

# 사회적 규범의 측정: 응답편향의 실험적 연구\*

김승현 | 서울과학기술대학교

사회과학은 자료수집방법 중 사회조사에 크게 의존하고 있지만 특정 문항에서는 응답자의 진정한 답변을 구하기가 쉽지 않다. 응답자들은 자신이 느끼는 것 또는 행태를 보이기도는 사회적으로 바람직하다고 생각하는 것이나, 극단적인 응답 또는 중도적 응답 등으로 답하려는 경향을 보인다. 이 문제는 주로 응답편향 또는 응답세트(response set)라는 용어로 포괄적으로 지적되었다. 이 연구에서는 사회적 신뢰, 공익지향성 등의 규범적 개념을 전통적 가치인 가족주의와 대비하여 응답편향의 문제점을 분석한다. 비교적 민감한 사회문제에 대한 참지지도를 알아보기 위해 쓰이는 열거식 실험적 조사를 통해 세 가지 개념을 측정하고, 그 결과를 전국적 확률표본에 의한 인터뷰 조사의 결과와 비교한다. 분석결과 사회적 규범의 측정에서 체계적 오류를 발견하였으나 가족주의의 측정에서는 유의미한 편향을 찾을 수 없었다. 이는 사회적 조사에 의한 사회적 규범의 행태적 이해도 잘못되지만, 두 가지 대립적 가치의 상관성 연구가 잘못될 가능성이 크다는 것을 의미한다.

**주제어:** 응답세트, 사회적 소망성편향, 열거식 실험조사

## I. 서론

오늘날 사회과학에서 자료수집방법 중 가장 널리 이용되는 방법은 사회조사임이 분명하다. 직접 인터뷰에 의하기도 하지만 자기보고식으로 문항에 답하게 하기도 한다. 최근에는 급속하게 발달하는 인터넷에 의해 응답을 수집하는 방법도 있다. 그렇지만 이러한 자료수집방법에 대한 방법론적 논란도 계속되고 있다. 연구자들은 응답자의 진정한 답변

\*이 논문은 2009년도 정부재원(교육과학기술부 인문사회역량강화사업비)으로 한국연구재단의 지원(KRF-2009-327-B00033)을 받아 연구되었음.

을 구하기가 쉽지 않다는 점을 오래 전부터 강조해왔다. 이러한 문제들은 주로 응답편향(response bias) 또는 응답세트(response set)라는 용어로 포괄적으로 지적되었다(Cronbach 1946; Schuluf et al. 2008). 여기서 응답편향이란 응답자가 설문이 측정하고자 하는 내용과는 상관없는 다른 기준에 의해 응답하려는 체계적인 경향을 말한다.

기존의 다양한 연구에 의하면 응답자들은 자신이 느끼는 것 또는 행태를 보이기보다는 사회적으로 바람직하다고 생각하는 것이나, 극단적인 응답 또는 중도적 응답 등으로 답하려는 경향을 보인다. 물론 이러한 응답편향이 무작위적인 경우에는 별 문제가 되지 않지만 체계적 오차로 작용하는 경우를 말한다. 이 중에서 소망성편향에 대한 연구가 가장 활발하고, 그 배경에는 특정 사회의 문화적 배경이 중요한 결정요인이라는 사실도 잘 알려져 있다. 집단주의적 문화에서는 이러한 문제점이 더 크게 나타난다고 한다. 이는 여러 연구전통에서도 지적하는 바로, 방법론, 정치학, 심리학 등에서 활발하게 연구되고 있다. 예를 들어 가치와 행태의 괴리에 관한 다양한 이론적·경험적 연구들도 왜곡의 가능성을 비교적 관점에서 논하고 있다(예: Bardi and Schwartz 2003). 특히 보편적 가치나 박애와 같은 개념은 행태와의 간극이 상대적으로 크고, 보수적이며 전통적 가치의 측정에서는 이러한 문제점이 비교적 작게 나타난다는 점을 지적하고 있다. 이는 가치의 측정이 행태를 짐작하게 하는 소위 예측타당성(predictive validity)을 크게 훼손하는 것이다. 많은 연구를 통해서 확인된 이러한 응답세트는 규범적 개념을 측정하려는 시도를 어렵게 할 뿐 아니라 개념 간의 상관성 연구에도 커다란 장애가 된다. 다시 말해서 사회현상의 이해를 어렵게 할 뿐 아니라 현상 간의 인과적 이해와 추론도 오류에 빠뜨리게 된다.

