기대빈도, 관찰빈도(-8점을 판정기준 점수로 사용)와 χ^2

피검사자의 유/무죄상태		Lykken(1988)		χ^2	Lykken(1998)		χ^2
		거짓	진실		거짓	진실	
유죄 (거짓)	기대빈도	106.68	20.32	58.62***	108.59	18.41	44.76***
	관찰빈도 ¹⁾	98	29		98	29	
무죄 (진실)	기대빈도	38.07	42.93		31.99	49.01	
	관찰빈도 ¹⁾	5	76		5	76	

1) 한유화, 박광배(2008)의 연구에서 산출한 판정기준 점수(-8점)에 의한 판정 빈도

*** $p < .001$

났다. 수식(1) ~ 수식(4)를 사용하여 Lykken(1988, 1998)이 예측하는 각 셀의 기대빈도를 산출한 결과는 표 3에서 제시한 기대빈도와 동일하였다. 이렇게 산출된 관찰빈도와 기대빈도가 합치하는지 알아보기 위해 χ^2 검증을 실시한 결과를 표 4에 제시하였다.

표 4에서 보는 바와 같이 -8점을 판정기준 점수로 사용하였을 경우 Lykken(1988, 1998)이 예측하는 거짓말탐지검사의 각 판정에 대한 기대빈도와 관찰빈도의 차이는 유의미한 것으로 나타났다($\chi^2=58.62$, $df=1$, $p < .001$; $\chi^2=44.76$ $df=1$, $p < .001$).

논 의

본 연구에서는 두 번의 분석을 실시하였는데, 두 가지의 판정기준 점수(-12점과 -8점)를 사용하여 CQT를 사용한 거짓말탐지검사의 판정을 내렸을 때 각각의 관찰빈도가 Lykken(1988, 1998)이 예측하는 기대빈도와 부합하는지 판단함으로써 Lykken(1988, 1998)이 CQT의 오류공정 비율을 과대추정하고 있다는 것을 증명하였다. Lykken(1988, 1998)은 실제로 진실을 말하는 피검사자 중 47% 또는 39.5%가 CQT에 의해서 거짓으로 판정받을 것이라고 예측하였지만, 연구 결과 -12점의 판정기준 점수 사용했을 경우에 실제로 진실을 말하는 피검사자 중 거짓 판정을 받은 피검사자는 1.23%에 불과하였고 -8점의 판정기준 점수 사용했을 경우에는 6.17%만이 거짓 판정을 받은 것으로 나타났다. 따라서 Lykken(1988, 1998)은 CQT의 오류공정 비율을 매우 과대추정하고 있다고 할 수 있다. Lykken(1988, 1998)은 CQT가 논리적으로 타당하지 않기 때문에 실제로

진실을 말한 사람을 CQT에서 거짓으로 판정하는 비율(오류공정 비율)이 우연수준인 50%에 가깝다고 주장하였다. 그러나 본 연구에서 CQT의 오류공정 비율은 1~7%에 지나지 않아 Lykken(1988, 1998)이 주장하는 오류공정 비율은 과대추정되었다는 것을 증명하였다. 오류공정 비율이 과대추정되었다는 것은 다시 말하면 CQT에서 실제로 진실을 말하는 피검사자는 거짓 판정을 받을 확률이 매우 낮으며(1~7%), CQT가 논리적으로 타당하기 때문에 이러한 결과가 나타날 수 있다는 것을 의미한다.

본 연구에서는 Elaad(2003)의 연구에서 나타난 결과 해석의 모호함을 제거하기 위하여 가정 중 하나였던 실제 거짓말여부를 추정하여 실제로 거짓말하는 피검사자와 실제로 진실을 말하는 피검사자의 비율에 대한 근거를 제시하였다. 실제 거짓말여부를 추정하기 위해 한유화와 박광배(2008)가 제안한 잠재계층분석 결과를 사용하여 본 연구의 분석자료에 포함된 기소여부와 거짓말탐지검사 결과를 잠재계층분석 결과에 따라 할당하여 분석에 사용한 분석자료들의 실제 거짓말여부를 추정하였다. 거짓말탐지검사의 현장 연구들에서는 피검사자들의 실제 거짓말여부를 판단하기 위해 일반적으로 두 가지 방법이 사용된다(Raskin, 1987). 하나는 피검사자들의 자백을 기준으로 판단하는 것이고, 다른 하나는 사건 개요에 대해 자세히 검토한 전문가들의 결정을 기준으로 판단하는 것이다. 본 연구에서 실제 거짓말여부 추정을 위해 잠재계층분석을 사용한 것은 통계적으로 타당하지만 피검사자들의 실제 거짓말여부를 나타내는 완벽한 지표는 아니므로 자백한 피검사자들의 사례를 분석자료로 사용하여 본 연구와 동일한 절차의 연구를

수행한다면 현재 대한민국에서 이루어지고 있는 CQT가 논리적으로 타당한지에 대한 좀 더 강력한 증거가 될 수 있을 것으로 생각된다.

