

범죄수사를 위한 거짓말탐지 검사(polygraph test)의 판정기준과 정확성*

한 유 화

박 광 배†

충북대학교 심리학과

한국의 검찰에서는 거짓말탐지검사(polygraph test)의 최종판정을 위한 기준 점수로써 관행적으로 -12점을 사용하고 있다. 이 판정기준 점수는 검사기법을 개발한 Backster(1963)가 제안한 판정기준 점수(-13점)와는 약간 다른 것으로 한국에서 거짓말탐지 검사를 표준화하기 위해서는 판정기준 점수에 대한 과학적 근거와 그 정확성에 대한 검증이 필요하다. 본 연구에서는 실제 범죄수사를 위하여 검찰에서 이루어진 거짓말탐지 검사 자료를 이용해서 거짓말탐지 검사의 판정기준 점수를 신호탐지이론에 기초하여 설정하고자 하였다. 또한 Backster가 제안한 판정기준 점수와 현재 검찰에서 사용하는 판정기준 점수, 신호탐지이론에 기초해서 본 연구가 설정한 판정기준 점수의 정확성을 비교하여 범죄수사를 위한 거짓말탐지검사의 검사 상황에 가장 적절한 판정기준 점수를 제시하고자 하였다. 신호탐지이론에 기초해서 거짓말탐지검사의 정확성을 최대화할 수 있는 판정기준 점수를 산출한 결과 -8점으로 나타났다. 또한 판정기준 점수를 도출한 표본자료와 다른 별도의 표본자료에서 판정기준 점수에 따른 정확성을 비교한 결과, -8점을 판정기준 점수로 사용했을 경우의 정확성(83.17%)이 가장 높았고, 검찰의 판정기준 점수(80.20%), Backster의 판정기준 점수(76.24%) 순으로 정확성이 감소하였다. 그러나 오류긍정과 오류부정의 비율을 비교한 결과에서는 -8점을 판정기준으로 사용할 때에 오류긍정이 가장 높은 것으로 나타났다. 따라서 범죄수사를 위한 거짓말탐지검사서 전체적인 판단 정확률을 확보하는 것이 중요하다면 -8점을 기준 점수로 사용하고, 진실을 말하는 사람을 거짓이라고 판단하지 않는 것이 중요하다면 -12 또는 -13점을 판정기준 점수로 사용할 것이 권고 되었다.

주요어 : 거짓말탐지 검사(polygraph test), 판정기준 점수, 신호탐지 이론, ROC 곡선

* 이 논문은 2007년도 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

한국의 검찰에서는 Backster(1963)가 개발한 zone comparison test(ZCT) 기법을 사용하여 거짓말탐지검사(polygraph test)를 실시하고 있다. 이 기법은 비교질문 검사(Comparison Question Test)의 일종인데, 2개의 관련질문을 이용하는 방법과 3개의 관련질문을 이용하는 방법으로 구분할 수 있다. 하나의 차트에 포함되는 관련질문의 개수와 최종판단을 위해 채점에 사용하는 차트의 개수에 따라서 최종판정을 위한 기준점 혹은 소위 '절단점(cutoff point)'이 다르게 적용된다. 현재 한국의 검찰에서는 2개의 관련질문을 포함하는 3개의 차트를 채점하여 최종판단을 내리는 방법을 사용하고 있다.

Backster(1963)는 2개의 관련질문을 포함하는 3개의 차트를 사용할 경우에 총점이 -13점 이하이면 '거짓말하는 피검사자'(이하 '거짓'), -12점 ~ +6점 사이이면 '판단불능', +7점 이상이면 '진실을 말하는 피검사자'(이하 '진실')로 판단하도록 기준을 정하고 있다. Backster의 ZCT 및 기타 수정된 ZCT를 사용하던 미국 각 기관에서는 Backster가 제시한 판정기준 점수가 실제 현장에서 사용해 본 결과 부적합하다고 판단하여 각 기관별로 그 기준을 따로 정해서 사용하고 있고, Backster 또한 그 이후에 몇 차례 기준점을 변경한 바 있다(Capps & Ansley, 1992). 한국에서도 마찬가지로 거짓말탐지검사가 도입되어 사용되기 시작한 초기에는 Backster가 제시한 기준을 판정기준 점수로 사용하였으나 현재 한국의 검찰에서는 수년 동안 검사를 수행해 온 결과 이 판정기준이 한국의 검사 상황 또는 검사관 특성에 부적합하다고 판단하여 이 기준을 약간 조정하여 사용하고 있다. 현재 한국에서 사용하고 있는 기준은 -12점 이하이면 '거짓', -11점 ~ 0점

사이이면 '판단불능', 0점 이상이면 '진실'로 판정한다.

현재 한국에서 사용하고 있는 판정기준 점수는 오랜 기간의 경험을 가진 실무자들에 의해서 암묵적으로 굳어지면서 사용하게 된 것으로 현재까지는 이 판정기준 점수를 사용하여 거짓말탐지검사의 최종판정을 내리는 것의 타당성에 대한 과학적인 근거가 전혀 제시되지 않고 있다. 거짓말탐지검사를 표준화하기 위해서는 최종판정을 위한 판정기준 점수의 설정에 대한 과학적 근거와 그러한 판정기준 점수가 가지는 유용성에 대한 검증이 요구된다.

거짓말탐지 검사의 정확성

거짓말탐지검사는 검사차트에 나타나는 피검사자의 생리반응을 보고 피검사자가 거짓을 말했는지를 판단하는 진단검사의 일종이라고 할 수 있다. 거짓말탐지검사는 거짓 또는 진실이라는 판단(진단)의 범주가 있고, 특정한 지표들을 기준으로 사용하여 판단한다는 점에서 다른 진단검사들과 유사하다. 대부분의 진단검사들에서는 '양성' 또는 '음성'의 판단이나 여러 범주 중 어떤 범주에 속하는지에 대한 판단이 검사 결과가 된다. 거짓말탐지검사는 거짓과 진실 중 피검사자의 반응이 어떤 범주에 속하는지를 판단하는 진단검사에 속한다고 할 수 있다. 거짓말탐지검사는 진단검사의 특성들을 가지기 때문에 진단검사의 정확성을 측정하기 위해 사용되는 방법들을 거짓말탐지검사의 정확성을 측정하기 위해 적용할 수 있다(National Research Council, 2003).