이 연구는 정치학의 연구관심에서 사회적 소망성편향을 중점적으로 확인하고 그 효과가 어느 정도 우려할만한 것인지를 알아보고자 한다. 정치학의 중요한 연구관심분야 중의 하나는 비교적 시각에서 보편적 가치나 사회적 또는 시민적 규범을 측정하는 것이다. 이를 위해서 인종적 태도, 이민문제 등 비교적 민감한 사회문제에 대한 솔직한 답변 또는 내면화된 가치인 참지지도(true support)를 알아보기 위한 실험적 방법인 열거식 실험조사(list experiment 또는 item count technique)를 이용한다. 이 연구에서는 가족주의, 사회적 신뢰 및 공익추구 등의 세 가지 개념을 대상으로 한다. 반세기 전에 출간된 시민문화의 비교연구(Almond and Verba 1963)에서 안정적 민주제도를 가능하게 하는 조건으로서 신뢰의 중요성을 강조하였다. 정치참여에 필요한 이차적 집단이 발달하는 데 가장 기본적 조건은 사회적 신뢰이고, 이를 통해 일차적 집단의 장벽을 무너뜨리고 다른 사람들과의 유대나 의무감을 확립한다. 그리고 이 신뢰는 가족집단에만 충성심과 협동을 한정하는 가족주의에 반

한다. 1980년의 유사한 반복적 연구(Almond and Verba 1980)에서도 민주적 문화에 필요한 정치사회적 태도가 발달하기 위해서는 가정의 민주화가 중요하다고 주장하였다. 1990년대 본격화된, 한국에서 아시아적 가치가 갖는 역할에 관한 연구에서도 이들 개념은 중요한 측정과 분석의 대상이 된다.<sup>1)</sup> 이렇게 전통적 가치로서 가족주의와 사회적·시민적 규범은 항상 대척점에서 이해되고 평가된다. 그럼에도 불구하고 사회조사에 기초한 상관성연구에서 부정적인 결과를 찾기 어렵다. 의심할 수 있는 가능성은 사회적 규범의 측정에 체계적인 오류가 개입함으로써 상관성의 크기를 왜곡하는 억제효과가 작용한다는 가정을 해볼 수 있다. 이러한 연구관심과 응답편향에 관한 연구들을 고려하면, 사회적 신뢰나 공익지향성 등의 사회적 규범을 전통적 가치인 가족주의의 개념과 대비하여 측정오류의 문제를 알아볼 필요가 있다. 전통적 가치와 사회적 규범의 측정에서 어떤 유형의 체계적인 측정오차가 있다면 사회적 현실을 이해하는 것뿐 아니라 두 종류의 개념 간의 상관성 추론도 잘못될 가능성이 크기 때문이다.

본 연구에서는 실험적 조사와 그 결과를 전국적 확률표본추출에 의해 조사한 유사한 문항들의 결과와 비교한다. 이러한 비교를 통해서 사회적 소망성에 의한 측정오차가 무작위적이지 않고 체계적 오차인지를 확인하고 궁극적으로 상관성 추정에 어떤 영향을 미칠 것인지 알아본다. 나아가 관련 문제점을 보다 확연하고 체계적으로 이해할 수 있을 것이다.

## II. 전통적 가치와 사회적 규범의 측정과 응답편향

리커트척도와 응답편향에 대한 다양한 연구들은 많은 요인들에 의해 정확한 응답을 얻기가 얼마나 어려운가를 강조하고 있다. 일찍이 크론바하(Cronbach 1946, 475)는 응답자가 동일한 내용의 질문이 다른 형태로 주어지면 또 다르게 응답하려는 경향을 응답세트라고 정의하였다. 이러한 응답세트는 사회적 규범을 측정하는 데서 빈번하게 나타나고 이는 주로 사회적 소망성이나 묵시적 동의에서 기인한다(Edwards 1957).

1) 광범위하게 연구되는 사회자본론에서도 두 가지 개념은 상반되게 이해된다. 특수한 신뢰는 주로 가족, 친척, 이웃 등 한정된 집단 내에서 배타적인 성격을 갖는 신뢰이다. 이렇게 집단 내 유대감이 강화되면 자연히 낯선 사람이나 사회적 배경, 종교 등이 다른 사람들에게 대해서는 배타적일 수밖에 없다는 논리적 주장이 경험적 연구결과와 함께 제시되어 왔다. 이러한 맥락에서 가족주의는 특수한 신뢰의 가장 기초가 되는 부정적 용어로 사용되어 왔다.

그런데 연구자들은 변수측정에 있어서 편향의 가능성이 있다는 것을 알면서도 문제의 심각성을 무시하는 경향이 있다. 특히 종속변수의 측정에서 편향이 있는 경우에는 이러한 무관심이 두드러진다. 왜냐하면 종속변수의 측정오차는 주로 추정치의 효율성을 낮춤으로써 계수의 통계적 유의미성만을 감소시키는 것으로 알기 때문이다(King et al. 1994, 158-163). 단지 통계적으로 유의미한 상관성을 찾기 어렵다고 생각하지만 측정오차가 무작위적이지 않고, 하위집단별로 체계적인 차이를 보인다면 현상에 대한 이해는 물론 상관성에 관한 추론도 잘못된다. 그러므로 편향의 원인과 이에 의해 예상되는 문제를 기존 연구에 비추어 구체적으로 분석할 필요가 있다.

