

연구논문

베이저안 진실 표출법(Bayesian Truth Serum)을 활용한 공무원 윤리행위인식 설문 의 응답 편향성 연구

한승훈* · 이준석**

인간과 사회에 대한 주관적 인식과 의견을 탐구하는 사회과학연구의 설문조사들은, 특히 공무원의 비윤리적 행위 인식에 대한 설문과 같이 응답 노출 시 사회적 비난 가능성이 있을 경우, 설문의 응답자가 본인의 솔직한 의견보다는 사회적으로 바람직한 의견으로 응답하여 실제와 괴리되는 편향성의 우려가 있다. 본 연구에서는 사회적 바람직성을 넘어 응답의 진실성을 유도하는 방법으로, 응답자가 솔직히 응답할 때 가장 보상이 커지도록 경제적 유인을 제공하는 베이저안 진실 표출법의 효용성을 탐색하고자 하였다. 597명의 공무원들을 무작위로 배정한 후, 처치집단에게는 주어진 가상의 윤리적 딜레마 상황에서 비윤리적 행위를 할 확률을 베이저안 진실 표출법 방식으로 설문을 하였고, 통제집단에게는 유인 기제가 없는 일반적인 방식으로 같은 설문을 하였다. 분석결과, 처치집단이 통제집단보다 비윤리적 행위를 할 가능성을 높게 응답한 경우가 존재하며, 특히 4~5급 공무원들의 경우 그러한 경향성이 뚜렷하였다. 이는 설문 응답자의 솔직한 응답을 끌어내는 베이저안 진실 표출법의 효용성을 보여주는 결과라 할 수 있다.

주제어: 공무원 윤리행위인식, 베이저안 진실 표출법, 사회적 바람직성 편향, 무작위 배정, 사고 실험

* 이 논문은 2023년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2023 S1A5A8080663).

** 중앙대학교 공공인재학부 부교수(sehan@cau.ac.kr), 제1저자.

*** 중앙대학교 국가정책연구소 연구교수(stoppingstonez@outlook.com), 교신저자.

I. 논의의 배경

한국에서의 공공부문 부패(corruption) 문제는 학문적 차원을 넘어 오랫동안 국가적 차원의 중요한 문제로 인식되어 왔다(박중훈 외 2017). 그러나 높은 관심에도 불구하고, 공공부문 부패와 관련된 실증적 연구들은 아직도 매우 부족한 실정이다. 무엇보다 정부가 그동안 시행해 온 반부패 정책의 효과 평가는 차치하고서라도, 공공부문에서 부패 및 부정행위가 얼마나 널리 퍼져 있는지 현황을 파악하는 데에만도 어려움을 겪고 있다(박중훈 외 2017).

현재 우리나라의 공공부문 부패 및 부정행위 현황은 주로 매년 시행하는 국제투명성기구(Transparency International)의 ‘부패인식지수(Corruption Perception Index, CPI)’ 설문조사와 한국행정연구원의 ‘정부부문 부패실태에 관한 연구’ 설문조사의 결과들을 이용하여 파악하고 있다. 이 두 조사는 공공부문 종사자에 대한 직접조사가 아닌, 공공부문과 관련된 기업체 종사자 등 제3자의 경험을 바탕으로 하는 간접조사라는 한계가 있다. 그럼에도 이들은 공공부문의 부패 및 부정행위가 감소 추세에 있음을 보여준다. 예컨대, CPI에 따르면 대한민국은 2012년 이후 100점 만점 중 50점대 중반을 기록하며 전 세계 약 180개국 중 반부패 국가 순위 40~50위권에 머물다가, 최근 2020년부터는 60점을 넘어서며 상위권인 30위권 초반으로 올라섰다.¹⁾ 한국행정연구원의 설문조사에서도, ‘지난 1년간 공무원에게 금품/향응/편의 제공’한 개인적 부패 경험이 있는지 묻는 질문에 대해 ‘그렇다’고 대답한 응답자 비율이 2017년에 전체 응답자의 1% 이하로 떨어진 이래, 2020년에는 0.0%를 기록하기도 하였다(오세영 2020).

그러나 이와 달리, 마찬가지로 간접 설문조사임에도 불구하고, 공공부문의 부패 및 부정행위에 대한 일반 대중의 일반적 인식은 매우 부정적이다. 동일한 한국행정연구원의 설문조사에 따르면, 일반 대중인 응답자의 절반 이상은 2010년대 이후에도 꾸준히 우리나라 공공부문의 부패 수준이 심각하다고 응답하고 있다. 예컨대 2021년 조사에서는 ‘그렇다’는 응답자의 비율은 61.9%였는데, 이는 민간부문의 부패가 심각하다고 응답한 비율인 47.9%보다 10% 포인트 이상 높은 수준이었다(최

1) <https://www.index.go.kr/unify/idx-info.do?idxCd=4269>.

혜민 2021). 이러한 공공부문에서의 개인적 부패 경험과 일반적 부패인식 수준 간의 극단적인 격차는, 제3자에 대한 간접 설문조사를 활용하여 부패 현황을 정확히 파악하는 데 있어서의 심각한 한계를 보여준다.

위와 같이 공공부문과 연관된 제3자에게 묻는 간접적인 설문지 문제가 있다면, 공무원 등 공공부문 종사자에게 직접 설문하는 방식을 고려해 볼 수 있다. 그러나 이러한 직접 설문조사는 간접 설문조사보다 더 부정확할 수 있다. 현재 사회과학 분야에서 널리 통용되는 자기기입식(self-report) 직접 설문조사 방식은, 응답자가 본인이 정말로 그렇게 생각하는 응답을 솔직하고 성실하게 기입할 것이라는 매우 큰 가정에 기대고 있다. 그러나 현실에서는 이러한 가정이 성립하지 않는 경우가 많다. 예컨대 응답자들은, 특히 공무원의 위법 또는 비윤리적 행위에 관한 질문과 같이 솔직한 응답의 공개 시 거센 사회적 비난이 예상되는 경우 고의로 응답을 회피하거나, 설사 응답을 한다고 해도 자신의 체면과 사회적 위신을 생각하여 솔직한 답변보다는 사회적으로 바람직한 방향으로 편향(socially desirable bias)된 답변을 택하는 경우가 많다(권희경·이현주 2020; 김승현 2013; 송인점·조성덕 2013). 이렇게 설문 응답자가 솔직하고 성실하게 답변하지 않는 한, 설문조사 결과는 왜곡되어 타당성(validity)과 신뢰성(reliability)이 위협받을 수밖에 없다.

이러한 설문지의 문제점들은 비단 부패 및 부정행위 연구에서만 나타나는 것은 아니다. 인간과 사회에 대한 주관적 인식을 자기기입 방식으로 묻는 설문조사가 여전히 사회과학의 경험적 연구 전반에서 가장 많이 쓰이는 도구인 이상, 그 문제점은 사회과학 연구결과 전반에 걸쳐 있다고 할 수 있다. 그리고 사회과학 설문조사에서 응답자가 솔직하고 성실하게 답변하지 않는 근본적인 문제점은, 응답자가 그렇게 할 유인(incentive)이 없다는, 더 정확히는 응답자가 그렇게 하지 않을 유인이 있기 때문이다(Schoenegger 2023). 예컨대, 자신의 위신과 체면을 생각하면 남들이 보기에 비난하지 않을 답변을 하는 것이 유리하고, 한정된 본인의 시간 자원을 고려하면 보상이 없거나 보상이 있는 경우라도 내용에 대한 검증이 없다면 설문조사는 내용과 상관없이 빨리 끝내는 것이 여전히 유리한 선택이 된다.

따라서 학계에서는 그동안 설문조사와 응답자 간의 이해관계의 불일치에 주목하여, 양자를 일치시킬 수 있는 유인 기제(incentive mechanism)를 만들기 위한 노력을 해왔다(Schoenegger 2023). 이때 유인 기제의 핵심은 응답자가 얼마나 솔직하고 성실하게 응답하였는지 판단하는 기준(criterion)을 정하는 일이다. 그러나 현실에서

적용이 가능한, 보편적인 판단 기준을 설정하는 것은 매우 어려운 일이다. 다른 대안이 없는 상황에서 기존의 많은 연구는 소위 ‘다수의 법칙(majority rule)’을 적용하여, 다른 응답자가 가장 많이 택하는 답안을 기준으로 이에 가까울수록 보상을 주는 방안을 제시하였다(예컨대, Linstone & Turoff 1975;²⁾ Miller et al. 2005). 그러나 이러한 다수의 법칙 기준은, 응답자가 본인의 진실된 의견을 표출하기보다는 다른 사람들의 선호를 유심히 살펴 그에 맞춰 응답하게 하는, 오히려 왜곡된 유인 기제로 작용할 위험이 크다는 비판을 받아 왔다.

이에 다수의 법칙에 기대지 않는 여러 판단 기준이 탐구되었는데, 최근 그중 하나인 ‘베이지안 진실 표출법(Bayesian Truth Serum, BTS)’이 주목받고 있다(Prelec 2004; Prelec et al. 2017).³⁾ BTS는 설문 응답자의 응답이 솔직하고 성실한 응답인지 판단하는 기준을 설정하는 새로운 방법으로서, 뛰어난 한 명의 개인의 판단보다 평범한 여러 개인이 모인 집합적 판단이 더 나올 수 있다는 소위 ‘대중의 지혜(wisdom of crowd)’ 기법의 하나이다. 집단적 판단 기준으로서의 ‘의외의 일반성(surprising commonality)’이라는 개념을 이용하여 인간의 인지적 의사결정 과정상 상대적으로 가장 진실에 가까울 것으로 생각되는 응답에 높은 점수를 부여하고, 그 점수들을 합산하여 응답자들 간의 순위를 산출하며, 그 순위에 따라 설문 응답자에게 차등적으로 보상하는 방식이다. 이렇게 솔직한 응답에 가까울수록 큰 경제적 보상이 주어진다는 유인 기제가 있음을 설문 응답자에게 사전에 이해시키면, 설문 결과는 그러한 유인 기제가 없을 때와 달라질 수 있다. 실제로 BTS 방식의 설문조사 결과가 전통적 방식의 설문조사 결과와 유의미하게 달랐다는 실증연구결과가 마케팅이나 환경평가, 범죄학 등의 분야에서 보고되고 있다.

그러나 저자들이 아는 한, 아직 우리나라에서 이러한 BTS 방식이 학문적으로 소개된 적이 없었다. 이에 본 연구에서는 국내에서 처음으로 BTS 방식을 적용한 설문조사 연구를 진행하고자 한다. 특히 공무원의 부패 및 부정행위에 대한 설문과 같이 설문 응답자가 사회적 비난 가능성을 피하고자 본인의 솔직한 응답보다는 사

2) 이들이 제시한, 행정학에서 많이 사용하는 델파이기법(Delphi Method)도 이러한 다수의 법칙에 기반한 방법으로 볼 수 있다.

3) Truth serum은 전쟁 등 특수한 상황에서 음용 시 강제로 진실을 자백하도록 만드는 자백제를 뜻한다. 물론 여기서 실제 자백제를 사용하지 않으며, 방법론의 명칭에의 truth serum의 표현은 응답자가 보상의 극대화를 위해서 진실한 선호를 표출하는 것이 최선의 방안이 되도록 만든 유인 기제에 대한 은유적 표현이다. 명칭의 번역에 있어 오해의 소지가 있는 직역 대신 이해가 쉬운 의역을 거쳐 ‘베이지안 진실 표출법’으로 칭하기로 한다.