일반적인 진단검사의 판정기준은 판정의 목적에 따라 다르게 적용된다. 진단목적에 따라

서 진단대상의 존재를 놓쳐서는 안 되는 경우에는(예를 들면, 전쟁 중 적군의 간첩) 오류긍정(false positive)과 민감도(옳은 긍정)가 모두 높은 관대한 판정기준 점수(거짓말탐지검사의 경우 낮은 점수)를 설정하는 것이 적절하고, 오류긍정(false positive)을 피하는 것이 중요한 경우에는 오류긍정과 민감도가 모두 낮은, 엄격한 판정기준 점수(거짓말탐지검사의 경우 높은 점수)를 설정하는 것이 적절하다(National Research Council, 2003). 거짓말탐지검사의 경우도 마찬가지로 거짓말하는 사람을 정확히 알아내는 것이 중요한지, 진실을 말하는 사람을 거짓말하는 것으로 오해하지 않는 것이 중요한지에 따라 판정기준 점수를 다르게 적용할 필요가 있다. 따라서 판정기준 점수를 설정할 때에는 오류긍정 또는 오류부정(false negative)을 범했을 경우에 따르는 결과의 심각성을 고려해야 한다.

수용자특성곡선(receiver operating characteristic curve; ROC curve)은 진단검사에서 절단점의 변화에 따라서 진단의 정확성(민감도와 특이도)이 어떻게 변화할지를 보여주고 이에 따라 최적의 절단점을 제시하기 위한 목적으로 주로 이용된다(Park, Koo, Hwang & Youn, 2002). 민감도와 특이도는 신호탐지이론에 기초하고 있는 개념으로 ROC 곡선은 민감도와 특이도를 동시에 그림으로 제시해 준다. 많은 지표들이 ROC 곡선에 포함된 정보들을 요약하는 용도로 쓰여 왔고(Swets, 1986), 그 중 ROC 곡선 아래의 면적을 나타내는 지표인 AUC(Area Under the ROC curve)는 두 개의 모집단을 분리된 집단으로 구분하는 검사의 능력에 대한 측정치이다. 이 지표가 .5이면 검사가 두 집단을 구분하는 것이 우연수준과 같다는 것을 의미하고 1에 가까울수록 두 집단을 완벽하게 구분

한다는 것을 의미한다(Kraemer, 1992).

진단검사의 정확성을 나타내는 지표들 중에는 민감도(sensitivity)와 특이도(specificity)가 흔하게 사용되는데 민감도는 특정 범주에 속하는 사건(사람)을 그 범주에 속한다고 정확히 판단할 확률을 말하고, 특이도는 특정 범주에 속하지 않는 사건(사람)을 그 범주에 속하지 않는다고 정확히 판단할 확률을 말한다. 완벽한 민감도를 가진 검사에서는 특정 범주에 속하는 사건(사람)을 그 범주에 속하지 않는다고 판단하는 오류(오류부정)가 전혀 나타나지 않는 반면에 완벽한 특이도를 가진 검사에서는 특정 범주에 속하지 않는 사건(사람)을 그 범주에 속한다고 판단하는 오류(오류긍정)가 전혀 나타나지 않는다.

신호탐지이론에 기초해서 판정기준 점수를 설정하기 위해서는 사건(사람)이 실제로 어떤 범주에 속하는 것인지를 알 수 있는 준거(standard)가 필요하다. 거짓말탐지검사의 경우에 이 준거는 피검사자가 실제로 거짓말을 하는 사람인지, 혹은 실제로 진실을 말하는 사람인지에 대한 객관적이고 절대적인 실제적 진실(ground truth: GT)을 의미한다. 일반적으로 GT는 알 수 없는 것이지만, 최근에는 생물학, 화학 등의 자연과학(Albert, McShane & Shih, 2001)과 법심리학분야(Spencer, 2007)에서 잠재계층분석(latent class analysis: Lazarsfeld & Henry, 1968; Goodman, 2002)을 이용하여 그것을 추정하는 방법이 많이 활용되고 있다.

본 연구는 실제 범죄수사과정에서 이루어진 거짓말탐지검사 자료를 이용하여 피검사자의 GT를 추정하기 위해 수행된 연구 1과, 연구 1에서 추정된 GT를 이용하여 판정기준 점수를 설정하고 그 정확성을 비교하기 위해 수행된 연구 2로 이루어졌다. 연구 1에서는 검찰에서

이루어진 거짓말탐지검사 자료에 대하여 잠재 계층분석을 실시하여 피검사자의 GT를 나타내는 지표를 발견하고, 연구 2에서는 별도의 검찰자료에 대하여 연구 1에서 파악된 GT지표들을 적용하여 신호탐지이론에 기초한 분석을 실시한 후 거짓말탐지검사의 타당성(정확성)을 최대화 할 수 있는 판정기준 점수를 확인하고 기존의 판정기준과 정확성을 상호 비교하였다. 범죄수사 실무에서 거짓말탐지검사의 결과를 사용할 때의 관심은 ‘거짓’의 여부를 구분하는데 있으므로 본 연구에서는 거짓, 판단불능, 진실 범주로 구분되는 거짓말탐지검사의 최종판단 범주들 중 판단불능 범주를 진실 범주에 포함시켜 ‘거짓’과 ‘거짓 아님’의 두 가지 범주를 구분하기 위한 분석만을 실시하였다.