## 1. 문화적 배경과 사회적 소망성편향

응답편향에 관한 오래된 일반적 주장에 의하면 편향의 내용과 크기는 문화적 배경에 따라 달라진다(Crowne and Marlow 1964). 이러한 연구에서 가장 중요한 구분은 집합주의(collectivism)와 개인주의(individualism) 문화이다(Triandis 1995). 집합주의문화는 스스로를 독립적인 개체로 보기보다는 소집단과 동일시하고 이 집단의 목표가 개인적 목표보다 선행하며, 상호주의, 안전, 복종, 의무, 조화, 위계질서 및 개인적 관계를 중시한다. 이에 비해 개인주의문화에서는 독립성, 행복, 성취, 자유, 자율, 공정, 교환 등을 강조한다(Triandis 1994). 집합주의적인 문화적 배경을 갖는 응답자는 사회적 조화와 인정에 민감해서 소위 인상관리(impression management)를 하는 경향이 있다는 것이 잘 알려져 있다(Lalwani et al. 2006). 이들은 인간관계를 고려하고 체면을 중시할 뿐 아니라 스스로를 비판적으로 인식한다. 예를 들어 일본인은 스스로를 다른 사람 보다 낮다고 평가하기보다는 상호의존적이라는 점에 더 집중한다.

이에 비해 개인주의적인 문화에서는 타인과의 비교에서 스스로를 지나치게 긍정적으로 평가하고 다른 사람, 사건, 대상 등에 대하여 아주 많이 아는 것처럼 행세한다. 이는 자기기만적 고양(self-deceptive enhancement)이라고 부르며, 무의식적으로 스스로를 과장되게 평가하는 것을 말한다. 즉 도덕적이며 이기적인 성향이 과장된 응답을 하도록 유도하는데, 집합주의적 문화에서는 도덕적 부분이 강조되고 개인주의적인 문화에서는 이기적인 부분이 강조된다(Paulhus 1991; 2002). 그러므로 집합주의적인 문화에서 사회적 규범을 측정하는 경우 응답편향은 보다 커지고 체계적일 가능성이 높다고 한다.

이러한 설명과는 다른 맥락에서 이 문제를 이해하고 검증하는 경험적 연구의 전통이 있

다. 주로 사회학이나 사회심리학의 영역에서의 가치와 행태에 관한 연구이다(Hitlin and Piliavin 2004). 여기서도 가치와 행태는 연관되어 있으나 가치가 행태의 유일한 결정요인이 아니고 가치로 행태를 추정하기도 어렵다고 한다. 행태는 하나 이상의 가치에 의해 결정되기도 하고, 상황적 요인이 가치를 압도할 수도 있다(Maio et al. 2001). 여기서 특정 사회의 문화는 개인적 가치가 행태로 표현되는 중요한 상황적 요인으로 볼 수 있다. 아울러 개인들은 강력한 인지적 기반 없이 가치를 견지하는 경우도 많다. 대부분 뻔한 소리(truism)로 치부하고 어떤 체계적인 방어논리도 없이 간직하거나 믿기 때문에 가치에 반하는 사회적 비교와 주장에 쉽게 무너진다(Maio and Olson 1998).

이렇게 가치와 행태 간의 괴리에 대해서는 잘 알려져 있지만, 특정 가치와 행태를 연계하는 연구는 많은 편이 아니다. 가치에 관한 구분이나 모형은 주로 세 가지로 구분된다(Maio et al. 2003). 마슬로우(Maslow)의 이론에 기반을 둔 잉글하트(Inglehart 1971)의 가치모형은 물질주의적 가치와 탈물질주의적 가치로 구분한다. 비슷한 맥락에서 로키치(Rokeach 1973)는 18가지의 도구적 가치와 18가지의 궁극적 가치로 구분하였고, 실증적 연구를 통해서 이들 가치를 어떻게 서열화하는가에 따라 개인들의 인권이나 종교에 대한 태도와 정치참여를 포함하는 여러 행태를 예측할 수 있는 점을 보여주었다. 이 연구와 관련하여 중요한 가치모형은 슈와르츠(Schwartz 1992)의 분류모형이다. 이는 지금까지 70개국 이상의 다양한 문화권에서 표출된 200개 이상의 표본집단들로 구성된 대규모 참여자들을 통해 검증된 가치분류이다. 이는 개인의 내재된 동기에 따라 구분되는 10개의 가치들과 이들 간의 관계를 도식적으로 보여준다: 안전(security), 동조(conformity), 전통(tradition), 박애(benevolence), 보편주의(universalism), 자율(self-direction), 자극(stimulation), 쾌락주의(hedonism), 성취(achievement), 권력(power). 이 10가지 가치들은 다시 4가지 차상위의 가치차원으로 나뉜다: 보수(conservation), 개방(openness to change), 자아증진(self-enhancement), 자아초월(self-transcendence). 즉 전통, 동조, 안전의 상위가치는 보수이며 박애와 보편주의의 상위가치는 자아초월이다. 여기서 개방은 보수와 대칭적이며, 자아증진은 자아초월과 대칭적이어서 이 중 한쪽의 가치를 증시하는 개인은 다른 가치를 등한시하는 패턴을 발견할 수 있었다.