회적으로 바람직한 방향으로 편향된 응답을 할 가능성이 큰 경우에, BTS 방식으로 경제적 유인을 주는 경우 그러한 왜곡 가능성을 극복할 수 있는지 그 가능성을 탐구해 보고자 한다. 다만 공무원의 부패 및 부정행위는 명백한 위법 시 발각되면 무거운 법적 처벌을 받을 위험이 크므로, 공무원들은 약간의 경제적 유인이 있다고 해도 본인의 위법 행위를 설문에서 그대로 노출하기는 매우 꺼릴 것이다. 따라서, 본 연구의 설문에서는 공무원의 실제 부패 및 부정행위를 묻는 대신, 가상의 윤리적 딜레마 상황에서 응답자가 어떠한 행위를 택할 것인지 묻는 사고 실험(thought experiment) 방식으로 공무원의 윤리행위인식(perception on ethical behaviors)을 조사하기로 한다.

본 논문의 구성은 다음과 같다. 먼저 다음 장에서 BTS 방식에 대한 이론적 논의와 실증적 연구를 간단히 소개하고 공무원 부패 및 부정행위와 관련된 선행연구들을 간단히 검토한다. 제3장에서는 본 연구에서 이용한 BTS 방식 설문의 사고 실험 방법에 대하여 설명하며, 이어지는 제4장에서는 분석한 설문 결과를 보고한다. 마지막으로 제5장에서 연구결과의 함의를 언급하고 연구의 한계와 결론을 제시한다.

II. 이론적 배경 및 선행연구의 검토

1. 베이지안 진실 표출법(BTS)의 소개

BTS 방법을 이해하기 위해서는 먼저 Prelec(2004)의 ‘의외의 일반성(surprising commonality)’ 개념에 대한 이해가 필요하다. 이는 다수의 응답자에 의한 집합적 평균 추측치보다 실제 평균 응답치가 더 큰 경우를, 즉 실제 응답이 추측보다 더 일반적인 경우를 의미한다. 예컨대, 응답자 집단이 어떠한 Yes/No 설문 문항에 대하여 Yes로 응답할 비율을 평균 60%로 추측하였으나 실제로는 그보다 높은 응답자의 70%가 Yes로 응답하였을 때, Yes 응답은 ‘의외로 일반적인(surprisingly common)’ 응답이 된다. 반면, 설문 응답자 전체의 집합적 추측치(40%)보다 실제 응답치(30%)가 더 적은 No의 응답은 ‘의외로 일반적이지 않은(surprisingly uncommon)’ 응답이 된다. 이때 중요한 것은 특정 응답이 갖는 실제 응답치 평균의 추측치 평균 대비 상대적 크기이지, 한 특정 응답이 다수 또는 소수에 속하는지를 결정하는 절

대적 수치(즉, 50% 이상인지 아니면 이하인지)가 아니다. 즉, 특정 응답이 소수 의견이라도 의외의 일반성을 갖는 데에 아무런 문제가 없다. 따라서, 응답자가 어떠한 솔직한 특정 응답을 하여도, 그 응답은 ‘의외로 일반적’일 수 있는 가장 최적의 기회를 얻게 된다(Prelec 2004).

이러한 의외의 일반성은 다른 정보가 부재한 상황에서 사람들은 일반적으로 다른 사람들 또한 본인과 동일한 견해를 가지고 있을 것이라 여기는 경향이 있다는 사회심리학의 ‘허위 합의 효과(false consensus effect)’와 관련이 있다(Schoenegger, 2023). 허위 합의 효과에 따르면, 솔직하게 특정 응답을 한 응답자는 다른 응답을 한 응답자들보다 본인 응답의 대중성을 높게 평가하는 경향을 보이는데, 이는 특정 응답을 택한 본인 자체가 정보를 가지고 있는 표본(informative sample)이기 때문이다. 반면, 그러한 정보를 갖지 못한 채 타 응답을 한 응답자들은 그 특정 응답의 대중성을 높게 평가할 이유가 없으므로, 집단 전체적으로는 그 특정 응답의 진정한 대중성은 과소평가(underestimation) 될 가능성이 크다. 즉, 특정 응답에 대한 응답자의 높은 대중성 평가는, 기존의 특정 응답의 응답률이 낮을 것이라는 일반적 견해에 본인이 그렇게 생각한다는 새로운 경험 정보가 더해져 업데이트된 베이지안 추론(Bayesian reasoning) 결과이다. 그러나 비록 응답자가 본인의 솔직한 특정 응답의 대중성이 실제로는 높을 것으로 보고 이를 고수하지만, 타 응답자들은 그렇게 생각하지 않을 것임을 알기에 응답자의 특정 응답에 동의하는 타 응답자의 비율을 낮게 추측하게 된다. 따라서, 응답자 본인도, 타 응답자도 특정 응답의 동의 여부를 낮게 추측하여 그 특정 응답의 집합적 평균 추측치는 낮을 수밖에 없다. 그런데, 특정 응답의 실제 응답치가 평균 추측치보다 커서 의외로 일반적이 된다는 것은, 타 응답자의 낮은 동의 비율이 예상됨에도 불구하고 본인의 특정 응답을 고수한 사람들이 많다는 뜻이다. 만일 본인의 진실된 선호가 아닌 응답이라면 다수의 의견과 배치되는 이러한 응답을 굳이 고수하지 않았을 것을 고려하여, BTS에서는 이러한 의외로 일반적인 응답은 대중의 지혜 기법에 의해 솔직한 응답이라 집합적으로 판단한다.

BTS와 같은 설문 의 유인 기제 설정과 관련된 방법론에서의 핵심은, 유인 기제에 직면한 설문 대상이 실제로 솔직히 반응하는지를 연구자가 어떻게 아느냐에 있다. Prelec(2004)은 구체적인 판단 기준으로서 (1) ‘의외의 일반성’에 기반한 특정 응답에 부여되는 정보 점수(information score)와 (2) 타 응답자의 응답률에 대한 추측의 정확도에 따른 점수(prediction accuracy score)를 계량화하여 이들 점수의 합에 따라 보상 여부를 결정할 것을 제안하고 있다. 이러한 정보 점수와 추측치 정확도 점

수의 구체적인 수식은 논문 끝에 추가로 첨부한 <부록 1>을 참조한다.

BTS 방식에서 의외의 일반성은 응답자 전체에 의하여 집합적으로 결정되는 것이기에, 응답자 개인의 독자적인 전략적 행동에 따라 쉽게 바뀌지 않는다는 장점이 있다. 즉, 설문 응답자가 추가적인 보상을 받기 위하여 높은 수준의 의외로 일반적인 특정 응답을 택하고자 하여도, 응답자 전체의 응답 비율과 평균 추측치들에 대한 완비된 정보(complete information) 없이는 어떠한 개인의 전략적 행동도 높은 수준의 의외로 일반적인 응답을 인위적으로 만들어 내기 어려우며, 오히려 그러한 행동은 의외로 일반적이지 않은 응답으로 이어져 보상받지 못할 위험을 키울 수 있다. 따라서 개인 수준에서 높은 수준의 ‘의외로 일반적’인 응답을 만들 최선의 전략은 응답자 본인의 의견을 솔직하게 표출하는 것이 된다.⁴⁾ 응답자가 본인의 응답이 다수의 응답에 속하든 소수의 응답에 속하든 상관없이 진실한 선호를 표출하는 것이 게임이론(game theory)에서 말하는, 불완비 정보 게임(incomplete information game) 하에서의 베이지안 내쉬 균형(Bayesian Nash equilibrium)이 된다. 이처럼 응답자가 솔직히 응답하는 전략을 보상이 가장 크게 되는 최선의 전략으로 만들어 설문조사와 설문 응답자의 이해관계를 일치시키는 유인구조를 만들어 낸다는 점이, BTS 방식이 다수의 법칙 등의 기존 방식과 구별되는 가장 큰 특징이다.

다만 베이지안 진실 표출법도 설문의 응답자가 진실한 선호를 표출하도록 유인하기 위해서는 두 가지 기본적인 가정이 필요하다(Prelec 2004). 첫 번째는 충분히 큰 표본의 크기 즉, 설문의 응답자가 충분히 많은 수가 있어야 한다. 만일 설문의 응답자가 충분히 많지 않으면 개별 응답자의 특정 응답에 따라 응답자 전체의 평균 추측치가 바뀔 수 있어, 응답자들은 솔직하지 않은 응답으로 본인이 받을 수 있는 보상을 인위적으로 최대화하는 전략적 행동을 할 수 있다. 두 번째는 응답자들은 자신과 동일한 선호를 가지고 있는 다른 응답자들도 응답자 전체의 응답률에 대하여 비슷하게 추측할 것이라고 믿어야 한다. 즉, 응답자들은 자신의 응답 선호를 비개인적인, 즉, 모든 사람에게 공통적으로 적용 가능한 정보 신호(impersonally informative signal)로 취급하여야 한다.

이러한 기준과 보상 방법을 이용하여, 일군의 사회과학자들은 그동안 현실에서 관측이 어려워 연구가 어려웠던 문제들 또는 주관적이어서 객관적인 현황치가 없는 사회과학적 문제들 - 예컨대, 대중들의 특정 상품에 대한 진실한 선호도, 대학생 또

4) 이에 대한 수학적 증명은 Prelec et al. (2017: Supplementary Information)을 참조한다.

는 연구자들의 학문적 부정행위, 전문가들의 에너지 자원에 대한 미래 예측 등 - 을 실증적으로 탐구하기 시작하였다(Barrage & Lee 2010; de-Magistris & Pascucci 2014; Frank et al. 2017; Howie et al. 2011; John et al. 2012; Loughran et al. 2014; Weaver & Prelec 2013; Zhou et al. 2019). 그중 하나인 Frank et al.(2017)은 BTS 방식이 실제로 솔직한 응답을 이끌어내는 효과가 있는지를 실증적으로 검증했다는 점에서 눈여겨볼 필요가 있다. 이들은 참가자들이 무작위로 동전 또는 주사위를 던져, 동전이 앞면일수록 또는 주사위 숫자가 클수록 큰 보상을 받을 수 있는 실험을 설계하였다. 이때 참가자들은 본인의 실험 결과를 설문 형식으로 스스로 보고 하는데, 보상을 높게 받기 위하여 실제 본인의 실험 결과에 상관없이 임의의 결과를 설문에 기입할 수 있었고 또한 응답이 거짓이라 하더라도 추가적인 사후 검증 절차를 통한 벌칙을 부여하지 않았기 때문에, 참가자들은 응답을 왜곡하여 보고할 유인이 컸다. 그럼에도 불구하고, BTS 방식을 이용한 설문 결과는 그렇지 않은 설문의 결과보다 참가자들이 실험 결과를 왜곡하여 보고한 비율이 낮았고, 따라서 BTS 방식이 응답자의 솔직한 응답률을 통계적으로 유의미하게 높일 수 있다는 결론을 내렸다. Frank et al.(2017) 외에도, 위의 여러 실증 연구들에서 BTS 방식을 이용한 설문 결과는 전통적인 설문 결과와는 유의미하게 다른 양상을 보였다.

2. 공공부문의 부패, 부정행위, 윤리행위에 대한 선행연구

전술한 바와 같이, 공공부문의 부패와 부정행위가 얼마나 심각하냐는 사회과학으로서의 행정학의 중요한 주제이기는 하나, 그 현황을 파악하기 매우 어렵고 실제로 실증적 연구도 부족한 편이다. 기존의 실증적 연구들은 현황 파악을 위해 대체로 다음 세 가지 방식의 자료수집 방법들을 사용하고 있다. 먼저 주관성을 배제하고 객관적인 수치를 얻기 위하여, 공무원 범죄의 사법처리 결과 등과 같은 행정자료를 이용할 수 있다. 그러나 이러한 행정자료는 적발되지 않고 숨겨진 부패 및 부정행위가 많을 수 있어 정확성의 문제가 있으며, 또 사법처리 결과는 부패의 현황 자체를 나타내기보다는 정부의 반부패 활동 수사 의지를 반영하는 측면이 크다는 점에서 (박중훈 외 2017), 부패 및 부정행위의 실증 경험적 연구자료로는 잘 쓰이지 않는다.