연구 1

신호탐지이론에 기초하여 거짓말탐지검사의 판정기준 점수를 설정하기 위해서는 피검사자가 실제로 거짓말을 하였는지 혹은 실제로 진실을 말하였는지를 알 수 있는 준거를 확보하는 것이 필수적이다. 따라서 연구 1에서는 잠재계층분석을 실시하여 GT를 추정하였다. 잠재계층 분석을 실시하기 위해서 검찰의 기소여부와 거짓말탐지 검사의 최종판정을 지표(indicator)로 사용하였다. 검찰의 기소여부는 그 정확성이 알려져 있지 않지만 실제 재판에서 기소된 피의자가 유죄판결을 받는 비율로써 추정할 수 있다. 대검찰청(2008)의 보고에 따르면 검찰에 의해 기소된 피의자가 재판에서 유죄판결을 받는 비율은 95% ~ 99%이다. 재판에서의 유죄판결 비율로 보아 검찰의 기소

여부는 피의자의 범죄여부에 대한 매우 정확한 지표가 될 수 있다. 거짓말탐지 검사의 정확성은 검사기법과 채점방법에 따라 약간의 차이가 있다. 현재 한국 검찰에서 사용하고 있는 검사기법인 비교질문 검사의 정확성에 대한 많은 실험 연구들(Driscoll, Honts & Jones, 1987; Ginton, Netzer, Elaad & Ben-Shakher, 1982; Honts, Raskin & Kircher, 1994; Horowitz, Kircher, Honts & Raskin, 1997; Kircher & Raskin, 1988; Podlesny & Raskin, 1978; Podlesny & Truslow, 1993; Raskin & Hare, 1978; Rovner, Raskin & Kircher, 1979)의 결과를 종합하면 거짓말탐지검사가 진실을 말하는 사람과 거짓말을 하는 사람을 정확하게 구분하는 정도는 66.5% ~ 92.5%로 나타났고, 평균적으로는 91%의 피검사자를 정확하게 구분하는 것으로 나타났다. 또한 현장 연구들(Honts, 1996; Honts & Raskin, 1988; Patrick & Iacono, 1991; Raskin, Kircher, Horowitz & Honts, 1989)의 결과에서는 67% ~ 91.5%의 정확성을 보였으며 평균적으로는 실제 피검사자의 90.5%를 정확하게 구분하는 것으로 나타났다. 따라서 검찰의 기소여부와 거짓말탐지 검사의 판정은 완벽하지는 않으나 실제로 거짓말을 하는 사람과 그렇지 않은 사람을 파악할 수 있는 지표로 사용될 수 있고, 이 두 개의 지표들은 서로 상당히 독립적으로 생성된다(여기에서 형용사 “독립적”의 의미는 인과관계에서의 독립성을 의미하는 것이고 통계적 상관관계에서의 독립성을 의미하는 것이 아니다). 거짓말탐지검사는 재판과정에서 증거 가치를 인정받지 못하기 때문에 검사는 거짓말탐지 검사에 근거하여 기소결정을 할 수 없다. 따라서 실제 검찰실무에서 검사는 다른 직접증거 혹은 정황증거들에 의해 피의자 범행에 대하여 상당한 확신을 가진 후에 피의자

혹은 고소인에 대한 거짓말탐지검사를 의뢰하는 것이 일반적인 수사관행이다. 그때 거짓말탐지검사를 실시하는 주된 목적은 기소결정을 위한 것이기 보다 피의자로부터 자백을 유도하여 검사가 다른 증거들에 의해 이미 형성한 심증에 대한 보강을 하고, 재판에서 유죄판결을 보장하기 위한 경우가 많다. 수사사례들 중에는 거짓말탐지검사에서 “진실” 판정이 있었음에도 불구하고 기소되는 사례도 많고, 그 반대의 경우도 많다. 따라서 검사의 기소결정이 거짓말탐지검사의 결과로부터 완전히 독립적으로 이루어지지 않는 것이라든가 거짓말탐지검사에 준하여 기소결정이 이루어지는(촉발되는) 경우는 극히 적다.

방 법

분석자료

피검사자의 GT를 추정하기 위한 잠재계층 분석을 위하여 2006년 한 해 동안 한국의 대검찰청 및 지방검찰청에서 이루어진 거짓말탐지검사의 실제 자료 중 기소여부가 알려진 182개 사례를 선정하여 사용하였다. 거짓말탐지검사를 실시하고 차트를 채점한 검사관들은 모두 전문 교육을 이수하였으며 평균 3년 이상의 실무 경력을 가지고 있었다. 분석자료에 포함된 피검사자들의 연령은 17세~82세(평균: 44.94세, 표준편차: 12.21세)까지로 다양하였으며 이 중 남성이 127(69.8%)명, 여성이 55(30.2%)명이었다. 피검사자의 대부분은 피의자(114명, 62.6%) 또는 고소인(41명, 22.5%)이었고, 피해자(14명, 7.7%), 참고인(7명, 4.9%) 등도 포함되어 있었다. 피검사자들이 거짓말탐지검

사를 받았을 당시에 내려진 최종판정이 ‘거짓’인 사례는 80(44%)개, ‘진실’인 사례는 58(31.9%)개, ‘판단불능’인 사례는 44(24.2%)개 포함되어 있었다. 또한 해당 사건의 피의자가 기소된 사례가 85(46.7%)개, 불기소된 사례가 84(46.2%)개, 기소중지된 사례가 13(7.1%)개였다. 피검사자들의 거짓말탐지검사의 총점은 -23 ~ 10점 사이에 분포하였고, 평균은 -8.4점(표준편차: 7.42)이었다.

분석

Albert, McShane와 Shih(2001)의 연구와 엄진섭, 지형기와 박광배(2008)의 연구에서 사용한 방법과 같이 피검사자의 GT를 추정하기 위해서 2개의 잠재계층(실제로 거짓말한 계층과 진실을 말한 계층)을 가정하는 잠재계층모형을 검증하였다. 모형은 두개의 관찰변인을 포함하는데, 관찰변인 중 하나는 3개의 범주(거짓/진실/판단불능)로 이루어진 거짓말탐지검사의 최종판정이고 다른 하나는 역시 3개의 범주(기소/불기소/기소중지)로 이루어진 해당사건의 피의자에 대한 검사의 기소여부이다. 거짓말탐지검사의 피검사자가 피의자인 사례에서는 그 피의자에 대한 검사의 기소여부를 곧바로 분석에 사용하였다. 반면, 거짓말탐지검사의 피검사자가 고소인 또는 피해자인 사례에서는 해당 사건의 피의자가 기소된 경우에는 기소여부 변인에서 ‘불기소’로 코딩되었고, 피의자가 불기소된 경우에는 ‘기소’로 코딩되었다. 피검사자가 고소인 또는 피해자인 사례에서 해당 사건의 피의자가 기소중지된 경우는 없었다.