뿐만 아니라 50개 이상의 국가에서 표본추출된 사람들의 가치 평가를 비교해보면 일관적으로 특정 가치를 가장 중요시하고 어떤 가치는 가치체계의 하위로 취급하는 경향을 알 수 있다(Schwartz and Bardi 2001). 거의 모든 사람들이 박애와 보편주의를 가장 중요한 가치로 들지만 권력, 전통, 자극은 항상 저급한 가치였다. 그럼에도 불구하고 개인이나 문

화적 차이에 따라 정도의 차이를 발견할 수 있는데, 서부유럽국가의 개인들은 다른 국가의 사람들보다 박애와 보편주의를 더 중시하고 권력을 덜 중시한다. 동아시아 사람들은 상대적으로 동조와 전통을 더 중시하는 경향이 있다.

그러므로 동아시아인들에게 동조와 전통과는 동떨어진 보편주의적 가치나 박애적 가치는 내면화하기 어려운 가치로 볼 수 있다. 어느 집단에 속해 있는가와 관계없이 모든 사람을 수용하고 관용하는 보편적 가치로서 사회적 신뢰를, 사람들에게 대한 관심과 배려를 지칭하는 박애적 가치로 자기희생적 공익추구를 들 수 있다. 이 두 가지 개념을 측정하기 위해 조사대상자에게 묻는다면 행태적 경향과는 다르게 포장하거나 도덕적 답변을 할 경향이 다분하다. 이에 비해 전통적 가치로서의 가족주의는 이러한 응답편향의 가능성이 희박하다고 볼 수 있다.

## 2. 사회조사에서의 측정결과와 상관성

위의 논의를 실제 국제적 조사결과와 비교해보면 그 가능성이 더욱 분명해진다. 예를 들어 2004년 시민권에 대한 조사(International Social Survey Program, Citizenship 2004)를 비교해보면 한국은 거의 모든 면(병역, 조세, 준법, 희생 등)에서 선진국의 평균치와 크게 다르지 않다(김승현 2010, 61). 시민성에 대한 조사 보다 더 많은 조사가 이루어진 것은 사회적 신뢰에 관한 것이다. 가장 오랜 역사를 갖고 있을 뿐 아니라, 국가 간의 비교 연구에 가장 광범위하게 쓰이는 보편적 신뢰를 측정하는 설문이 있다. 이는 반세기 이상(Rosenberg 1956) 거의 모든 서베이에서 쓰이는 질문으로 “일반적으로 대부분의 사람들을 신뢰할 수 있다고 믿는가 아니면 사람들을 조심스럽게 대해야 하는가?”(Generally speaking, do you believe most people can be trusted or can't you be too careful in dealing with people?)이다. 그리고 대부분 조사의 응답은 신뢰와 조심으로 이분적이다. 그런데 구체적인 조사결과들은 문항의 타당성을 의심하게 만든다. 1990년대 중반의 세계가치조사(WVS: World Values Survey)에 의하면 사회적 신뢰의 수준은 미국 35%, 일본 40%, 한국 30%, 대만 40% 등으로 큰 차이를 보이지 않는다(Norris 2002, Table 8-1). 이에 비하여 가장 높은 신뢰수준을 보이는 나라들은 잘 알려져 있듯이 북구국가들(스웨덴 57%, 핀란드 48%)이다. 그런가 하면 2003년의 동아시아 바로미터(EAB: East Asia Barometer)는 또 다른 결과를 제시하고 있다. 여기서 한국의 신뢰수준은 39.3%이다. 그러나 2006년의 아시아 바로미터 서베이(Asia Barometer Survey)에 의하면 신뢰한다는 응답이 한국은 66.6%이고



일본은 36.3%인데, 2003년의 조사결과에 의하면 한국 51.1%, 일본은 36.6%이다.

다시 말해서 이들 조사결과를 바탕으로 가족주의와 두 가지 규범적 개념 간의 상관성 분석을 한다면 의미 있는 부정적인 관계를 찾기 어려울 것이다. 소망성편향의 연구결과에 의하면 상관성 연구에서 한 가지 개념의 측정에 체계적인 오차가 개입된다면 억제효과가 나타나서 상관성의 크기가 실제보다 상당한 폭으로 줄어들기 때문이다(Ganster et al. 1983). 조작화와 측정의 측면을 살펴보면 이러한 가능성은 더욱 커진다. 사회적 신뢰로 지칭되는 것은 보편적 신뢰(*generalized trust*)로서 “우리와 다른 사람들에 대한 신뢰”(trust in people who are likely to be different from ourselves, rather than trust in people like ourselves)를 일컫는다(Uslaner 2004, 502). 때로는 많은 조사 자료가 사회적 신뢰를 측정하기 위하여 사회일반에 대한 신뢰정도를 묻고 있다. 그러나 사회자본론에서 지칭하는 사회적 신뢰는 모르는 사람(*strangers*)에 대한 신뢰를 지칭하는 것이지 사회일반에 대한 신뢰를 말하는 것은 아니다. 이렇게 보면 본래의 개념과 설문들이 정확하게 일치한다고 보기 어렵다. 2005년의 WVS에서 한국과 중국의 결과를 비교해보면 이러한 의문은 더욱 증폭된다. 중국의 경우 전통적인 이분적 설문에 긍정적으로 답한 것은 48.6%에 이르고 이는 2000년의 결과와 크게 다르지 않다. 한국은 28.2%이고 이 역시 앞에서 언급한 1990년대 중반이후 크게 변하지 않은 수치이다. 그런데 동일한 조사에서 처음 본 사람 또는 모르는 사람(*strangers*)에 대한 측정치는 4점 척도에서 두 가지 긍정적인 답변(*trust completely, trust a little*)을 합하면 중국은 11.3%, 한국은 14.9%로 크게 차이가 난다.