두 번째로는 현재의 CPI 조사와 ‘정부부문 부패 실태에 관한 연구’와 같이, 제3자에게 공공부문의 부패 및 부정행위 평가를 요구하는 간접 설문조사 결과들을 이

용한 연구들이 있다. 이 방식에서는 일반 시민 또는 민간 사업체 종사자가 공무원의 부패 및 부정행위와 이해관계가 없는 제3자일 것을 가정하여, 이들이 직·간접적으로 경험한 바를 바탕으로 설문지에 진실하게 응답할 것이라고 기대한다. 그러나 간접조사 방식은 기대와 달리 여러 문제점을 노출하고 있다. 먼저, 가정과 달리, 공무원의 부패 및 부정행위에 있어서 관련 민간 사업자는 제3자가 아닌 부패 당사자로서 공범이라는 이해관계를 갖는 경우가 많다. 그렇다면 민간 사업자도 수사가 확대되면 자신도 처벌받을 위험이 있기에, 설문조사에서 솔직하게 응답할 유인이 없고 오히려 감추고 왜곡된 응답을 할 유인이 크다. 두 번째 한계는, 설사 이들이 본인이 아닌 타인이 연루된 간접 경험에 대해 솔직히 응답한다고 해도, 부패 및 부정행위가 은밀히 내부자끼리만 공유된다는 점을 고려하면 이러한 간접 경험의 범위와 정확성은 매우 제한적이라는 한계가 크다. 아울러 최근에는 간접 설문방식의 또 다른 심각한 문제로서, 공공부문 부패 및 부정행위에 대한 인식의 편향 가능성이 주목받고 있다(강정석·고재권 2016; 고길곤·조수연 2012; 고길곤 외 2019; 박중훈 외 2017; 장용진 2015). 즉, 일반 시민들의 인식 형성에 있어, 실제 공공부문의 부패 및 부정행위 외의 다른 요인들이 매우 큰 영향을 미친다는 것이다. 예컨대, 일반 시민들은 영화나 드라마와 같은 미디어의 영향으로 공공부문에는 수많은 부패 및 부정행위가 만연해 있을 것으로 지레짐작하여 편견을 갖거나 ('반향실 효과'), 부패 및 부정행위에 대한 정의(definition)에 있어서 특정 행위를 비위행위가 아닌 관행으로 보기도 하고, 또한 본인이 현 정부에 대한 이념적 신뢰 정도에 따라 부패인식이 달라지기도 한다.

따라서 이들 행정자료와 간접조사의 한계를 고려하면, 최선의 방법은 공공부문 종사자인 공무원을 대상으로 하는 직접 설문조사 방식일 수 있다. 그러나 전술한 바와 같이, 이 경우 설문 응답의 진실성 문제가 매우 크게 불거지므로 많은 연구에서는 이러한 직접 설문방식을 선호하지 않는다. 그럼에도 불구하고, 행정학에서의 몇몇 실증 연구들은 공무원에게 직접 설문하는 방식으로 공무원의 비윤리적 행위와 관련된 데이터를 모아 연구하기도 하였다. 예컨대 박현환(2019)은 중앙 및 지방행정기관 공무원 1,129명을 대상으로 공직봉사 동기와 회색부패에 대한 인식을 설문조사하여 양자의 관계를 탐색하였으며, 박천오(2014)는 경기도 6개 기초지방자치단체 소속 공무원 470명을 대상으로 자치단체장의 부당한 정책추진과 관련된 윤리의식에 대하여 직접 설문조사를 실시하였다.

3. 본 연구의 방향: BTS 방식을 이용한 공무원 직접 설문조사

종합하면, 공공부문의 부패 및 부정행위의 실태를 파악하기 위해 현재 쓰이고 있는 세 가지 방식인 행정자료, 간접 설문조사, 직접 설문조사들은 모두 한계가 있다. 특히 응답자에게 공무원의 부패 및 부정행위와 관련하여 직·간접적인 경험과 인식을 묻는 설문조사의 경우, 현행 방식에서는 응답자가 솔직히 응답할 것이라는 막연한 기대와 가정 외에 이를 끌어내기 위한 어떠한 유인도 없어, 설문조사 결과에 대한 타당성과 신뢰성에 대한 문제 제기가 계속되고 있다(박중훈 외 2017).

이에 본 연구에서는, 직접 설문방식으로 공무원에게 부패 및 부정행위와 관련하여 묻되, BTS 방식을 도입하여 이들이 솔직하게 응답할 경제적 유인을 주는 경우 설문 결과가 달라지는지 확인하고자 하였다. 다만, 부패 및 부정행위는 대체로 명백한 위법이라 누설 시 파면과 해임, 심지어 구속 등의 처벌로 이어질 가능성이 크므로, 공무원들은 자신들의 실제 행위가 위법한 부패 및 부정행위에 해당한다면 아무리 경제적 유인이 있다고 하더라도 이를 솔직하게 응답할 가능성은 작다. 따라서 본 연구에서는 공무원들의 과거와 현재의 실제 부패 및 부정행위를 묻는 대신, 가상의 윤리적 딜레마(ethical dilemma) 상황에서 어떠한 행위를 택할지를 살피는 사고 실험을 행하여 공무원들의 윤리행위에 대한 일반적 인식을 묻기로 하였다.

공무원의 비윤리적 행위들은 사회적 비난 가능성이 있으므로, 다른 유인이 없는 경우 공무원들은 그러한 비윤리적 행위를 할 가능성을 본인의 진실한 생각보다 낮게 보고하는 하향 편향(downward bias)의 가능성이 크다. 따라서, 만일 BTS에서의 경제적 유인이 솔직한 응답을 끌어낸다면, 공무원들의 응답은 유인이 없을 때보다 비윤리적 행위를 할 가능성이 크게 나올 것이다. 종합하면, 본 연구에서는 BTS 방식의 효용성에 관한 연구로서, “BTS 방식의 설문에서 공무원들이 응답한 비윤리적 행위를 할 가능성은 기존 방식의 설문에서 응답한 비윤리적 행위 가능성보다 클 것이다”라는 가설을 실증적으로 검증해 보고자 한다.

Ⅲ. 연구방법

1. 설문자료의 수집

본 연구를 위한 설문자료의 수집은 한 전문 설문업체의 도움을 받아 이뤄졌다. 업체는 자신들이 보유한 설문 패널 집단(panel pool)의 43,810명 공무원 중 1,000명을 무작위로 선정(random sampling)하여 비식별화(de-identified) 과정을 거친 후 잠재적 설문 대상자 명단을 제공하였다. 단, 업체의 설문 패널 참여자 모집이 광고나 기존 회원의 추천 등을 통한 상시 모집 방식으로 이루어졌으므로, 본 연구자료의 표본추출방식은 편의 표본추출(convenience sampling)이라고 할 수 있다.⁵⁾ 연구자는 이들을 난수발생기(random number generator)를 이용한 무작위 배정(random assignment)을 통해 처치집단(treatment group) 500명과 통제집단(control group) 500명으로 나눈 후, 다시 별개로 작성된 두 집단의 명단을 업체에 주어 설문을 진행하게 하였다.

설문은 비용과 시간상의 편의를 위해 온라인 설문방식을 채용했다. 다만 배정에 따라 나뉜 두 집단이 동일한 설문질문들에 대하여 다른 설문방식으로 설문에 응답할 수 있도록, 온라인 설문 플랫폼을 집단별로 따로 구성하였다. 두 설문 플랫폼에서의 가장 큰 차이는 ‘참여에 대한 보상 안내’ 부분이었다. 설문의 잠재적 응답자가 설문에 적극 참여할 동기를 부여하는 것이 중요한 본 연구의 특성을 고려하여, 기본적으로 모든 응답자에 대하여 설문을 끝까지 완료 시 무조건 1만 원을 지급하고 추가로 완료한 응답자의 1/3에 대하여 2만 원의 보너스를 주는 것으로 보상 구조를 설계하였다. 그러나 2만 원의 추가 보너스를 누구에게 지급하느냐에 있어서, 통제 집단에게는 특별한 다른 언급 없이 무작위로 선정(random selection)하여 지급할 것으로 간략히 안내한 반면,⁶⁾ 처치집단은 베이지안 진실 표출법(BTS)을 소개하고 응답에 따라 계산된 BTS 점수의 상위 득점자 순으로 보너스가 제공될 것이며 이때 BTS 점수를 높일 최선의 전략은 솔직하게 응답하는 것임을 안내하였다.⁷⁾ 즉, 처치

5) 본 연구의 원래 계획은 할당표본추출(quota sampling)을 통해 새로이 공무원 전체 모집단에 대한 대표성을 높인 표본을 수집하려 했으나, 예산의 한계로 인하여 새로이 설문 응답자를 찾지 않고 업체의 기존 설문 패널들을 이용하는 대신, 잠재적 설문 응답자를 최대 1,000명까지 늘리기로 하였다.

6) “본 설문에 참여하는 응답자들은 모든 설문을 빠짐없이 성실히 응답 시 참여에 대한 보상으로 10,000원(1만 원)의 보상을 무조건 받게 됩니다. 아울러, 설문에 성실히 참여한 전체 응답자 중 무작위로 선정된 33%는 추가적으로 20,000원(2만 원)의 보상을 받게 됩니다.”

7) “본 설문에 참여하는 응답자들은 모든 설문을 빠짐없이 성실히 응답 시 참여에 대한 보상으로 10,000원(1만 원)의 보상을 무조건 받게 됩니다. 아울러, 설문에 대한 응답을 ‘베이지안 진실 표출법(Bayesian Truth Serum, 줄여서 BTS: 아래 설명 참조)’이란 방법으로 평가하여, 가장 높은 BTS 점수를 받은 상위 33% 응답자들은 추가로 20,000원(2만 원)의 보상을 받게 됩니다. ... 설문 응답자는 이러한 알고리즘에 따라 계산된 (1) 본인의 답변에 대한 점수와 (2) 다른 참여자들이 어떻게 응답할 것인지 예측한 것에 대한 점수를 합한, 최종 BTS 점수를 받게 됩니다. 이 최종 BTS

집단의 응답은 새로운 BTS 방식의 설문 결과를, 통제집단의 응답은 기존 방식의 설문 결과를 나타낸다.

실제 온라인 설문조사는 2023년 12월 21일부터 27일까지 일주일 동안 진행되었다. 첫날에 1,000명에게 온라인 설문 플랫폼 링크가 포함된 이메일이 발송되었고, 이들 중 총 634명(처치집단 312명, 통제집단 322명)이 기한 내 응답을 완료하였다. 이들 중 설문에 응답은 했으나 정보 활용에 동의하지 않은 18명(처치집단 6명, 통제집단 12명)과 현재 중앙정부 부처와 지방자치단체 외의 공공기관(교사나 경찰, 공기업 등)에 재직 중인 19명(처치집단 13명, 통제집단 6명)을 제외한, 총 597명(처치집단 293명, 통제집단 304명)이 본 연구의 최종 분석표본(analysis sample)이 되었다.

2. 변수의 설정

처치집단과 통제집단에게 보상에 대한 안내 외에는 동일한 설문이 주어졌으며, 설문 문항은 크게 세 부분으로 구성되었다. 첫 번째 부분은 승진 및 보상 만족도나 직무만족도 등 공무원의 공직 실태를 파악하기 위한 설문이나, 이들은 본 연구의 주 관심사가 아니므로 여기서는 설명을 생략하기로 한다. 두 번째는 본 연구의 주 관심 사항인 공무원들의 윤리행위인식에 대한 설문이었고, 마지막은 공무원들 개인의 인적 특성들에 대한 설문이었다.