잠재계층분석을 이용하여 2잠재계층 모형을 검증하기 위해서 두 가지 제약(restriction)을 사

용하였다. 하나는 두 계층 모두에서 검사관의 판단이 판단불능으로 나타날 확률이 동일하다는 제약이고, 다른 하나는 실제로 진실을 말하는 계층에서 거짓말 탐지검사의 판정이 ‘거짓’으로 나타날 확률과 실제로 진실을 말하는 계층에 속하는 사람이 검찰에 의해 기소될 확률이 동일하다는 제약이다. 이 두 가지 제약 중 첫 번째 제약은 판단불능 판정을 할 확률이 진실로 거짓을 말하는 사례나 진실로 진실을 말하는 사례에서 유사하다는 미국 의회의 기술평가국(U.S. Congress, Office of Technology Assessment, 1983)의 연구 결과에 기초한 것이다. 또한 두 번째 제약은 거짓말탐지 검사에서는 오류 긍정이 비교적 드물다는 미국 의회의 기술평가국의 연구결과와 한국의 형사재판에서 무죄 판결이 나올 확률(즉, 무고한 사람이 검찰에 의해 기소될 확률)도 이와 유사하게 드물다는 국내의 연구(대검찰청, 2008) 결과에 기초하였다.

결과 및 논의

잠재계층분석을 이용하여 두 가지 제약을 포함하는 2잠재계층 모형을 검증한 결과 모형 합치도는 매우 우수하였으며(likelihood ratio $\chi^2 = 0.63$, $df=1$, ns), 실제로 거짓말하는 계층의 확률은 .64였고, 실제로 진실을 말하는 계층의 확률은 .36이었다. 사후확률에 따라 잠재계층 분석에 사용한 182개 사례를 각 잠재계층에 할당한 결과 110(60.4%)개 사례가 실제로 거짓말 하는 계층에, 72(39.6%)개 사례가 실제로 진실을 말하는 계층에 할당된 것으로 나타났다. 또한 182개의 사례 중 95.56%가 옳은 잠재계층에 정확히 할당된 것으로 추정되었다. 관찰변인의 범주패턴에 따른 사후확률과 그에 따라 할당된 잠재계층을 표 1에 제시하였다.

표 1에 제시된 바와 같이 잠재계층분석에 의하여 검찰이 불기소 결정을 하고 거짓말탐지 검사에서 ‘진실’판정이 된 피검사자, 검찰이 기소중지 결정을 하고 거짓말탐지 검사에서 ‘진실’판정이 된 피검사자, 검찰이 불기소

표 1. 관찰변인의 범주패턴에 따른 사후확률과 잠재계층할당

관찰변인의 범주패턴			사후확률		잠재계층할당
기소여부	거짓말탐지 검사 최종판정	n	“거짓” 잠재계층	“진실” 잠재계층	
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

결정을 하고 거짓말탐지 검사에서 ‘판단불능’ 판정이 된 피검사자는 실제로 진실을 말하는 계층에 속하는 것으로 추정되었다. 이들의 잠재계층(‘진실’)이 틀리게 추정되었을 확률은 0.103이다. 이 오류확률은 관점에 따라서 작게 또는 크게 평가될 수 있다. 그러나 표 1을 보면 알 수 있듯이, 이들 중 대다수(72명 중 50명)를 차지하는 사람들은 검찰에 의해 불기소되고, 거짓말탐지 검사에서 진실 판정을 받은 사람들이고 이들에 대한 잠재계층 할당이 오류일 확률은 0.04에 불과하다. 따라서 ‘진실’계층으로 할당된 사례들이 오류일 경우 발생하는 문제의 심각성은 실질적으로 크지 않을 것으로 판단된다.

그 밖의 모든 피검사자는 실제로 거짓을 말하는 피검사자로 추정되었다. 실제로 거짓을 말하는 것으로 추정된 피검사자들의 잠재계층이 틀리게 추정되었을 확률은 0.006이다. 검찰이 기소중지 결정을 하고 거짓말탐지 검사에서 ‘진실’판정이 된 3명의 피검사자는 실제로 진실을 말하는 계층에 속하는 것으로 추정되었는데, 이 추정이 틀린 것일 확률(0.48)이 매우 높은 편이지만 피검사자의 숫자가 매우 작으므로 전체 잠재계층 추정 정확성에는 거의 영향을 주지 않는다.

연구 2

연구 2에서는 연구 1에서 추정된 GT와 별개의 검사자료를 바탕으로 신호탐지이론에 기초하여 거짓말탐지 검사의 정확성을 최대화할 수 있는 최종 판단기준 점수를 설정하였다. 또한 Backster의 판정기준 점수, 검찰의 현행 판정기준 점수와 신호탐지이론에 기초하여 설

정한 판정기준 점수의 정확성과 오류긍정 및 오류부정 비율을 비교하여 어떤 판정기준 점수를 사용하는 것이 가장 적합한지를 검증하였다.

방 법

분석자료

분석자료는 2006년 한 해 동안 대검찰청 및 지방검찰청(대전, 부산, 창원, 청주 등)에 거짓말탐지 검사가 의뢰된 사건과 관련된 피의자, 피해자 또는 고소인 등을 대상으로 거짓말탐지 검사를 실시한 후 얻은 채점자료였다. 연구 2의 분석자료에는 앞서 연구 1에서 사용된 자료가 포함되지 않았다. 연구 1에서 사용된 자료를 제외한 약 400여 명의 피검사자에 대한 검사 결과 중 기소여부와 거짓말탐지 검사의 최종판정 결과가 명확히 확인되는 사례 208개를 선별하여 연구 2를 위한 분석에 사용하였다. 거짓말탐지 검사를 실시하고 차트를 채점한 검사관들은 모두 전문 교육을 이수하였으며 평균 3년 이상의 실무 경력을 가지고 있었다.