즉, 일관적이지 않은 측정치도 문제이지만 결과가 참값을 반영한다고 보기도 어렵다. 따라서 연구중심의 한국의 문화적 배경을 고려해본다면 보다 구체적으로 보편적 신뢰를 측정하는 문항을 적용하는 것이 바람직하다고 할 수 있고, 측정오차가 체계적인지 무작위적인지를 확인하는 작업이 필요하다.

### III. 실험적 조사설계

실험적 조사는 사회과학 중 주로 심리학의 연구방법이었지만 정치학을 비롯한 다른 사회과학 분야에서도 차츰 중요한 연구방법으로 간주된다(Druckman et al. 2011). 이중 열거식 실험(*list experiment*) 또는 항목수 응답형의 방법(*item count technique*)이라 칭하는 실험적 조사는 응답자들이 여러 가지 이유에서 스스로 느끼는 대로 올바른 답변을 꺼리는

사회적 이슈나 반사회적 행태에 대한 조사에서 광범위하게 사용되고 있다(Kuklinski et al. 1997; Sniderman and Carmine 1997; Streb et al. 2008). 대량의 실험적 조사는 여러 가지 장점이 있다. 이러한 조사는 실험과 서베이의 장점을 두루 갖추고 있다. 무엇보다도 실험실 내의 소수 대상자가 아닌 모집단 표본을 대상으로 조사할 뿐 아니라 작위적인 실험실적 상황이 아니라 실제 상황에서 조사한다는 점이다. 패널충원과정에서 구성원들의 개인적 특성을 어느 정도 파악하고 있기 때문에 특정하위집단을 대상으로 하는 실험에 유용하다(Mutz 2011). 특히 대립적 가설을 검증하는 데 이러한 실험설계는 지극히 유용하다. 유의미한 실험결과를 얻는 것은 F값의 공식( $F = \text{Sum of squares between cells} / \text{sum of squares within cell}$ )에서 보듯 분모보다는 분자가 커야한다. 그런데 심리학적 전통에서는 비교적 동질적인 집단을 대상으로 함으로써 분모를 최소화하는 경향이 있다. 그러나 모집단기반의 실험에서는 정반대로 비동질적 집단들을 포함시킴으로써 통계적 유의미성을 얻기 힘든 설계이다.

아울러 분모를 줄이기 위해서 실험집단에 무작위배정이 이루어지기 전에 중요한 변수로 블록킹(blocking)을 하는 방법이 있다(Mutz 2011, 95). 블록킹의 기반이 되는 변수들이 종속변수의 변이를 설명하는데 중요한 변수인 경우 오차분산(설명되지 않는 변이성)을 줄인다. 무엇보다도 무작위블럭설계(randomized-block design)는 처리 전에 선택된 독립변수에 의해 실험집단과 통제집단 사이의 무작위 차이뿐 아니라 체계적 차이도 효과적으로 줄인다. 이는 블록화하지 않은 다른 설계에 비해 통계적 추정에서 보다 효율적이다(Horiuchi et al. 2007). 따라서 이러한 설계는 실험설계의 통계적 검정력을 증진시킨다.

실험조사연구의 성패는 실험적 조사를 위한 예비조사 및 온라인 조사과정에서 표본추출과 무작위 배정에 달려 있다. 예비조사에 의한 통제문항의 선정, 층화표본추출에 의한 응답자 선정과 무작위배정, 문항의 무작위 배열 등에 유의해야 한다. 보다 구체적으로 본 조사 이전에 선행되어야 할 작업은, 천정효과나 바닥효과<sup>2)</sup>를 예방하기 위하여 응답자들 대부분이 전체적으로 부정적이거나 긍정적으로 답하지 않을 통제문항들을 선별하는 것이다. 그러기 위해서는 부정적 상관관계에 있는 문항들이 바람직하다. 설계효과(규범적 문항이 통제문항의 선택에 영향을 미침)를 피하기 위해서는 통제문항이 애매해서는 안 되고

2) 많은 응답자가 모든 문항에 대해 긍정적 또는 부정적으로 응답한다면, 실험집단의 다수 응답자가 연구관심이 되는 문항에 어떤 식으로 응답하는지를 스스로 인식함으로써 실험이 목표로 하는 솔직한 답을 하기 어렵게 한다. 이 설계의 근본가정에 대한 논의는 블레어와 이마이(Blair and Imai 2012) 참조.



응답자들의 견해가 분명한 문항이어야 한다. 이러한 중요 가정이 위배되는 일이 없도록 예비조사를 통해서 평가가 이루어져야 한다.