공무원들의 윤리행위인식과 관련된 질문들은, 전·현직 공무원들과의 인터뷰를 바탕으로 하여, 공무원들이 업무와 관련하여 실제로 경험할 가능성이 큰 총 5개의 윤리적 딜레마 상황을 나타내도록 구성하였다. 이때 공무원의 비윤리적 행위가 반드시 개인적인 이익만을 위한 것은 아니며, 개인이 속한 부처 등 비개인적인 이익을 위해서도 행해질 수 있다는 점을 고려하여(박현환 2019), 4개의 상황에서는 동일한 비윤리적 행위라 하더라도 전자를 ‘사적 목적’의 행위로, 후자를 ‘공적 목적’의 행위로 구분하여 두 개의 별개의 질문을 하였다. 이에 따라, 공무원 윤리행위인식의 10개 질문은 ‘공적 목적을 위한 승진청탁(Q1)’과 ‘사적 목적을 위한 승진청탁(Q2)’, ‘공적 목적을 위한 수의계약 남용(Q3)’과 ‘사적 목적을 위한 수의계약 남용(Q4)’, ‘자의적 규정해석(Q5)’, ‘긴급입찰 남용(Q6)’, ‘공적 목적을 위한 인허가 남

점수를 높이는 가장 최선의 전략은, 응답자가 질문에 대하여 실제로 본인이 생각하는 바를 가장 솔직하게 대답하는 것이며, 또한 다른 응답자들이 어떻게 응답할 것인지 신중히 고려하여 대답하는 것입니다.”

용(Q7)'과 '사적 목적을 위한 인허가 남용(Q8)', '공적 목적을 위한 (타인의) 비위행위 목인(Q9)'과 '공적 목적을 위한 (타인의) 비위행위 목인(Q10)'에 대한 것으로 작성하였다. 각각의 구체적인 질문 문장들은 <부록 2>를 참조 가능하다.

설문은 먼저 가상의 시나리오를 주고 그러한 상황에서 설문 응답자가 각각의 비윤리적 행위를 할 가능성이 얼마인지를 확률을 묻는 방식으로 구성하였다. 이때 BTS 점수 계산을 위하여, 본인이 그러한 비윤리적 행위를 할 가능성뿐만 아니라, 본인 외의 타인이라면 그러한 행위를 할 가능성도 함께 질문하였다. 질문에 대한 응답은 '0%'(=1)부터 '100%'(=11)까지의 11단계의 리커트 척도(Likert Scale)로 측정하였다.⁸⁾ 척도 간의 간격은 동일하게 10%를 나타내어 각각의 질문에 대한 응답 변수들은 등간(interval) 변수로 볼 수 있으므로, 이들을 따로 변환할 필요 없이 0부터 11까지의 값을 갖는 양적 변수 그대로 사용하기로 하였다. 다만 '공적 목적을 위한 (타인의) 비위행위 목인(Q9)'과 '사적 목적을 위한 (타인의) 비위행위 목인(Q10)'은 원래 설문에서 목인이 아닌 신고를 할 가능성이 얼마인지를 물었었기에, 이들의 응답 변수는 신고할 가능성이 100%이었다면 '목인' 가능성 변수값은 1로, 신고할 가능성 0%는 '목인' 가능성 변수값은 11로 역 코딩(reverse coding)하여 구성하였다. 따라서 모든 질문에 대한 응답 변수들은, 설문 대상자가 응답에서 더 높은 척도를 선택할수록 관련된 비윤리적 행위를 할 가능성이 더 큰 것으로 해석할 수 있다.

아울러 공무원들의 인적 특성들을 질문하여 이들의 응답을 별개로 변수화하였다. 포함된 인적 특성 변수들은 '성별'(남성=0, 여성=1의 이항변수), '연령'(연속변수), '교육수준'(고졸 이하, 전문대졸, 대졸, 대학원졸의 범주형 변수), '근속기간'(연속변수), '채용방식'(경력경쟁채용/개방형직위=0, 공개경쟁채용=1의 이항변수), '근무조직'(중앙정부 부처=0, 지방자치단체=1), '직급'(1~3급, 4~5급, 6~9급의 범주형 변수)이다. 처치집단과 통제집단이 무작위 배정된 만큼, 이론적으로 인적 특성 변수들은 두 집단 간에 통계적으로 유의미한 차이가 나타나지 않아야 한다. 그러나 <표 1>에서 보는 바와 같이, 카이제곱 검정과 이분산 가정하 t검정을 시행한 결과 대부분의 인적 특성 변수들은 두 집단 간에 차이가 통계적으로 유의하지 않았으나, '근속기간'과 '채용방식'의 두 변수에 대하여는 유의미한 차이가 존재하였다. 처치집단

8) 아울러 BTS 점수의 간소화를 위해, 11단계의 응답들은 해당 비윤리적 행위를 할 것이냐에 대한 'No'(응답이 0%(=1)~50%(=6)인 경우)와 'Yes'(60%(=7)~100%(=11))의 이항변수로 재구성되어 기록되었다.

에 비하여 통제집단 공무원들의 근속기간이 평균 약 1.6년 정도 길었으며, 또한 공개 채용으로 입직한 비율이 약 9.3%p 정도 더 높았다.

<표 1> 기술통계: 인적 특성 변수

	처치(BTS)집단	통제집단	차이 (=처치-통제)
성별(여성=1)	0.269 [0.445]	0.260 [0.439]	+0.010 (0.036)
나이(단위: 년)	50.567 [7.303]	50.891 [7.146]	-0.325 (0.592)
교육	고졸 이하 30명	0.102, 30명	$\chi^2(3)=2.044,$ $p=0.563$
	전문대 졸 55명	0.188, 55명	
	대학 졸업 143명	0.539, 164명	
	대학원 졸업 65명	0.222, 65명	
근속기간(단위: 년)	21.850 [10.470]	23.487 [9.834]	-1.637** (0.832)
채용방식(공개경쟁채용=1)	0.703 [0.458]	0.796 [0.404]	-0.093*** (0.035)
근무조직(지자체=1)	0.710 [0.455]	0.704 [0.457]	+0.006 (0.037)
직급	1~3급 22명	0.078, 22명	$\chi^2(2)=0.797,$ $p=0.671$
	4~5급 92명	0.314, 92명	
	6~9급 179명	0.611, 179명	
총 표본 수	$N=293$	$N=304$	$N=597$

참고: 꺾쇠 안의 숫자는 표준편차(standard deviation), 괄호 안의 숫자는 표준오차(standard error)를 의미함. 차이의 검정은 범주형 변수(교육, 직급)는 카이제곱 검정을, 연속 및 이항변수는 이분산 가정하의 t 검정을 이용함. 통계적 유의성은 * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

3. 분석방법

처치집단과 통제집단은 무작위로 배정되었으므로, 이론적으로는 관측 여부와 관계없이 모든 잠재적인 혼동요인(confounding factor)들이 두 집단 간에 비슷하게 균형화(balanced)되었을 것으로 볼 수 있다. 이 경우 분석에서 별개의 통제변수를 고려할 필요가 없으며, 단순 t 검정(t -test)만으로도 처치집단과 통제집단 간의 윤리행위인식 변수 응답 차이에 대한 분석이 가능하다. <표 1>에서 본 바와 같이, 관측 가능한 인적 특성 변수 중에서도 ‘근속기간’과 ‘채용방식’에 대하여 두 집단 간 균형화가 불완전한 상태이므로, 단순 t 검정 시 이 변수들이 분석결과를 왜곡할 가능성이 있다. 따라서 분석에서는 무작위 배정 실험이라 하더라도 이러한 변수들을 명시적으로 통제변수(control variable)로 넣어 내적 타당성(internal validity)을 높일 필요가 있으며, 이에 통제변수가 포함된 다중회귀분석(multivariate regression)을 추가로 시행하였다. 이러한 통제변수가 추가된 다중회귀분석 접근은 연구결과의 내적 타당성(internal validity)을 높이는 장점이 있을 뿐 아니라, 아울러 표준오차를 줄여 추정(estimation)에서의 정확성(precision)을 높인다는 추가적인 장점이 있다.

다중회귀분석 모형으로는, 종속변수인 윤리행위인식 변수값이 1부터 11까지의 등간 척도로 측정되어 연속된 수치(continuous number)로 해석이 가능하다는 특성을 고려, 최소자승회귀분석(ordinary least square regression, OLS)을 사용하였다. 이때 회귀분석에서의 주요 독립변수는 처치집단 여부(통제변수=0, 처치집단=1)이며, ‘근속기간’과 ‘채용방식’ 즉, ‘공개채용여부’는 ‘통제변수’로 고려되었다. 그리고 잔차(residual)의 이분산성(heteroskedasticity)을 보정하기 위하여 강건한 표준오차(robust standard error)를 사용하였으며, 변수들의 다중공선성(multicollinearity) 존재 여부를 확인하기 위하여 계산한 분산팽창계수(variance inflation coefficient, VIF)는 모두 3.5 이하로 낮게 나왔다.⁹⁾

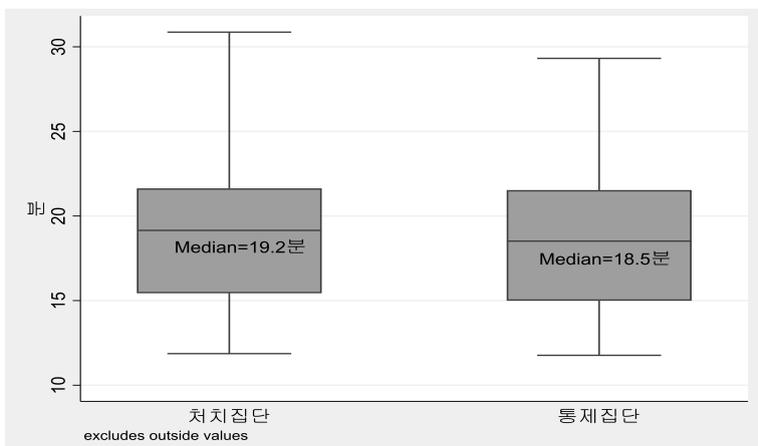
9) 분석방법의 설명에서 다시 상기할 중요한 점은, t 검정과 회귀분석 등 모든 실제 분석에서 BTS 점수는 사용되지 않는다는 것이다. BTS 점수는 분석과는 별개로 추가적 보상을 누구에게 수여하느냐와만 연관이 있다.

IV. 결과

1. 응답시간 분석

만일 BTS 방법이 설문 응답자의 신중하고 솔직한 응답을 유도했다면, 처치집단은 통제집단에 비하여 응답시간이 길 것으로 예상할 수 있다. 그러나 설문 시작 직후에 23시간 이상 방치한 이후 다시 설문을 재시작하는 등 몇몇 응답자의 설문 응답시간이 비정상적으로 길었다. 이러한 이상치(outlier)들로 인하여 평균(mean)이 왜곡되는 점을 고려, 평균 대신 중위값(median)을 이용하여 응답시간을 비교하였다.

실제로 처치집단과 통제집단은 응답시간에 있어서 약간의 차이를 보였다. <그림 1>은 두 집단의 응답시간을 나타낸 상자수염그림(Box-Whisker Plot)이다. 처치집단의 응답시간 중위값은 약 19.2분으로 통제집단의 18.5분보다 약간 길었다. 중위값 외에도 25%(15.4분 vs. 15.0분)나 75%(21.6분 vs. 21.5분) 백분위값 등 다른 지표들도 처치집단이 통제집단보다 약간 길었다. 그러나 이러한 차이들은 통계적으로 유의미하지는 않았다. 예컨대 분위 회귀분석(quantile regression)으로 분석 시, 처치집단과 통제집단 간의 중위값 차이(약 0.6분)는 5% 수준에서 통계적으로 유의하지 않았다 ($p=0.288$).



<그림 1> 상자수염그림(Box-Whisker Plot): 처치-통제집단 간 응답시간 비교

참고: 이상치(Outlier)는 그림에서 제외함.

2. 윤리행위인식 분석: 전체집단 분석

<표 2>는 10개의 주어진 윤리적 딜레마 상황에서 설문 응답자 본인이 비윤리적 행동을 취할 가능성을 11개의 리커트 척도로 측정한 결과를 처치집단과 통제집단을 나누어 나타내고 있다. 이러한 비윤리적 행위들에 대한 응답은 사회적 바람직성을 고려하여 그러한 행위를 할 가능성을 축소하여 보고하는 하향 편향의 유인이 있으므로, 만일 BTS 방법이 설문 응답자의 솔직한 응답을 유도할 수 있었다면, 처치집단이 스스로를 평가한 비윤리적 행위 가능성은 통제집단의 응답보다 높게 나타날 것으로 예상할 수 있다.