분석자료는 총 208개 사례였고, 실제로 검사를 받은 피검사자들의 연령은 17세 ~ 82세(평균: 44.5세, 표준편차: 11.92세)까지로 다양하였다. 이 중 남성이 147(70.7%)명, 여성이 61(29.3%)명이었으며 피검사자의 대부분은 피의자(138명, 66.3%), 고소인(42명, 20.2%), 피해자(14명, 6.7%) 신분이었다. 피검사자들이 실제 거짓말탐지 검사를 받았을 당시에 내려진 최종판정이 ‘거짓’인 사례가 95(45.7%)개, ‘진실’인 사례가 69(33.2%)개, ‘판단불능’인 사례가

44(21.2%)개 포함되어 있었다. 또한 분석자료 중 해당 사건의 피의자가 기소된 사례가 98(47.1%)개, 불기소된 사례가 97(46.6%)개, 기소중지된 사례가 13(6.3%)개였다. 피검사자들의 거짓말탐지 검사 총점은 -23점~10점 사이에 분포하였고, 평균은 -8.50점(표준편차: 7.64)이었다.

분석

연구 2의 분석을 위하여 연구 1에서 파악된 잠재계층이 준거로 사용되었다. 즉, 검찰이 불기소 결정을 하고 거짓말탐지 검사에서 ‘진실’ 판정이 된 피검사자, 검찰이 기소중지 결정을 하고 거짓말탐지 검사에서 ‘진실’판정이 된 피검사자, 그리고 검찰이 불기소 결정을 하고 거짓말탐지 검사에서 ‘판단불능’판정이 된 피검사자는 실제로 진실을 말하는 사람으로 간주되었다. 그 밖의 모든 피검사자는 실제로 거짓을 말하는 피검사자로 간주되었다. 분석 자료에 포함된 사례 중 연구 1의 결과에 의해 실제로 거짓말하는 것으로 판단되는 사례는 127(61.1%)개, 실제로 진실을 말하는 것으로 판단되는 사례는 81(38.9%)개인 것으로 나타났다.

연구 1에서 파악된 준거를 이용하여 연구 2의 분석은 두 가지 부분으로 나누어진다. 첫째는 거짓말탐지 검사의 정확성을 최대화 할 수 있는 판정기준 점수를 설정하기 위해 신호탐지이론에 기초한 분석을 실시하는 것이었으며, 둘째는 Backster가 제시한 판단기준 점수, 현재 검찰에서 사용 중인 판단기준 점수, 그리고 신호탐지 이론에 기초하여 설정한 판단기준 점수를 각각 사용하여 정확성과 오류긍정 및 오류부정의 비율을 비교하는 것이었다. 연구 2에서 이루어진 두 개의 분석은 각기 다

른 자료에 의해 이루어졌다. 판정기준의 설정을 위한 분석에는 전체 208개 사례의 분석자료 중 무선적으로 107개 사례를 선택하여 사용하였고, 나머지 101개 사례를 사용하여 정확성과 오류긍정 및 오류부정의 비율을 비교하였다.

결과 및 논의

신호탐지이론에 기초한 판정기준 점수의 설정

신호탐지 이론에 기초하여 거짓말탐지 검사의 민감도와 특이도를 평가하고 최적의 판단기준 점수를 산출하였다. ROC 곡선은 곡선 아래의 면적(AUC)이 넓을수록 더 높은 정확성을 갖는 것으로 해석할 수 있는데 이 값은 .921로 나타났다. 정확성이 가장 높은 최적의 절단점은 -8점으로 산출되었으며, 이 절단점에서의 민감도와 특이도는 각각 .84과 .86으로 나타났다. 이것은 -8점 이하의 총점을 가진 사례를 거짓으로 판단하고 -8점을 초과하는 총점을 가진 사례를 “진실”로 판단한다면 실제로 거짓말하는 사람을 “거짓”으로 판정할 확률이

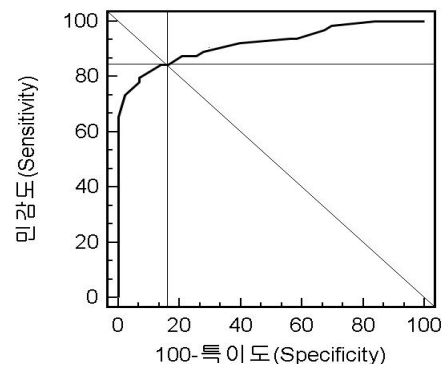


그림 1. 거짓말탐지검사의 ROC 곡선

표 2. 판정기준 점수에 따른 민감도와 특이도

기준 점수	민감도(sensitivity)	특이도(specificity)
≤ -13	0.66	31.00
≤ -12	0.73	0.98
≤ -11	0.78	0.93
≤ -10	0.80	0.93
≤ -9	0.81	0.91
≤ -8*	0.84	0.86
≤ -7	0.84	0.84

.84이며, 실제로 진실을 말하는 사람을 “진실”로 판정할 확률이 .86이라는 것을 의미한다. 신호탐지 이론에 기초한 분석 결과 도출된 ROC 곡선을 그림 1에 제시하였고 판정기준 점수의 변화에 따른 민감도와 특이도를 표 2에 제시하였다.

정확성 비교

Backster의 판정기준 점수, 검찰의 현행 판정기준 점수와 신호탐지이론에 기초한 판정기준 점수를 사용하여 101개의 거짓말탐지검사 결과를 거짓과 진실로 다시 분류하여 각 경우의 정확성을 세부적으로 비교하였다. 정확성은 전체 사례 중 실제 거짓말 여부가 진실인 사례를 진실로 판정한 경우와 실제 거짓말 여부가 거짓인 사례를 거짓으로 판정한 경우의 합 비율로 산출하였다.