통제집단조사에서 쓰일 통제문항을 선별하기 위해 두 세트의 문항으로 공무원 및 대학생을 상대로 수차례 예비조사를 실시하였다. 이 조사를 통해서 통제문항이 선별되었다. 기본적으로 남녀, 20~50대까지 통계청 자료에 근거하여 층화표본추출을 하여 연령대별, 남녀별로 비례할당을 했다.<sup>3)</sup> 이는 무작위블럭설계(randomized block design)이며 8개의 블록으로 구분된 하위집단으로 구성되었다. 소득은 특별히 할당을 하지 않았으나 표본추출결과를 봤을 때 200~400만 원대가 가장 많았고, 도시 평균 가구소득이 330만 원 수준에 가까웠으므로 모집단과 유사한 분포라고 판단된다.<sup>4)</sup> 순차적으로 메일 발송을 하여 응답자가 응답을 결정하는 순간 무작위로 네 개의 집단으로 배정된다. 설문문항의 순서 역시 무작위로 제시된다. 이러한 방식으로 이 실험조사의 근본적인 가정에 충실할 수 있다.

질문은 어느 문항에 찬성하는가를 묻지 않고 몇 개의 문항에 동의하는지를 물음으로써 익명성이 보장된다. 이 실험적 조사에서 참지지도는 통제집단의 세 문항에 대한 평균을 실험집단의 네 문항에 대한 답변과 비교함으로써 계산된다. 즉 통제집단의 응답평균이 2이고 실험집단의 응답평균이 2.5라면 참지지도는 50%라고 할 수 있다. 이러한 계산이 가능한 이유는 무작위배정에 의해 집단 간 동질성이 확보되었기 때문이다. 그리고 참지지도는 평균의 차 검증에 의해 참지지도의 통계적 유의미성을 확인한다. 만약 하위집단 비교에서 통계적으로 유의미한 음의 값이 있다면 이는 근본적으로 통제문항이 실험문항에 영향을 미친 결과로 설계효과를 의심할 수 있다.

세 가지 개념에 대한 구체적인 문항의 구성에도 주의가 요구된다. 가족주의 문항은 부정적 함의를 가지면서도 규범적 응답을 유도하지 않아야 한다. 이러한 점에서 일반적으로

3) 참여율(participation rate)은 28~42%에 이르고, 답변완료율(completion rate)은 97%이다. 이 과정에서 설문을 마치는 비율이 무작위배정에 의한 집단 간 동질성 확보에 관건이 된다. 이 조사의 경우 설문이 비교적 간단한 것도 이유이겠지만, 이 비율이 지극히 높아서 문제가 없다. 해당 조사는 패널인사이드사(2012) 패널의 평균 응답 완료율을 기준으로 목표 표본크기 2400명의 3배수를 추출하였다. 총 7200명을 추출하되, 각 집단은 600명이며 총 2400명의 응답을 수집하였다. 2012년 3월 둘째 주부터 10여 일에 걸쳐 조사가 진행되었다.

4) 무작위배정이 성공적이었는지를 각 변수별로 비교하는 과정은 대부분의 결과분석에서 무의미하다는 점이 잘 알려져 있다(Mutz 2011, Cp.7). 이 조사에서는 관심이 되는 주요변수가 블록킹 변수로 이용되어 소위 무작위효과분석(randomization check)이 별 의미가 없는 대신, 하위집단 별 모집단 특성이 반영된 정도를 알아보려고 소득변수를 살펴보았다.

로 자신의 이익보다는 가족의 이익을 우선시해야 한다는 문항보다 ‘가족을 위해서는 무슨 일이든 할 수 있다’를 사용한다. 사회자본론에서 지칭하는 사회적 신뢰는 모르는 사람(strangers)에 대한 신뢰를 지칭하는 것임을 반영하여 모르는 사람에 대한 신뢰의 정도를 묻는다. 소위 사회적 신뢰는 보편적 신뢰(generalized trust)로서 우리와 다른 사람들에 대한 신뢰를 일컫는다(Uslander 2004, 502). 자기희생을 기반으로 하는 공익지향성은 자주 쓰이는 문항들과 크게 차이가 없다. 이들 세 문항이 세 가지 통제문항과 더불어 동의하는 문항의 개수를 묻는 세 실험집단의 네 가지 문항을 구성하게 된다([부록] 참조).

이러한 실험적 조사가 갖는 한계도 있다. 무엇보다도 다수의 문항을 실험문항으로 조사하기 어렵다. 하나의 문항은 하나의 실험집단을 확보해야하므로 문항이 늘어나면 조사비용도 비례해서 늘어나고 설계도 복잡해진다. 아울러 분석의 방법이 평균의 차나 비율의 차를 검증하는 것이 주된 방법일 수밖에 없다. 참지지도는 통제 및 실험집단의 평균의 차 검증에 의해 분석하고, 편향의 통계적 유의미성은 비율의 차 검증에 의존한다.<sup>5)</sup> 그러나 익명성이 보장되는 인터넷에 의한 사회조사가 전화나 면담조사보다 사회적 소망성에 의한 응답편향이 적거나 거의 없다는 사실은 기존 연구에서 여러 차례 확인되었다(Evans et al. 2003; Richman et al. 1999). 예를 들어, 사회조사에 의한 투표율은 실제 투표율보다 높다. 전화조사와 온라인조사에 의해 투표참여에 대한 열거식 실험조사를 시행한 결과 온라인 실험조사에서만 유의미한 소망성편향을 찾을 수 없었다(Holbrook and Krosnik 2010).