그러나 예상과 달리, <표 2>의 두 집단 간 응답을 단순 t 검정으로 비교한 결과에 따르면 두 집단 간의 가능성 정도의 차이는 대체로 뚜렷이 나타나지 않았다. ‘승진 청탁(Q1, 2)’이나 ‘수의계약 남용(Q3, 4)’, ‘인허가 남용(Q7, 8)’은 두 집단의 차이가 거의 0에 가까웠으며, ‘자의적 규정해석(Q5)’과 ‘긴급입찰 남용(Q6)’은, 비록 통계적으로 유의미하지 않았지만, 처치집단의 응답 평균이 통제집단보다 오히려 낮았다.

그러나 ‘(타인의) 비위행위 묵인(Q9, 10)’에 대하여는 예상과 같이 처치집단의 응답 평균이 통제집단의 응답 평균보다 높았다. 특히 마지막 질문인 ‘사적 목적을 위한 비위행위 묵인(Q10)’에 대한 처치집단의 응답은 통제집단 응답보다 0.573만큼 높았고, 이 차이는 5% 수준에서 통계적으로 유의미하였다. 윤리행위인식 변수의 측정에서 각 단계의 간격, 즉 1만큼의 차이가 10%의 가능성을 의미한다는 점을 고려하면, 그 차이는 처치집단이 사적 목적을 위한 비위행위를 묵인할 가능성이 통제집단보다 약 5.7%p 정도 높다는 것으로 해석할 수 있다.

그러나 <표 1>에서 본 바와 같이, 비록 처치집단과 통제집단은 무작위 배정되어 처치 여부 외에 다른 모든 인적 특성들이 비슷했어야 하지만 실제로는 ‘근속기간’과 ‘채용방식’에 있어서 두 집단이 통계적으로 유의미하게 차이가 있었다. 따라서 <표 2>에서의 결과, 특히 ‘사적 목적을 위한 비위행위 묵인(Q10)’의 통계적으로 유의한 차이는, 처치 여부가 아니라 ‘근속기간’과 ‘채용방식’ 등 무작위 배정으로도 균형이 맞춰지지 않은 잔여 인적 특성의 차이에 기인했을 수 있다. 이러한 우려를 검증하기 위하여, 대안적 설명인 두 변수를 통제변수로 넣은 다중회귀분석을 실시하였다.

<표 2> *t*검정: 윤리행위인식의 집단 간 평균 비교

	처치(BTS)집단	통제집단	차이(=처치-통제)
Q1. 승진청탁: 공적 목적	3.522 [2.942]	3.428 [3.054]	+0.095 (0.245)
Q2. 승진청탁: 사적 목적	3.102 [2.720]	3.013 [2.962]	+0.089 (0.233)
Q3. 수의계약 남용: 공적 목적	2.993 [2.976]	2.898 [2.884]	+0.095 (0.240)
Q4. 수의계약 남용: 사적 목적	3.321 [2.943]	3.237 [2.934]	+0.084 (0.241)
Q5. 자의적 규정해석	2.666 [2.565]	2.776 [2.809]	-0.111 (0.220)
Q6. 긴급입찰 남용	2.689 [2.676]	2.911 [2.860]	-0.222 (0.227)
Q7. 인허가 남용: 공적 목적	2.918 [2.592]	2.974 [2.819]	-0.056 (0.222)
Q8. 인허가 남용: 사적 목적	3.164 [2.598]	3.178 [2.855]	-0.014 (0.223)
Q9. 비위행위 묵인: 공적 목적	5.771 [2.984]	5.447 [2.971]	+0.324 (0.244)
Q10. 비위행위 묵인: 사적 목적	5.843 [2.774]	5.270 [2.987]	+0.573** (0.236)
총 표본 수	<i>N</i> =293	<i>N</i> =304	<i>N</i> =597

참고: 꺾쇠 안의 숫자는 표준편차(standard deviation), 괄호 안의 숫자는 표준오차(standard error)를 의미함. 각 윤리행위인식 변수는 11단계의 리커트 척도로 측정됨 ('0%'=1, '10%'=2, ~ '100%'=11). 차이의 검정은 이분산 가정하의 *t*검정을 이용함. 통계적 유의성은 * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ 을 의미함.

<표 3>은 통제변수들이 포함된 다중회귀분석결과를 나타낸다. 그러나 통제변수가 포함된 경우에도 앞서의 단순 *t*검정 결과와 크게 달라지지 않았다. 여전히 '사적 목적을 위한 비위행위 묵인(Q10)'에 대해서만 처치집단의 응답 평균이 통제집단의 응답 평균보다 5% 신뢰수준에서 0.586점만큼(약 5.9%) 통계적으로 유의하게 높았으며, 다른 9개의 윤리행위인식 변수들에 대하여는 통계적으로 유의한 차이가 없었

<표 3> 회귀분석: 윤리행위인식의 집단 간 평균 비교

전체	Q1. 승진청탁: 공직	Q2. 승진청탁: 사직	Q3. 수의계약 남용: 공직	Q4. 수의계약 남용: 사직	Q5. 자의적 규정해석	Q6. 긴급입찰 남용	Q7. 인허가 남용: 공직	Q8. 인허가 남용: 사직	Q9. 비위행위 목인: 공직	Q10. 비위행위 목인: 사직
처치집단 여부	0.086 (0.249)	0.076 (0.238)	0.021 (0.242)	0.027 (0.242)	-0.141 (0.221)	-0.275 (0.228)	-0.134 (0.219)	-0.071 (0.223)	0.351 (0.245)	0.586** (0.237)
근속기간	-0.010 (0.012)	-0.006 (0.012)	-0.043*** (0.012)	-0.035*** (0.013)	-0.027** (0.011)	-0.022** (0.011)	-0.040*** (0.011)	-0.037** (0.011)	0.020 (0.013)	0.019 (0.012)
공개채용 여부	0.086 (0.286)	-0.039 (0.283)	-0.037 (0.307)	0.003 (0.301)	0.161 (0.261)	-0.183 (0.272)	-0.132 (0.261)	0.038 (0.264)	-0.064 (0.296)	-0.202 (0.272)
상수	3.597*** (0.367)	3.187*** (0.355)	3.937*** (0.366)	4.059*** (0.366)	3.293*** (0.331)	3.577*** (0.343)	4.029*** (0.327)	4.015*** (0.335)	5.029*** (0.368)	4.981*** (0.346)
	<i>N</i> =597 <i>F</i> =0.30 <i>P</i> > <i>F</i> =0.8243 <i>R</i> ² =0.0014	<i>N</i> =597 <i>F</i> =0.19 <i>P</i> > <i>F</i> =0.9018 <i>R</i> ² =0.0008	<i>N</i> =597 <i>F</i> =4.70 <i>P</i> > <i>F</i> =0.0030 <i>R</i> ² =0.0229	<i>N</i> =597 <i>F</i> =2.90 <i>P</i> > <i>F</i> =0.0343 <i>R</i> ² =0.0149	<i>N</i> =597 <i>F</i> =2.00 <i>P</i> > <i>F</i> =0.1130 <i>R</i> ² =0.0105	<i>N</i> =597 <i>F</i> =2.11 <i>P</i> > <i>F</i> =0.0981 <i>R</i> ² =0.0101	<i>N</i> =597 <i>F</i> =5.45 <i>P</i> > <i>F</i> =0.0011 <i>R</i> ² =0.0251	<i>N</i> =597 <i>F</i> =4.03 <i>P</i> > <i>F</i> =0.0074 <i>R</i> ² =0.0185	<i>N</i> =597 <i>F</i> =1.42 <i>P</i> > <i>F</i> =0.2364 <i>R</i> ² =0.0744	<i>N</i> =597 <i>F</i> =2.83 <i>P</i> > <i>F</i> =0.0376 <i>R</i> ² =0.0142

참고: 각 종속변수는 11단계의 리커트 척도로 측정됨 ('0%'=1, '10%'=2, ~'100%'=11). 표의 셀(cell) 안의 숫자는 각각의 개별 OLS 회귀분석에서의 통제집단 대비 처치집단 변수의 계수를 나타내며, 괄호 안은 강건한 표준오차(robust standard error)임. 통계적 유의성은 * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ 을 의미함.

다.¹⁰⁾ 아울러, 통제변수인 ‘근속기간’은 6개의 윤리행위인식 변수(Q3, 4, 5, 6, 7, 8)와 음(-)의 연관성을 보인 반면, ‘채용방식(공개채용여부)’은 어떠한 윤리행위인식 변수와도 통계적으로 유의한 연관성이 없었다.

3. 윤리행위인식 분석: 직급별 부분집단 분석

앞의 다중회귀분석에서 ‘근속기간’과 ‘채용방식’ 외에 다른 통제변수를 모두 넣는 경우, ‘직급’, 특히 6~9급에 대비하여 4~5급 공무원인지를 나타내는 통제변수의 계수는 다수의 경우에 통계적 유의성을 가졌다. 이 경우 전체집단 분석에서 4~5급 공무원들이 가지는 응답의 특성이 1~3급 및 6~9급 공무원들의 응답들과 다른 패턴을 가졌을 가능성이 있기에, 본 연구에서는 직급에 따른 3개의 부분집단에 대한 분석(subgroup analysis)을 추가로 시행하였다.

직급별 부분집단에 대한 *t*검정 결과는 <표 4>가 나타내고 있다. 예상과 같이, 1~3급이나 6~9급과 달리, 4~5급 고위직 공무원들의 경우 처치 여부, 즉 BTS 방법에 따른 설문 여부에 따른 집단 간 응답 평균의 차이가 두드러졌다. 4~5급 공무원들의 처치집단은 같은 직급의 통제집단보다 6가지 상황(Q1-5, Q10)에서 비윤리적 행위를 할 가능성이 크다고 응답하였다. 구체적으로는, 이들 4~5급 공무원들의 처치집단은 통제집단보다 ‘공적 목적을 위한 승진청탁’(Q1)을 할 가능성을 약 12.5%p, ‘사적 목적을 위한 승진청탁(Q2)’ 가능성을 약 9.8%p, ‘공적 목적을 위한 수의계약 남용(Q3)’ 가능성을 약 7.5%p, ‘사적 목적을 위한 수의계약 남용(Q4)’ 가능성을 약 8.3%p 높게 평가하였다. 그리고 ‘자의적 규정해석(Q5)’을 할 가능성은 약 9.9%p, ‘사적 목적을 위한 (타인의) 비위행위 묵인(Q10)’ 가능성은 약 9.3%p 높게 응답하였다.

반면 1~3급 및 6~9급 공무원 집단들은 대부분의 윤리행위인식 변수에 대한 응답에 있어서 처치집단과 통제집단 간의 차이가 불분명했다. 오히려 다수인 7가지(Q1-4, Q6-8)의 윤리적 딜레마 상황에서는, 비록 통계적으로 유의하지는 않았지만, 비윤리적 행위를 할 가능성에 대하여 처치집단이 통제집단보다 응답 평균이 낮았다. 심지어

10) ‘근속기간’과 ‘채용방식’ 외에 ‘나이’와 ‘교육수준’ 등 개인적 요인들을 통제변수로 모두 넣은 다중회귀분석결과도 이와 큰 차이가 없이, ‘사적 목적을 위한 비위행위 묵인(Q10)’ 하나에 대해서만 집단 간 차이가 통계적으로 유의했다.