Backster의 판정기준 점수를 사용하여 분류한 경우 정확성은 76.24%(77/101)로 나타났고, 검찰의 현행 판정기준 점수를 사용하여 분류한 경우 정확성은 80.20%(81/101)로 나타났다. 마지막으로 신호탐지이론에 기초한 판

정기준 점수를 사용하여 분류한 경우 정확성은 83.17%(84/101)였다(표 2). 정확성을 비교한 결과, 본 연구에서 설정한 판정기준 점수를 사용했을 경우의 정확성이 가장 높았지만, 이 경우 진실한 사람을 거짓말하는 사람으로 판단하는 비율이 5.3%로 나타났다. 반면, Backster의 판정기준 점수와 검찰의 현행 판정기준 점수를 사용했을 경우의 정확성은 신호탐지이론에 기초한 판정기준 점수를 사용하는 경우보다 낮았으나 진실한 사람을 거짓말하는 사람으로 판단하는 오류를 전혀 범하지 않는 것으로 나타났다.

오류긍정(false positive)과 오류부정(false negative) 비율의 비교

Backster의 판정기준 점수와 현행 판정기준 점수를 사용했을 경우에는 실제 진실을 말한 사람을 거짓말하는 사람으로 판단한 사례가 하나도 존재하지 않았던 것으로 나타나 오류긍정 비율이 0%였으며, 신호탐지이론에 기초한 분석의 판정기준 점수를 사용했을 경우에는 실제 진실을 말한 사람을 거짓말하는 사람으로 판단한 사례가 2개 존재하는 것으로 나타나 오류긍정 비율이 5.3%였다. 오류긍정비율에 대하여 쌍비교를 위한 t-검증을 수행한 결과 Backster 기준과 현행 판정기준의 차이는 존재하지 않고, Backster의 판정기준과 신호탐지이론에 기초한 판정기준의 차이는 유의미하지 않았고($t=1.434, df=37, p=0.160$), 현행 판정기준과 신호탐지이론에 기초한 판정기준의 차이도 동일하게 유의미하지 않았다. 반면에 Backster의 판정기준 점수를 사용했을 경우에는 실제 거짓말한 사람을 진실을 말한 것으로 판단한 사례가 24개 존재한 것으로 나타나 오

표 3. 판정기준에 따라 “진실”, “거짓”으로 분류되는 사례 수 (%)

실제 거짓말 여부(GT)	Backster 판정기준			현행 판정기준			신호탐지이론에 기초한 판정기준		
	진실	거짓	전체	진실	거짓	전체	진실	거짓	전체
진실	38 (100)	0 (0)	38 (100)	38 (100)	0 (0)	38 (100)	36 (94.7)	2 (5.3)	38 (100)
거짓	24 (38.1)	39 (61.9)	63 (100)	20 (31.7)	43 (68.3)	63 (100)	15 (23.8)	48 (76.2)	63 (100)
전체	62 (61.4)	39 (38.6)	101 (100)	58 (57.4)	43 (42.6)	101 (100)	51 (50.5)	50 (49.5)	101 (100)

류부정 비율이 38.1%였고, 검찰의 현행 판정 기준 점수를 사용했을 경우에는 20개 존재한 것으로 나타나 31.7%, 신호탐지이론에 기초한 판정기준 점수를 사용했을 경우에는 15개 존재한 것으로 나타나 23.8%였다(표 3). 오류부정비율에 대하여 쌍비교를 위한 t-검증을 실시한 결과 Backster의 판정기준과 현행 판정기준의 차이는 유의미 하였고 ($t=2.050$, $df=62$, $p=0.045$), Backster의 판정기준과 신호탐지이론에 기초한 판정기준의 차이도 또한 유의미하였으며($t=3.215$, $df=62$, $p=0.002$), 현행 판정기준과 신호탐지이론에 기초한 판정기준의 차이도 유의미하였다($t=2.312$, $df=62$, $p=0.024$).

비록 본 연구에서 신호탐지이론에 기초하여 설정한 기준점인 -8점이 Backster의 판정기준이나 현행 판정기준에 비해서 통계적으로 유의미한 정도로 오류공정비율을 높이지는 않지만, 실제 거짓말 여부(GT)가 “진실”인 사례의 수(38)가 상대적으로 적고, 오류공정비율의 기저선이 매우 낮아서 유의수준에 도달하지 못했을 가능성을 배제할 수 없다. 따라서 본 연구에서 신호탐지이론에 기초하여 설정한 기준점인 -8점이 오류공정비율을 높일 수 있는 가능

성을 완전히 배제하기는 어려운 것으로 판단된다. 결론적으로, 오류공정비율은 신호탐지이론에 기초한 판정기준을 사용했을 경우에 가장 높았지만 오류부정 비율은 신호탐지이론에 기초한 판정기준을 사용했을 경우에 가장 낮았다. 종합하면 판정기준이 높아질수록 오류공정 비율은 높아지며 오류부정 비율은 낮아지는 것으로 나타났다.

종합논의

한국의 검찰에서 범죄수사를 위하여 사용하는 거짓말탐지검사에 대하여 신호탐지이론을 적용한 분석을 시도한 결과, 총점이 -8점 이하이면 거짓으로 판단하고 -8점보다 높으면 진실로 판단할 경우에 거짓말탐지검사의 종합적인 정확성이 최대가 되는 것으로 나타났다. 반면, -8점을 판정기준으로 사용하면 오류공정비율이 다소 높아질 수 있는 것으로 나타났는데, 그 이유는 신호탐지이론에 기초한 판정기준 점수(-8점)가 세 가지 판정기준 점수 중 가장 거짓 판정을 내리기 쉬운 판정 기준이기

때문이다. 오류긍정과 오류부정에 의해 초래되는 사회적 비용이 동일하다고 간주하면, 거짓말탐지검사의 종합적인 정확성을 높일 수 있는 판정기준 점수를 설정하기 위해서 신호탐지이론에 기초한 -8점의 판정기준 점수를 사용하는 것이 가장 합당할 것이다. 그러나 거짓말탐지 검사에서는 거짓말하는 사람을 정확히 가려내는 것과 동시에 거짓말하지 않는 즉, 실제로 진실한 사람을 거짓말한다고 판단하지 않는 것이 더 중요할 수도 있다. 따라서 범죄수사 실무에서 거짓말탐지 검사의 기능과 목적에 따라 종합적인 정확성이 중요할 때는 -8점을 판정기준 점수로 사용하고, 진실을 말하는 사람을 거짓말한다고 판단하지 않는 것이 중요할 때는 오류긍정 비율이 0%인 -12점 또는 -13점을 판정기준 점수로 사용할 것이 권고된다.