#### IV. 분석결과

비록 최근의 연구경향에서는 무작위배정의 효과분석(randomization check)이 불필요하다는 입장(Mutz 2011)이지만, 네 집단으로 무작위배분이 성공적인지의 여부를 파악하기 위해 중요한 인구학적 변수이며 집단배정의 기준이 되지 않았던 소득, 교육 변수를 중심으로 하위집단별 분포를 살펴보면 상당히 균등한 것을 알 수 있다.<sup>6)</sup> 이와 더불어 실험조사의 중요한 가정인 천정효과와 바닥효과가 적절히 통제될 수 있는 통제항목의 응답분포인지를 알아보았다. 예비조사를 통해서 선별된 통제문항([부록]의 <통제집단문항> 참조)에 대

5) 블레어와 이마이(2012)는 개인적인 수준에서 측정된 참지지도를 종속변수로 분석하는 방법과 통계패키지를 개발하였다. 그러나 소망성편향이 분석대상(종속변수)이면 이용할 수 없다.

6) 재분석의 목적으로 다른 연구자가 요구하는 경우 수집된 자료를 제공할 수 있다.

〈표 1〉 가족주의

	참지지도	05 한국종합사회조사 <sup>a</sup>	03 EAB <sup>b</sup>	직접설문 <sup>c</sup>
남자	0.74(0.06)***	0.77	0.58	0.77
여자	0.68(0.05)***	0.71	0.62	0.67
20대	0.66(0.08)***	0.63	0.57	0.79*
30대	0.73(0.07)***	0.71	0.58	0.72
40대	0.69(0.07)***	0.74	0.60	0.77*
50대	0.77(0.09)***	0.79	0.63	0.75
고졸 이하	0.77(0.06)***	0.78	0.62	0.79
대재 이상	0.70(0.05)***	0.69	0.57	0.73

a) 2005년 한국종합사회조사(김상욱 2005)에서는 “사람은 자기 자신보다는 가족의 안녕과 이해를 우선시해야 한다”고 묻고 7점 척도(전적으로, 상당히, 약간, 중립 등)로 답하게 함.

b) 2003년 East Asia Barometer의 경우 “가족을 위해서라면 내 이익은 둘째로 접어두어야 한다”고 묻고 중간답항 없이 4점 척도로 답하게 함. 두 조사는 동일하게 다단계 층화 무작위 표본추출로 표본크기는 1500여 명임.

c) 통제집단의 문항 중 직접설문에 대한 답이며 \*는 온라인조사에서의 소망성편향의 크기에 대한 비율의 차 검증 결과이다.

주: 한국종합사회조사, EAB의 모든 비율  $p < .001$  (표준오차생략).

괄호 안의 숫자는 표준오차이며, \*  $p < .05$ , \*\*  $p < .01$  \*\*\*  $p < .001$ .

한 분포는 지극히 이상적이었다(0, 8.2%; 1, 66.3%; 2, 22%; 3, 3.5%). 천정효과나 바닥효과를 유발할 가능성이 거의 없는 분포를 보이고 있다.

우선 세 표(〈표 1~3〉)에서 공통적으로 발견되는 것은 통제집단에 포함된 직접설문의 결과가 하위집단별로 통계적으로 유의미한 편향이 줄어들었지만 전반적으로 신뢰와 공익의 두 변수에서 더 일관적인 편향을 볼 수 있다. 이는 온라인 조사가 소망성편향의 개입이 적은 조사라는 기존 연구결과와 다르지 않다. 이를 배경으로 세 가지 개념에 대한 연구질문을 차례로 살펴보자. 가족주의의 경우 열거식 조사에 의한 참지지도는 앞에서 설명한 것처럼 통제집단과 실험집단 1(부록 참조)의 평균의 차이를 계산하였고 평균의 차 검증에 의해 통계적으로 유의미하다.

전국적 확률표본에 근거한 면담조사의 결과<sup>7)</sup> 중 2005년의 한국종합사회조사(김상욱

7) 일반조사와 온라인조사를 비교하기 위해서는 후자의 외적 타당성이 어느 정도 확보되는가가 중요하다. 교육수준이 낮고 고령자들이 조사대상에서 제외된 점을 감안하면 전체적인 표본이 모

〈표 2〉 사회적 신뢰

	참지지도	2006 KDI조사	03 EAB	직접설문 <sup>a</sup>
남자	0.19(0.06)***	0.41	0.41	0.19
여자	0.13(0.05)*	0.40	0.38	0.21**
20대	0.21(0.09)*	0.39	0.36	0.14
30대	0.15(0.07)*	0.40	0.40	0.18
40대	0.14(0.07)*	0.41	0.41	0.20
50대	0.16(0.09)	0.45	0.41	0.26*
고졸 이하	0.15(0.06)*	0.40	0.39	0.19
대재 이상	0.18(0.05)***	0.41	0.40	0.21

a) 통제집단의 문항 중 중 직접설문에 대한 답이며 \*는 온라인조사에서의 소망성편향의 유의성검증 결과이다.