어 ‘자의적 규정해석(Q5)’ 행위를 할 가능성에 대하여는 처치집단의 응답 평균이 통제집단보다 통계적으로 유의미하게 낮기도 했다. 따라서 이러한 1~3급 및 6~9급의 집단 간 응답 평균의 음(-)의 차이들은 4~5급의 집단 간 응답 평균의 양(+)의 차이들을 상쇄하여, <표 2>의 전체집단 분석결과에서 본 바와 같이 대부분의 윤리 행위인식 변수들의 집단 간 차이를 통계적으로 유의하지 않게 만들고 심지어 그 차이값을 0에 가깝게 만든 것으로 보인다.

그러나 <표 4>의 *t*검정 분석결과도 앞서와 같이 처치 여부가 아니라 ‘근속기간’과 ‘채용방식’과 같은 잔여 인적 특성의 차이에 의해 나타났을 가능성을 배제할 수 없었다. 특히 부분집단 분석의 경우, 무작위 배정을 통해 원래의 처치-통제집단 간의 비교에서 모든 잠재적 혼동요인들이 제거된다는 장점이 더 이상 유지되지 않으므로, 그러한 혼동요인을 명시적으로 통제할 필요가 크다. 그러므로 이러한 우려를 검증하기 위하여 추가적으로 ‘근속기간’ 및 ‘채용방식’을 통제변수로 넣은 다중회귀 분석을 실시하였다. <표 5>는 그러한 직급별 부분집단 회귀분석결과를 나타내고 있다. 다만 지면의 한계상, <표 5>에서는 회귀분석에서의 모든 변수의 계수를 보고하는 대신, 관심 대상인 처치집단 여부의 계수만 나타내었다.¹¹⁾

통제변수를 넣은 다중회귀분석결과를 대체로 *t*검정 결과와 유사하였다. 4~5급 공무원 집단의 경우, 여전히 ‘승진청탁(Q1, 2)’과 ‘자의적 규정해석(Q5)’, ‘사적 목적을 위한 (타인의) 비위행위 묵인(Q10)’의 4가지 비윤리적 행위 가능성에 대하여 처치집단이 통제집단보다 통계적으로 유의하게 높은 응답 평균을 보였다. 다만 *t*검정 결과와 달리, ‘수의계약(Q3, 4)’의 두 가지 비윤리적 행위 가능성에 대하여는 두 집단 간 응답 평균 차이가 더 이상 통계적으로 유의하지 못했다. 반면, 1~3급 및 6~9급 공무원 집단들은 여전히 ‘자의적 규정해석(Q5)’에 대하여는 처치집단의 응답 평균이 예상과 달리 통제집단보다 낮았다. 아울러 6~9급 공무원 집단의 경우, *t*검정과 달리 다중회귀분석에서는 ‘긴급입찰 남용(Q6)’ 또한 10% 수준에서나마 통계적으로 유의하게 처치집단의 응답 평균이 통제집단의 응답 평균보다 낮았다.

11) 직급별 부분집단 회귀분석의 모든 변수에 관한 결과는 교신저자에게 요청 가능하다.

<표 4> 부분집단 *t*검정: 직급별 윤리행위인식의 집단 간 평균 비교

		Q1. 승진청탁: 공적	Q2. 승진청탁: 사적	Q3. 수의계약 남용: 공적	Q4. 수의계약 남용: 사적	Q5. 자의적 규정해석	Q6. 긴급입찰 남용	Q7. 인허가 남용: 공적	Q8. 인허가 남용: 사적	Q9. 비위행위 목인: 공적	Q10. 비위행위 목인: 사적
1~3급	처치집단 (<i>N</i> =22)	4.136 [2.731]	3.000 [2.047]	2.864 [2.867]	3.227 [2.487]	1.818 [1.736]	1.727 [1.751]	2.727 [2.004]	3.136 [2.122]	6.136 [2.513]	6.136 [2.475]
	통계집단 (<i>N</i> =26)	4.538 [3.153]	3.538 [2.983]	3.154 [3.016]	3.923 [3.136]	2.962 [2.807]	2.731 [2.539]	2.885 [2.487]	2.808 [2.400]	4.769 [2.566]	5.308 [2.542]
	차이	-0.402 (0.849)	-0.538 (0.730)	-0.290 (0.851)	-0.696 (0.812)	-1.143* (0.663)	-1.003 (0.622)	-0.157 (0.648)	+0.329 (0.653)	+0.329 (0.653)	+0.829 (0.726)
4~5급	처치집단 (<i>N</i> =92)	4.326 [3.305]	3.815 [3.077]	3.250 [3.370]	3.728 [3.332]	3.207 [2.881]	3.011 [2.764]	3.250 [2.728]	3.413 [2.750]	3.413 [2.750]	6.011 [2.514]
	통계집단 (<i>N</i> =103)	3.078 [3.002]	2.835 [2.894]	2.505 [2.578]	2.903 [2.703]	2.214 [2.569]	2.650 [2.750]	2.893 [2.814]	3.049 [2.847]	5.583 [3.120]	5.078 [3.089]
	차이	+1.248*** (0.454)	+0.980** (0.429)	+0.745* (0.434)	+0.825* (0.438)	+0.993** (0.393)	+0.360 (0.396)	+0.357 (0.397)	+0.364 (0.401)	+0.374 (0.432)	+0.933** (0.402)
6~9급	처치집단 (<i>N</i> =179)	3.034 [2.667]	2.749 [2.533]	2.877 [2.776]	3.123 [2.772]	2.492 [2.434]	2.642 [2.704]	2.771 [0.193]	3.039 [0.192]	5.631 [3.076]	5.721 [2.938]
	통계집단 (<i>N</i> =175)	3.469 [3.045]	3.040 [3.005]	3.091 [3.023]	3.331 [3.026]	3.080 [2.907]	3.091 [2.968]	3.034 [2.881]	3.309 [2.928]	5.469 [2.940]	5.377 [2.997]
	차이	-0.435 (0.304)	-0.291 (0.296)	-0.214 (0.309)	-0.209 (0.309)	-0.588** (0.285)	-0.449 (0.302)	-0.263 (0.291)	-0.269 (0.293)	+0.163 (0.320)	+0.163 (0.320)

참고: 꺾쇠 안의 숫자는 표준편차(standard deviation), 괄호 안의 숫자는 표준오차(standard error)를 의미함. 윤리행위인식 변수는 11단계의 리커트 척도로 측정됨 ('0%'=1, '100%'=11). 차이의 검정은 이분산 가정하의 *t*검정을 이용함. 통계적 유의성은 * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ 을 의미함.

<표 5> 부분집단 회귀분석: 직급별 윤리행위인식의 집단 간 평균 비교

	Q1. 승진청탁: 공직	Q2. 승진청탁: 사직	Q3. 수의계약 남용: 공직	Q4. 수의계약 남용: 사직	Q5. 자의적 규정해석	Q6. 긴급입찰 남용	Q7. 인허가 남용: 공직	Q8. 인허가 남용: 사직	Q9. 비위행위 묵인: 공직	Q10. 비위행위 묵인: 사직
처치(BTS)집단 여부 계수 (vs. 통제집단)										
1~3급 (N=48)	-0.678 (0.886)	-0.508 (0.775)	-0.285 (0.926)	-0.757 (0.894)	-1.368* (0.710)	-1.105 (0.658)	-0.319 (0.684)	0.099 (0.677)	1.358 (0.827)	0.858 (0.773)
4~5급 (N=195)	1.347*** (0.463)	1.054** (0.446)	0.670 (0.445)	0.737 (0.448)	0.941** (0.401)	0.355 (0.403)	0.220 (0.398)	0.264 (0.408)	0.332 (0.438)	0.819** (0.409)
6~9급 (N=354)	-0.489 (0.308)	-0.336 (0.302)	-0.320 (0.308)	-0.278 (0.310)	-0.637** (0.286)	-0.542* (0.303)	-0.364 (0.285)	-0.348 (0.290)	0.232 (0.324)	0.422 (0.318)

참고: 각 종속변수는 11단계의 리커트 척도로 측정됨 ('0%'=1, '10%'=2, ~'100%'=11). 표의 셀(cell) 안의 숫자는 각각의 개별 OLS 회귀분석에서의 통제집단 대비 처치집단 변수의 계수를 나타내며, 괄호 안은 강건한 표준오차(robust standard error)임. 모든 회귀분석은 '근속기간'과 '채용방식'의 통제변수들을 포함한 다중회귀분석이나, 본 표에서는 통제변수의 계수들은 생략함. 통계적 유의성은 * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ 을 의미함.

V. 토의 및 결론

본 연구에서는 공무원의 비윤리적 행위 여부와 같이 응답자가 솔직하게 대답하지 않을 가능성이 큰 설문에서 베이지안 진실 표출법(BTS)이 설문 참여자의 보다 솔직한 응답을 끌어낼 수 있는지에 대하여 실증적으로 검증하고자 하였다. 설문 응답의 하향 편향 가능성을 고려하면, 무작위배정으로 나눈 처치집단과 통제집단을 비교 시 BTS 방식으로 설문을 처치집단이 그렇지 않은 통제집단보다 그러한 비윤리적 행위를 할 가능성을 크게 평가하는 경우 BTS 방식이 솔직한 응답을 끌어내는 효과가 있다고 결론지을 수 있다.

이러한 기준에 비추어 보면, 연구결과는 BTS의 효용성에 대하여 긍정적이었다. 전체 597명을 대상으로 하는 설문에서는, 총 10개의 윤리적 딜레마 상황 중 ‘사적 목적을 위한 (타인의) 비위행위 묵인(Q10)’ 행동을 할지의 한 가지에 대하여만 처치집단이 통제집단에 비하여 그렇다고 응답한 비율이 통계적으로 유의미하게 높았다. 그러나 직급에 따른 부분집단 분석결과를 보면, 전체집단 분석결과와는 4~5급 공무원들의 응답 경향과 6~9급 공무원들의 응답 경향이 상충하여 상쇄된 것으로 볼 수 있었다. 통제변수 영향을 함께 고려한 다중회귀분석결과를 기준으로, 4~5급 공무원들은 모든 행위 가능성에 대하여 처치집단 응답이 통제집단보다 높았고, 그 응답 차이도 ‘사적 목적을 위한 (타인의) 비위행위 묵인(Q10)’과 더불어 ‘공적 및 사적 목적을 위한 승진청탁(Q1, 2)’과 ‘자의적 규정해석(Q5)’의 총 4가지 비윤리적 행위에 대하여 10%p에 가까운 정도로 크면서 통계적으로도 유의하였다. 반면, 6~9급 공무원들은 예상과 달리 오히려 대부분(10개 중 8개) 상황에서 처치집단의 응답 평균이 통제집단보다 낮았으며, 심지어 ‘자의적 규정해석(Q5)’과 ‘긴급입찰 남용(Q6)’ 행위에 대해서는 통계적으로 유의한 차이를 보였다.

4~5급 공무원들과 달리 6~9급 공무원의 경우 예상과 반대의 결과가 나왔지만, 그렇다고 이것이 곧 BTS 방식의 효용성을 부정하는 것은 아니다. 직급별 설문 결과의 차이에 대한 여러 설명이 가능하겠지만, 그중 하나는 4~5급과 6~9급이 윤리적 딜레마 상황에서 가질 수 있는 권한 차이를 들 수 있다. 본 연구의 윤리적 딜레마 상황들은 대체로 조직이 어떻게 행동할지 실무적 결정권을 가지는 4~5급 공무

원들이 현실에서 자주 처하게 되는 상황이므로, 비록 가상의 상황이지만 이들은 자신들이 처했었던 과거 현실의 상황을 대입해 응답했을 것으로 짐작된다. 반면 6~9급 공무원들은 주어진 윤리적 딜레마 상황 대부분에서 판단하고 결정하는 권한을 가지는 경우가 별로 없다. 따라서 6~9급 공무원들의 경우, 통제집단은 응답에 본인의 진솔한 생각보다 현실에서 자주 목격하는 조직 내 상급 직급자들의 일상적인 결정을 반영하였지만, 처치집단의 경우 “나라면 그렇게 하지 않을 텐데”라는 본인의 솔직한 이상적(ideal)인 생각을 나타낸 것으로 짐작할 수 있다. 따라서 설문에서 6~9급 공무원들에게 맞춰진 다른 윤리적 딜레마 상황이었다면, 6~9급 공무원들의 설문결과는 다르게 나타났을 수도 있다.