연구 1(GT의 결정)에서 진실된 것으로 추정된 피검사자 중 검찰의 현행 판정기준 점수(-12점)를 사용했을 경우에는 “진실”로 판정된 피검사자 2명이 본 연구에서 설정한 판정기준 점수(-8점)를 사용하면 “거짓”으로 판정되었다. 또한 GT에서 거짓말 하는 것으로 추정된 피검사자 중 현행 판정기준 점수를 사용해서 “진실” 판정을 받았던 피검사자 5명이 본 연구에서 설정한 판정기준 점수를 사용했을 경우에는 “거짓”으로 판정되었다. 판정기준 점수를 -12점에서 -8점으로 바꾸는 경우에 거짓말탐지검사의 판정이 바뀌는 7개의 사례들을 추적하여 살펴본 결과, GT가 진실인 경우(2개 사례) 판정이 변화한 사례들은 -8점을 판정기준으로 사용할 경우 판정오류가 될 가능성이 높다. 왜냐하면 그 2개의 사례들이 검사에 의해 불기소된 사례들이기 때문이다. 그러나 GT가 거짓인 경우(5개 사례)에는 기소된 사례 4

개, 기소 중지된 사례 1개의 판정이 변화하였다. 이 사례들의 경우에는 -8점을 판정기준으로 사용할 경우에 오히려 거짓말 여부에 대하여 정확한 판정이 될 가능성이 높다. 따라서 본 연구에서 설정한 판정기준 점수(-8점)를 사용할 때, 검찰의 현행 판정기준 점수(-12점)를 사용할 때보다 오류긍정이 늘어나지만 오류부정이 줄어드는 정도가 더 크고, 그 때문에 전체 정확률이 더 높아지는 것으로 판단된다.

본 연구의 서론에서 거짓말탐지검사는 피검사자가 여러 개(3개)의 범주 중 어느 범주에 속하는지를 판단하는 진단검사의 일종이라고 하였으나 실무에서 거짓말탐지 검사를 사용하는데 있어서 중요한 것은 거짓과 거짓이 아닌 경우를 판단하는 것이므로 판단불능 범주와 진실 범주를 구분하기 위한 판정기준 점수를 알아내기 위한 분석은 실시하지 않았다. 다시 말하면 거짓말탐지 검사의 결과가 참고자료로 사용 될 때 피검사자 즉, 피의자 또는 피해자가 사건에 대해 거짓말을 하고 있는지 그렇지 않은지가 주된 관심의 대상이고 피검사자가 거짓말을 하는지 진실을 말하는지를 구분할 수 없는 상태를 밝혀내는 것은 실질적인 의미를 가지지 않는다. 따라서 판단불능과 진실 범주의 구분을 폐기하거나, 그 구분을 존속시켜야만 할 경우에는 그 구분을 위한 판정기준 점수는 오랫동안 거짓말탐지 검사를 실시해 온 검사관의 실무감각에 맡겨도 무방할 것으로 사료된다. 그러나 만약 수사실무에서 판단불능 범주와 진실 범주를 구분하는 것이 중요한 의미를 갖는다면 향후 두 범주를 구분하기 위한 적절한 판정기준 점수도 본 연구에서와 같은 방법으로 과학적인 분석이 이루어진 후에 설정할 필요가 있다.

본 연구의 의의는 현재 한국의 대검찰청 및

지방 검찰청에서 거짓말탐지 검사를 위해 실시하고 있는 검사 기법이 개발 되었을 당시에 정해진 판정기준 점수가 아닌 실무 현장에서 수정되어 온 판정기준 점수를 사용하고 있는 실정에서 각 판정기준 점수들을 사용했을 경우의 정확성과 오류긍정 및 오류부정 비율을 비교한 결과를 바탕으로 실무에 가장 적합한 판정기준 점수를 설정함으로써 실무 현장의 검사관들 사이에서 암묵적으로 수정되어 사용되고 있는 판정기준 점수의 유용성을 과학적 분석을 통해 설정한 판정기준 점수와 비교하고 현행 판정기준 점수의 우수성을 검증한 것에 있다.

실제 거짓말탐지 검사의 최종판정을 위해서는 검사 결과 얻어지는 점수를 비롯하여 검사 당시 피검사자의 상태 또는 행동징후(왜곡의 시도 여부) 등을 모두 고려하고 있다. 그러나 본 연구에서는 이러한 요인들은 고려하지 않은 거짓말탐지검사의 총점만을 사용하여 적절한 판정기준 점수를 파악하였으며, 정확성과 오류긍정 및 오류부정의 비율을 비교하기 위해서도 거짓말탐지검사의 점수만을 사용하였다. 따라서 본 연구의 결과를 해석할 때에는 거짓말탐지검사에서 모든 가능한 요인들을 함께 고려하여 내려지는 최종판정의 정확성을 비교하여 판정기준 점수를 설정한 것이 아니고 총점만을 사용하여 정확성을 비교하고 유용한 판정기준 점수를 제시하였다는 점을 유념할 필요가 있을 것이다.

마지막으로, 본 연구에서는 거짓말탐지검사의 각기 다른 판정기준 점수가 초래하는 정확성의 차이를 파악하기 위하여 연구 1에서 검사의 기소여부를 이용한 준거(standard)를 사용하였다. 한편, 재판에서의 유무죄 판결여부를 또 다른 관찰변인으로 포함시켰다면 더욱 타

당한 준거가 도출되었을 가능성도 생각해볼 수 있다. 그러나 현재 한국의 재판상황에서는 형사사건의 유죄 판결률이 95%-99%에 이르므로 기소여부와 유무죄 판결여부가 거의 동일한 변인이다. 따라서 판결여부를 또 하나의 관찰변인으로 포함하는 경우에도 크게 다른 결과가 나올 것으로 예상되지 않는다. 다만, 유무죄 판결여부가 관찰변인으로 포함되면 잠재계층분석의 결과에 대해 조금이나마 더 확신할 수 있을 것으로 생각되고, 그런 이유로 후속연구에서 재판의 결과를 준거로 사용하는 방법에 대하여 검토할 필요가 있다.