대한민국의 거짓말탐지검사 판정은 거짓말 탐지검사의 채점결과 뿐만 아니라 검사 당시 피검사자의 행동징후 또는 신체적 상태 등 여러 가지 변인들에 대한 분석 결과에 의해 영향을 받는다. 본 연구에서는 실제 검사자료를 사용하여 분석을 실시하였지만 거짓말탐지검사의 판정에 영향을 주는 다른 요인들은 고려하지 않았다. 그것은 본 연구의 목적은 거짓말탐지검사의 판정이 아닌 CQT의 가정에 대한 타당성을 오류긍정 비율을 사용하여 검증하는 것이었고, 여러 가지 요인을 고려한 거짓말탐지검사의 판정이 잠재계층분석 결과에 반영되어 있기 때문에 실제 검사자료에 포함된 판정을 그대로 사용한다면 잠재계층분석 결과와 중복되어 CQT의 오류긍정 비율을 과소추정 할 수 있기 때문이었다.

본 연구에서는 CQT의 오류긍정비율이 매우 낮다는 것을 검증하였다. 만약 CQT의 가정이 타당하지 않다면, 이러한 결과는 나타날 수 없기 때문에 본 연구의 결과는 CQT의 가정이 타당하다는 것에 대한 하나의 증거가 될 수 있을 것이다. 이것은 추론에 근거한 것이기 때문에 좀 더 실증적인 근거를 제시하기 위해서는 CQT의 기본적인 가정인 실제로 거짓말하는 사람은 관련질문에서, 실제로 진실을 말하는 사람은 비교질문에서 심리적인 부담과 생리적 반응이 더 크게 나타난다는 것을 증명하는 실험 연구가 필수적으로 이루어져야 할 것이다.

본 연구는 합리적인 추론을 바탕으로 CQT의 논리적 타당성을 검증하기 위한 Elaad(2003)의 연구 결과를 해석함에 있어서 발생할 수

있는 모호함을 제거하고 좀 더 명확한 결과 해석을 시도한 것에 의의가 있다고 하겠다. 그러나 모호함을 제거하기 위한 방법으로 사용한 실제 거짓말여부 추정을 위한 잠재계층 분석의 결과는 피검사자들의 실제 거짓말여부를 나타내는 완벽한 지표가 될 수 없으므로 본 연구의 결과 또한 해석과 적용에 신중한 검토가 요구된다.

참고문헌

- 한유화, 박광배 (2008). 범죄수사를 위한 거짓말탐지 검사(polygraph test)의 판정기준과 정확성. 한국심리학회지: 사회문제, 14(4), 103-117.
- Backster, C. (1963). The Backster chart reliability rating method. *Law and Order*, 1, 63-64.
- Ben-Shakhar, G. & Furedy, J. J. (1990). *Theories and Applications in the Detection of Deception: A Psychophysiological and International Perspective*. New York: Springer-Verlag.
- Elaad, E. (2003). Is the inference rule of the "control question polygraph technique" plausible? *Psychology, Crime & Law*, 9, 37-47.
- Horowitz, S. W., Kircher, J. C., Honts, C. R., & Raskin, D. C. (1997). The role of comparison questions in physiological detection of deception. *Psychophysiology*, 34, 108-115.
- Horvath, F. (1988). The utility of control questions and the effects of two control questions types in field polygraph technique. *Journal of Police Science and Administration*, 16, 198-209.
- Lykken, D. T. (1988). *The case against polygraph testing*. In: A. Gale (Ed.), *The Polygraph Test*:

- lies, truth and science. London: Sage.
- Lykken, D. T. (1998). *A Tremor in the blood. Uses and abuses of the lie detector*. Second edition. New York: Plenum Trade.
- Offe, H., & Offe, S. (2007). The comparison question test: Does it work and if so how? *Law and Human Behavior*, 31, 291-303.
- Raskin, D. C. (1986). The polygraph in 1986: Scientific, professional, and legal issues surrounding applications and acceptance of polygraph evidence. *Utah Law Review*, 1986, 29-74.
- Raskin, D. C. (1987). Methodological issues in estimating polygraph accuracy in field applications. *Canadian Journal of Behavioural Science*, 19, 389-404.
- Reid, J. E. (1947). A revised questioning technique in lie detection tests. *Journal of Criminal Law, Criminology and Police Science*, 37, 542-547.
- 1차원고접수 : 2009. 6. 18.
수정원고접수 : 2009. 8. 13.
최종게재결정 : 2009. 8. 14.

Validity of the basic assumption underlying the comparison question technique(CQT) for forensic lie detection

Yu Hwa Han

Kwangbai Park

Chungbuk National University

Comparison Question Technique (CQT) is the method of forensic lie detection that is the most widely used in Korea. CQT is based on the basic assumption that the psychophysiological responses vary depending on whether the subject receives a relevant or a comparison question and whether the subject is telling a lie or the truth. The present study tested the basic assumption underlying the use of CQT by means of a chi-square test for the goodness-of-fit between the false positive rates as expected by Lykken (1988, 1998) and those as estimated from actual data. The false positive rates in actual data were estimated to be generally low in the range of 1% to 7%. On the other hand, Lykken (1988, 1998) expected them to be in the range of 39.5% to 47%. Our analyses indicate that Lykken severely overestimated the false positive rates of CQT and as a result, the basic assumption underlying the use of CQT as a method of lie detection is not as ill-founded as Lykken proposed. The limitations of the present study and further studies to establish the validity of the rationales underlying CQT were discussed.

Key words : polygraph test, comparison question technique(CQT), logical validity, false alarm rate