BTS 방식이 설문 참여자의 더 솔직한 응답을 끌어낼 수 있었다는 본 연구결과는, 반대로 이야기하면 기존의 설문, 특히 사회적 비난 가능성이 있어 설문 참여자가 솔직하게 응답하지 않을 유인이 있는 기존 설문의 결과들이 어느 정도 편향되었을 가능성이 있음을 시사한다. 현실과 괴리된 설문 결과는 자칫 일반 대중들의 인식을 왜곡하거나(예컨대 과도한 범죄에 대한 두려움 확산) 국가 정책을 왜곡하는(예컨대 과대 표현된 이용 의도에 따른 지역 공항의 과다 건설) 등 우리 사회에서의 실제적 문제들을 일으킬 수 있다. 따라서 BTS 방식과 같이 설문 참여자에게 솔직하고 신중한 응답을 할 유인 기제를 설문에 적극적으로 도입하고, 아울러 이미 시행된 설문이라도 그 결과의 질(quality)을 평가하여 제고시키려는 부단한 노력이 필요하다.

다만 다른 연구들과 마찬가지로 본 연구 또한 한계를 가지고 있음을 주지할 필요가 있다. 첫째, 일반적인 BTS 방식의 한계로서, 이를 사용한다고 해도 설문 응답이 정말로 진실인지는 궁극적으로 알 수 없다. 특히 객관적인 정답이 없어 진위를 판단할 수 없는 주관적 인식에 대한 설문에서는, BTS 방법은 경제적 유인에 의해 응답자가 얼마나 솔직히 응답했을 가능성이 있는가를 나름의 기준에 따라 측정할 뿐이다. 이는 본 연구에서도 마찬가지로, BTS 방식으로 설문한 처치집단이 통제집단보다 비윤리적 행위의 가능성을 더 크게 응답하였다고 해도, 그것이 곧 처치집단의 응답이 진실이고 통제집단의 응답이 거짓임을 의미하지 않으며, 또한 그 차이의 크기가 곧 진실한 응답과의 괴리의 크기를 의미하지 않는다. 따라서 BTS 방식은 만능(panacea)이 아니며, 차후 연구에서 그의 유용성과 한계에 대하여 더 밝혀져야 할 부분이 아직 많다. 둘째, 본 연구만의 한계로, 설문 참여자의 모집을 편의 표본추출에 기대었기에, 본 연구의 결과는 우리나라의 모든 공무원에 대한 것으로 일반화할

가능성, 즉 외적 타당성(external validity)이 높지 않다. 따라서 차후 연구에서는 전체 공무원을 잘 대표할 수 있는 공무원 표본들을 선발할 수 있도록 확률적 표본추출 방법이나 최소 할당 표본추출 방법을 사용하여 다시 연구할 필요가 있다.

이러한 한계들에도 불구하고, 본 연구는 베이지안 진실 표출법(BTS)이라는 새로운 방법을 도입하여 공무원의 비윤리적 행위 가능성을 직접 설문조사 연구한 최초의 국내 연구라는 의의가 있다. 특히 BTS 방식을 사용하여 설문하는 경우 유인 기제가 없는 기존의 설문방식을 사용했을 때보다 공무원들이 비윤리적 행위를 할 가능성을 더 크게 평가하는 것으로 나타났고, 이는 기존의 공무원 부패, 부정행위, 윤리행위 등을 설문한 결과가 어느 정도 현실과 괴리되어 편향되었을 가능성을 시사한다. 그리고 사회과학 분야에서 설문이 아직도 주요한 연구도구로 사용되는 만큼, 설문에서의 유인 기제 설정과 연구결과의 평가에 대한 논의가 필요함을 제시하는 중요한 함의를 가진다고 하겠다.

참고문헌

- 강정석·고재권. 2016. “국민들의 정부부문 부패 인식에 영향을 미치는 요인 분석.” 《한국경제포럼》 9(3): 41-63.
- 고길곤·조수연. 2012. “관행수용도와 부패.” 《한국행정학보》 46(3): 213-239.
- 고길곤·박준·이시영. 2019. “부패경험이 인식된 공직부패수준에 미치는 영향의 크기: 확장 역확률 가중치 접근을 이용한 준비모수적 접근을 중심으로” 《한국행정학보》 53(2): 91-121.
- 권희경·이현주. 2020. “사회적 바람직성 편향이 자기보고식 측정에 미치는 영향.” 《한국 청소년연구》31(4): 251-282.
- 김승현. 2013. “사회적 규범의 측정: 응답편향의 실험적 연구.” 《한국정치연구》 22(2): 153-178.
- 박중훈·박준·정동재. 2017. 《청탁금지법 시대 한국사회의 부패 인식과 경험》. 한국행정연구원 연구보고서.
- 박천오. 2014. “한국 지방공무원의 윤리 의식에 관한 실증 연구: 정책윤리를 중심으로.” 《한국인사행정학회보》 13(2): 93-112.
- 박현환. 2019. “공직봉사동기와 부패인식의 연관성에 관한 연구.” 서울대학교 대학원 행정학 박사학위논문.

- 송인덕·조성겸. 2013. “서베이 모드와 사회적 바람직함 편향: 민감한 주제에 대한 전화 면접, 대면 면접, 온라인 서베이 응답 비교” 《조사연구》 14(2): 165-199.
- 오세영. 2020. 《정부부문 부패실태에 관한 연구》. 한국행정연구원 연구보고서.
- 장용진. 2015. “정부신뢰와 부패인식: 아시아 국가를 중심으로.” 《한국부패학회보》 20(2): 21-43.
- 최혜민. 2021. 《정부부문 부패실태에 관한 연구》. 한국행정연구원 연구보고서.
- Barrage, L. and M.S. Lee. 2010. “A Penny for Your Thoughts: Inducing Truth-telling in Stated Preference Elicitation.” *Economics Letters* 106(2): 140-142.
- de-Magistris, T. and S. Pascucci. 2014. “The Effect of the Solemn Oath Script in Hypothetical Choice Experiment Survey: A Pilot Study.” *Economics Letters* 123(2): 252-255.
- Frank, M.R., M. Cebrian, G. Pickard, and I. Rahwan. 2017. “Validating Bayesian Truth Serum in Large-scale Online Human Experiments.” *PLoS One* 12(5): e0177385.
- Howie, P.J., Y. Wang, and J. Tsai. 2011. “Predicting New Product Adoption Using Bayesian Truth Serum.” *Journal of Medical Marketing* 11(1): 6-16.
- John, L.K., G. Loewenstein, and D. Prelec. 2012. “Measuring the Prevalence of Questionable Research Practices with Incentives for Truth Telling.” *Psychological Science* 23(5): 524-532.
- Linstone, H.A. and M. Turoff. 1975. *The Delphi Method: Techniques and Applications*. Reading, MA: Addison-Wesley.
- Loughran, T.A., R. Paternoster, and K.J. Thomas. 2014. “Incentivizing Responses to Self-report Questions in Perceptual Deterrence Studies: An Investigation of the Validity of Deterrence Theory Using Bayesian Truth Serum.” *Journal of Quantitative Criminology* 30(4): 677-707.
- Miller, N., P. Resnick, and R. Zeckhauser. 2005. “Eliciting Informative Feedback: The Peer-Prediction Method.” *Management Science* 51(9): 1359-1373.
- Prelec, D. 2004. “A Bayesian Truth Serum for Subjective Data.” *Science* 306(5695): 462-466.
- Prelec, D., H.S. Seung, and J. McCoy. 2017. “A Solution to the Single-question Crowd Wisdom Problem.” *Nature* 541(7638): 532-535.
- Schoenegger, P. 2023. “Experimental Philosophy and the Incentivisation Challenge: A Proposed Application of the Bayesian Truth Serum.” *Review of Philosophy and Psychology* 14: 295-320.
- Weaver, R. and D. Prelec. 2013. “Creating Truth-telling Incentives with the Bayesian

Truth Serum.” *Journal of Marketing Research* 50(3): 289–302.

Zhou, F., L. Page, R.K. Perrons, Z. Zheng, and S. Washington. 2019. “Long-term Forecasts for Energy Commodities Price: What the Experts Think.” *Energy Economics* 84: 104484.

<접수 2024.02.19; 수정 2024.03.27; 게재확정 2024.04.17>

Investigating Response Bias in Survey on Public Employees’ Perception of Ethical Behaviors: A Bayesian Truth Serum Approach

SeungHoon Han
(Chung-Ang University)
Junseok Lee
(Chung-Ang University)

Social science survey research faces various response biases when investigating subjective opinions. One common bias is the social desirability bias, which leads survey participants to choose answers that align with societal norms rather than their true beliefs. This bias is particularly pronounced among specific groups, such as public officers, who may fear social repercussions if their true answers are revealed. In this study, we aimed to assess the effectiveness of the Bayesian Truth Serum (BTS) method in mitigating the social desirability bias. The BTS method involves providing economic incentives that are maximized when respondents provide honest answers. To achieve this, we surveyed 597 public officers about their perceived likelihood of engaging in unethical behaviors in hypothetical situations. Participants were randomly assigned to either the BTS treatment group or the non-BTS control group. The results showed that the treatment group exhibited a higher intention to engage in unethical behaviors compared to the control group. Moreover, this tendency was more pronounced among intermediate-level manager officers holding ranks 4 to 5, who are likely to experience such ethically conflicting situations, than among other rank officers. These findings suggest that the Bayesian Truth Serum method effectively encourages candid responses from survey participants.

Key words: ethical behavior perceptions among public officers, Bayesian Truth Serum, social desirability bias, random assignment, thought experiment

〈부록 1〉 BTS에서의 점수 계산

설문에서 총 n 명의 응답자 중 r 번째 응답자가 질문에 대한 총 m 개의 가능한 답변 중 k 번째 특정 답변($k=1, 2, \dots, m$)을 선택할 수 있다고 하면, 이때 응답자가 선택한 답변은 $x^r = (x_1^r, \dots, x_m^r)$ (단 $x_k^r \in \{0, 1\}$, $\sum_r x_k^r = 1$), 타 응답자에 대한 응답률 추측치는 $y^r = (y_1^r, \dots, y_m^r)$ (단 $y_k^r \geq 0$, $\sum_r y_k^r = 1$)와 같이 벡터(vector)로 표기할 수 있다. 이 경우 k 번째 특정 답변의 응답자 전체 표본에서의 실제 응답률 $\overline{x_k}$ 와 그러한 k 번째 특정 답변의 응답자 전체 표본에서의 평균 추측치 $\overline{y_k}$ 는 다음과 같다.

$$\overline{x_k} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_r x_k^r, \quad \log \overline{y_k} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_r \log y_k^r$$

이를 바탕으로 하여 질문 문항에 대한 응답자 r 의 정보 점수와 추측 정확성 점수는 각각 다음과 같이 계산된다.

$$\text{정보 점수:} \quad \sum_k (x_k^r \cdot \log \frac{\overline{x_k}}{y_k^r}) \quad (1)$$

$$\text{추측치 정확도 점수:} \quad \sum_k (\overline{x_k} \cdot \log \frac{y_k^r}{x_k^r}) \quad (2)$$

최종적으로 이 질문 문항에서 응답자 r 이 얻을 수 있는 BTS 총점수는 아래 (3)과 같이 정보 점수와 추측치 정확도 점수의 가중합으로 결정된다. 이때 추측치 정확도에 대한 가중치 α 를 조절함으로써, 연구자는 정보 점수와 추측치 정확도 점수 간의 상대적 영향력을 조정할 수 있다. Prelec(2004)에 따르면, α 가 충분히 작을 경우 BTS 총점수는 게임이론에서의 파레토 지배적(Pareto-dominated)인 균형이 되며, $\alpha = 1$ 일 때는 제로섬 게임(zero-sum game)이 된다.