참고문헌

- 대검찰청. (2008). 2007년 종합 심사 분석. 서울: 대검찰청
- 엄진섭, 지형기, 박광배 (인쇄중). 폴리그래프 검사의 정확도 추정. 한국심리학회지: 사회문제, 14(4).
- Albert, P. S., McShane, L. M., & Shih, J. H. (2001). *Latent class modeling approaches for assessing diagnostic error without a gold standard: With applications to p.53 immunohistochemical assays in bladder tumors*. Biometrics, 57, 610-619
- Backster, C. (1963). The Backster chart reliability rating method. *Law and Order*, 1, 63-64.
- Capps M. & Ansley N. (1992), Analysis of Federal Polygraph Charts By Spot and Chart Total. *Polygraph*, 21(2), 110-131.
- Driscoll, L. N., Honts, C. R. & Jones, D. (1987). The validity of the positive control physiological detection of deception

- technique. *Journal of Police Science and Administration*, 15, 46-50.
- Ginton, A., Netzer, D., Elaad, E. & Ben-Shakher, G. (1982). A method for evaluating the use of the polygraph in a real-life situation. *Journal of Applied Psychology*, 67, 131-137.
- Goodman, L. A. (2002). *Latent class analysis: The empirical study of latent types, latent variables, and latent structures*. In J. A. Hagenaars & A. L. McCutcheon (Eds.). *Applied Latent Class Analysis*. Cambridge: Cambridge University.
- Honts, C. R. (1996). Criterion development and validity of the control question test in field application. *The Journal of General Psychology*, 123, 309-324.
- Honts, C. R. & Raskin, D. C. (1988). A field study of the validity of the directed lie control question. *Journal of Police Science and Administration*, 16, 56-61.
- Honts, C. R., Raskin, D. C. & Kircher, J. C. (1994). Mental and physical countermeasures reduce the accuracy of the polygraph tests. *Journal of Applied Psychology*, 79, 252-259.
- 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.
- Kircher, J. C. & Raskin, D. C. (1988). Human versus computerized evaluations of polygraph data in laboratory setting. *Journal of Applied Psychology*, 73, 291-302.
- Kraemer H. C. (1992). *Evaluating medical tests, objective and quantitative guidelines*. London, Sage Publications, 63-95.
- Lazarsfeld, P. F. & Henry, N. W. (1968). *Latent Structure Analysis*, Boston: Houghton Mifflin.
- National Research Council (2003). *The Polygraph and Lie Detection*. Committee to Review the Scientific Evidence on the Polygraph. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- Park, Son Il, Koo, Hee Seung, Hwang, Cheol Yong and Youn, Hwa Young (2002). Application of Receiver Operating Characteristic (ROC) Curves for Clinical Diagnostic Tests. *Korean Journal of Veterinary Clinic*, 19(3), 312-315.
- Patrick, C. J. & Iacono, W. G. (1991). Validity of the control question polygraph test: The problem of sampling bias. *Journal of Applied Psychology*, 76, 229-238.
- Podlesny, J. A. & Raskin, D. C. (1978). Effectiveness of techniques and physiological measures in the detection of deception. *Psychophysiology*, 15, 344-358.
- Podlesny, J. A. & Truslow, C. M. (1993). Validity of the expanded-issue(modified general question) Polygraph technique in a simulated distributed-crime-roles context. *Journal of Applied Psychology*, 78, 788-797.
- Raskin, D. C. & Hare, R. D. (1978). Psychopathy and detection of deception in a prison population. *Psychophysiology*, 15, 121-136.
- Raskin, D. C., Kircher, J. C., Horowitz, S. W. & Honts, C. R. (1989). *Recent laboratory and field research on polygraph techniques*. In: J. C. Yuille (ed.) *Credibility Assessment*, (pp.

- 1-24). Deventer, The Netherlands: Kluwer.
- Rovner, L. I., Raskin, D. C. & Kircher, J. C. (1979). Effects of information and practice on detection of deception. *Psychophysiology*, 16, 197-198.
- Spencer, B. D. (2007). Estimating the accuracy of jury verdicts. *Journal of Empirical Legal Studies*, 4, 2, 305-329.
- Swets J. A. (1986). Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *Psychological Bulletin*, 99, 181-198.
- U.S. Congress, Office of Technology Assessment (1983). *Scientific Validity of Polygraph Testing: A Research Review and Evaluation - A Technical Memorandum*. OTA-TM-H-15, November, Washington D.C.: U.S. Government Printing Office.
- 논문투고일 : 2008. 10. 22
1차 심사일 : 2008. 10. 31
게재확정일 : 2008. 11. 20

The cutoff criterion and the accuracy of the polygraph test for crime investigation

Yu Hwa Han

Kwangbai Park

Chungbuk National University

The polygraph test administered by the Korean Prosecutors Office for crime investigations customarily uses the score of -12 as the cutoff point separating the subjects who lie from those who tell the truth. The criterion used by the KPO is different from the one (-13) suggested by Backster (1963) who invented the particular method for lie detection. Based on the signal detection theory applied to the real polygraph test data obtained from real crime suspects by the KPO, the present study identified the score of -8 as an optimal criterion resulting in the highest overall accuracy of the polygraph test. The classification of the subjects with the score of -8 as the criterion resulted in the highest accuracy (83.17%) compared with the accuracies of classifications with the Backster's criterion (76.24%) and the KPO's criterion (80.20%). However, the new criterion was also found to result in more false-positive cases. Based on the results from the present study, it was recommended to use the score of -8 as the criterion when the overall accuracy is important but the score of -12 or -13 when avoiding false-positive is more important than securing the overall accuracy.

Key words : polygraph test, cutoff point, signal detection theory, ROC curve