응답자 r 의 BTS 총점수:

$$\sum_k (x_k^r \cdot \log \frac{\overline{x_k}}{y_k^r}) + \alpha \cdot \sum_k (\overline{x_k} \cdot \log \frac{y_k^r}{x_k^r}) \quad (\text{단, } > 0) \quad (3)$$

〈부록 2〉 공무원 윤리행위인식 설문 문항

[가상 시나리오 1] 당신은 모 부처에서 과장으로 근무하고 있습니다. 최근 내년도 승진심사가 시작되었고, 당신은 7급 주사보인 과 내 부하직원 A씨를 6급 주사로 승진시키고자 합니다. 그러나 당신은 이번에 A씨 외에도 타과의 뛰어난 승진심사 대상자가 많아 A씨가 승진이 안될 수도 있다는 사실을 알게 되었습니다. 마침 고향 선후배로 종종 개인적인 친목 모임을 가질 정도로 친하게 지내는 인사과의 B과장이 이번 승진심사를 담당하고 있음을 알게된 당신은, B과장과 친목 술자리 모임을 만들고 그 자리에 A씨를 함께 불러 승진 청탁을 할 것인지 고민하고 있습니다.

Q1) 공적 목적을 위한 승진청탁

- (1) 당신은 “A씨가 능력은 보통이어서 일을 잘하는 것은 아니지만, A씨가 승진을 하게되면 우리 과 직원 간 업무 부담 격차를 조정할 수 있어 부처 전체의 업무 성과가 올라갈 수 있다”라고 생각하고 있습니다. 그렇다면, 당신은 이러한 이유로 B과장에게 A씨의 승진 청탁을 하겠습니까?
- (2) 위의 질문에 대하여, 본인 외 다른 설문 응답자들이라면 같은 상황에서 술자리 청탁을 할 가능성이 얼마나 된다고 응답하였을 것으로 생각하십니까?

Q2) 사적 목적을 위한 승진청탁

- (1) 당신은 이 “A씨가 능력은 보통이어서 일을 잘하는 것은 아니지만, 나와 B와 같은 고향 출신이라 A씨만큼 우리가 믿을 수 있는 사람이 없고, 그러니 우리가 밀어주어야 한다”라고 생각하고 있습니다. 그렇다면, 당신은 이러한 이유로 B과장에게 A씨의 승진 청탁을 하겠습니까?
- (2) 위의 질문에 대하여, 본인 외 다른 설문 응답자들이라면 같은 상황에서 술자리 청탁을 할 가능성이 얼마나 된다고 응답하였을 것으로 생각하십니까?

[가상 시나리오 2] 당신은 정부 모 부처에서 기술 관련 연구용역을 관리하는 업무를 맡고 있습니다. 최근 당신의 부처에서는 AI 기술과 관련하여 신규 연구용역을 발주하기로 결정하였습니다. 그러나 당신은 아래와 같은 이유로, 연구비 규모를 공개입찰 의무금액보다 살짝 줄여, 공개입찰 대신 당신의 친한 고교 동창인 모 대학 A교수와 수의계약으로 발주할 것인지 고민하고 있습니다.

Q3) 공적 목적을 위한 수의계약 남용

- (1) 당신은 “A교수는 연구실적이 뛰어나지도 않고 AI 전공자도 아니지만, 용역 규모가 작아 공개입찰을 해도 다른 지원자가 없을 것 같고 그럼에도 우리 부처는 반드시 이번 용역의 결과가 필요하므로, 친구인 A교수에게 부탁해서라도 용역을 진행해야겠다”라고 생각하고 있다고 가정합니다. 그렇다면, 당신은 이러한 이유로 공개입찰 대신 A교수와의 수의계약을 추진하겠습니까?
- (2) 위의 질문에 대하여, 본인 외 다른 설문 응답자들이라면 같은 상황에서 수의계약을 추진할 가능성이 얼마나 된다고 응답하였을 것으로 생각하십니까?

Q4) 사적 목적을 위한 수의계약 남용

- (1) 당신은 “A교수는 연구실적이 뛰어나지도 않고 AI 전공자도 아니지만, A교수가 본인의 정교수 승진에 연구용역 간접비 실적이 절실히 필요하다고 한번만 도와달라고 하고, 또 내가 도와줄 수 있는 위치에 있으니, 친구인 내가 도와줘야겠다”라고 생각하고 있다고 가정합니다. 그렇다면, 당신은 이러한 이유로 A교수와의 수의계약을 추진하겠습니까?
- (2) 위의 질문에 대하여, 본인 외 다른 설문 응답자들이라면 같은 상황에서 수의계약을 추진할 가능성이 얼마나 된다고 응답하였을 것으로 생각하십니까?

[가상 시나리오 3] 당신은 정부 모 부처에서 정부지원사업 외주 관리 업무를 담당하고 있습니다. 최근 당신의 부처에서는 신규 평가 분야 사업의 외주 발주에 있어서, 그동안 평가 분야는 아니었지만 우리 부처의 여러 교육 분야 외주 사업들에 참여하여 좋은 실적을 냈던 A업체를 선정해주길 바라고 있습니다. 그러나 담당자인 당신은 다음과 같은 이유로 A업체를 선정해야 하는지 그리고 한다면 어떤 방법으로 선정할지 고민하고 있습니다.

Q5) 자의적 규정해석

- (1) A업체 자격을 검토해 본 당신은, 자격 관련 조항인 “자격 (가) 또는 자격 (나) 등 중 하나를 갖추고 있을 것”에서의 자격 (가)와 (나) 모두 없어 A업체가 자격 미달임을 알게 되었습니다. 그러나 당신은 상사로부터 “자격 미달인 A업체를 꼭 선정해야 한다면, 자격 규정을 유연하게 해석하면 된다. 즉, 문구 중 ‘등’ 부분을 적극적으로 해석하여, A업체가 자격 (가)와 (나) 외의 ‘등’에 해당하는 자격을 갖춘 것으로 해석하면 된다.”라고 들었습니다. 그렇다면, 당신은 A업체의 자격을 이처럼 자의적으로 해석하여 A업체 선정을 추진하겠습니까?
- (2) 위의 질문에 대하여, 본인 외 다른 설문 응답자들이라면 같은 상황에서 자의적 해석으로 A업체를 선정을 추진할 가능성이 얼마나 된다고 응답하였을 것으로 생각하십니까?

Q6) 긴급입찰 남용

- (1) 당신은 A업체가 자격은 있으나, 예상되는 이번 신사업 외주 신청자가 많아 공정한 경쟁 입찰 시 A업체가 선정되지 않을 수 있다는 것을 알게 되었습니다. 이에 당신은 상사로부터 “A업체를 반드시 선정하려면, 긴급입찰로 입찰 공고 기간을 매우 짧게 기습적으로 내어 다른 업체들이 준비 못하게 만들면 된다. A업체에게만 공고 전에 미리 언질을 주고 입찰시키면 반드시 선정되도록 만들 수 있다.”라고 들었습니다. 그렇다면, 당신은 이와 같이 편법적인 긴급입찰 방식을 사용하여 A업체가 반드시 선정되도록 하겠습니까?
- (2) 위의 질문에 대하여, 본인 외 다른 설문 응답자들이라면 같은 상황에서 편법적 입찰 방식을 사용하여 A업체를 선정할 가능성이 얼마나 된다고 응답

하였을 것으로 생각하십니까?

[가상 시나리오 4] 당신은 정부 모 부처에서 인허가 업무를 담당하고 있다고 가정합니다. 그런데 평소에 개인적으로 친분이 있는 타 부처 A국장으로부터 연락을 받았습니다. 알고 보니 당신이 최근 신청을 불허한 업체에 근무하는 직원 B씨와 A국장이 가까운 친척 관계이고, 가능하면 이번 B씨 업체의 신청을 허가해달라는 청탁을 받았습니다. 이에 당신은 아래와 같은 이유로 A국장의 청탁을 들어주어야 하는지 고민하고 있습니다.

Q7) 공적 목적을 위한 인허가 남용

- (1) 당신은 당신 부처의 직속 상사인 C국장으로부터 “A국장이 속한 부처의 협조는 우리 부처의 업무 진행에 있어서 매우 중요한 만큼, 그 부처의 실세인 A국장을 불편하게 만들지 말라. 이번에도 그냥 허가해주고 좋게 좋게 넘어가라.”는 말을 들었다고 가정합니다. 그렇다면, 당신은 이러한 이유로 B씨 업체에 대한 불허 결정을 번복하여 다시 허가로 결정 내리겠습니까?
- (2) 위의 질문에 대하여, 본인 외 다른 설문 응답자들이라면 같은 상황에서 B씨 업체에 대하여 허가로 번복할 가능성이 얼마나 된다고 응답하였을 것으로 생각하십니까?

Q8) 사적 목적을 위한 인허가 남용

- (1) 당신은 이미 이전에 개인적인 일로 A국장에게 크게 신세를 진 일이 있어 “A국장의 부탁을 거절하기에는 너무 미안하고, 이번에 거절하면 A국장과 멀어질텐데...”라고 걱정하고 있다고 가정합니다. 그렇다면, 당신은 이러한 이유로 B씨 업체에 대한 불허 결정을 번복하여 다시 허가로 결정 내리겠습니까?
- (2) 위의 질문에 대하여, 본인 외 다른 설문 응답자들이라면 같은 상황에서 B씨 업체에 대하여 허가로 번복할 가능성이 얼마나 된다고 응답하였을 것으로 생각하십니까?

[가상 시나리오 5] 당신은 정부 모 부처 모 과에서 사무관으로 일하고 있다고 가정합니다. 그런데 우연한 기회에 당신은 과의 직속 상사인 A과장이 부처 거래처로부터 리베이트를 받았다는 사실을 알게 되었고, 또한 당신이 알았다는 사실을 A과장도 알게 되었습니다. 그러나 당신은 아래와 같은 이유로 A과장의 비위행위를 신고할지 고민하고 있습니다.

Q9) 공적 목적을 위한 (타인의) 비위행위 목인

- (1) 당신은 A과장으로부터 “내가 리베이트를 받은 것은 맞지만, 대부분은 내 개인적인 용도로 사용한 것이 아니라 우리 과의 회식비와 운영비, 경조사비 등으로 썼다. 당신도 그 혜택을 받은 게 아니냐. 그러니 당신은 이 사실을 모른척 해달라.”라는 부탁을 들었다고 가정합니다. 그렇다면, 당신은 A과장의 부탁에도 불구하고 A과장의 비위행위를 신고하겠습니까?
- (2) 위의 질문에 대하여, 본인 외 다른 설문 응답자들이라면 같은 상황에서 A팀장의 비위행위를 신고할 가능성이 얼마나 된다고 응답하였을 것으로 생각하십니까?

Q10) 사적 목적을 위한 (타인의) 비위행위 목인

- (1) A과장은 당신의 동문 선배이자 고향 선배여서 이전부터 당신에게 업무에서 여러 편의를 봐주던 사람이었고, 당신은 A과장으로부터 “내가 리베이트를 받은 것은 맞지만, 당신만 모른척 해주면 아무 문제 없을 것이니 그냥 모른척 해달라. 우리 사이가 어떤 사이인데 그 정도는 해줄 수 있지 않냐. 또 이 사실이 알려지면, 앞으로 당신도 회사 생활이 매우 고달파질 것이다”말을 들었다고 가정합니다. 그렇다면, 당신은 A과장의 부탁에도 불구하고 A과장의 비위행위를 신고하겠습니까?
- (2) 위의 질문에 대하여, 본인 외 다른 설문 응답자들이라면 같은 상황에서 A팀장의 비위행위를 신고할 가능성이 얼마나 된다고 응답하였을 것으로 생각하십니까?

* 응답 형식

안 한다						한다				
①	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩
0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%