주: 괄호 안의 숫자는 표준오차이며, \* p<.05, \*\* p<.01 \*\*\* p<.001

KDI, EAB의 경우 모든 비율이 p<.001 (표준오차 생략)

2005)의 문항의 하위집단 간 평균<sup>8)</sup>도 거의 비슷하고 통계적으로 의미 있는 차이를 발견할 수 없다.<sup>9)</sup> 2005년의 조사문항과 실험적 조사문항(나는 가족을 위해서는 무슨 일이든 할 수 있다) 간에는 상당한 의미의 차이나 강도가 느껴지는데도 이 정도의 근사값을 얻었다는 것은 조작화의 방향에 관계없이 전통적 가치로서 가족주의에 대한 측정이 무난하다는 것을 말해준다. 다시 말해서 소망성편향이 거의 나타나지 않는다는 사실을 알 수 있다. 이러한 점에서 가족주의라는 전통적 가치가 철저히 내재화되어 있음을 알 수 있다.

그러나 2003년 EAB의 결과는 문항의 유사성에도 불구하고 하위집단별 평균 지지도가 상당한 차이를 보이고 있다. 표본의 수도 비슷하므로 차이의 근거로 추정되는 것은 여기서

집단을 대표한다고 보기 어렵다. 하지만 주요 인구학적 변수를 기준으로 블록설계 및 층화표본 추출에 의해 통제집단과 실험집단 간의 동질성을 확보함과 동시에 전체표본의 인구특성을 잘 반영할 수 있다(Mutz 2011, Cp.8).

- 8) 세 표에서 평균값으로 제시된 것들은 전부 0-1 척도로 환산한 결과이다. 이는 다음과 같은 수식에 의한 환산이다. (응답자의 응답 - 1)/(답항수 - 1), 즉 4점 척도의 경우 응답자가 반대이면 (2-1)/(4-1)로 변환된다. 그러므로 1, 2, 3, 4의 4점 척도는 0, 1/3, 2/3, 1의 척도로 바뀐다. 즉 비교가능한 척도로 변환되었고 측정결과가 바뀐 것은 아니다.
- 9) 각 조사에서 계산된 비율과 참지지도와의 차이를 응답편향으로 간주하여 통계적 검증을 하였다. 예를 들어 실험적 조사의 20대는 150명이었고 2005 한국종합사회조사(김상욱 2005)의 20대는 283명이며 이를 기반으로 비율의 차 검증이 이루어졌다.

조사문항은 중간응답이 생략된 4점 척도였다는 사실뿐이다. 그렇지만 <표 2>의 사회적 신뢰의 결과 중 전국적 확률표본을 기반으로 조사된 두 결과가 응답항의 차이에도 불구하고 거의 동질적인 점을 고려하면 2003년 EAB의 가족주의 측정에 문제점이 있다고 생각된다. 즉 여러 측면에서 2003년의 가족주의 조사결과(EAB)를 신뢰하기 어렵다.<sup>10)</sup>

사회적 신뢰의 측정 결과를 보면 실험적 조사에 의한 참지지도는 다른 두 가지 면담조사와는 상당한 차이를 보이고 있어서 사회적 소망성의 개입이 두드러진다고 할 수 있다.<sup>11)</sup> 여기에 인용된 두 결과는 사회적 신뢰에 대한 최근 조사를 포함하는 여러 가지 조사결과를 감안한 것이다. 종교와 환경에 대한 2008년과 2010년의 국제사회조사(International Social Survey Programme 2008; 2010)에 삽입된 신뢰에 관한 문항은 “귀하는 일반적으로 사람들을 신뢰할 수 있다고 생각하십니까, 아니면 조심해야 된다고 생각하십니까”로 묻고 4점 척도로 답하게 하였다. 두 결과는 40.3%와 41.5%의 응답자가 긍정적 답변을 하였다. 이는 약 40%의 결과가 일반적일 뿐 아니라 어느 정도 긴 시간이 흐른 뒤에도 신뢰의 정도가 크게 변하지 않음을 보임으로써 2012년 실시된 온라인조사와의 직접비교에 무리가 없음을 보여주고 있다.<sup>12)</sup>

2006년 한국개발원(KDI)조사(김태종 외 2006)의 문항은 처음 만난 사람에 대한 신뢰를 묻고 0~10의 11점 척도로 응답하게 하였고, 2003년 EAB 조사는 가장 오랫동안 쓰였던 문항(Rosenberg 1956)인 “대부분의 사람들이 믿을 수 있다고 생각하십니까? 아니면 사람들을 대할 때 조심할수록 좋다고 생각하십니까?”라는 문항을 이분적으로 답하게 한 것이다. 이렇게 문항과 답항이 다름에도 불구하고 이들 하위집단별 평균점은 놀랍게 유사하다. 가족주의도 그렇지만 다양한 조작화방법이 크게 다르지 않은 결과를 보여주고 있다. 그럼에

10) 이를 뒷받침하는 것으로 이 조사에서 가족주의와 공익지향성의 상관계수는 0.244이지만 다른 조사결과에서 상응하는 양의 상관성을 찾기 어렵다. 통제집단에 삽입된 직접문항의 경우([부록] 참조), 두 변수 간의 상관성은 0.113이다. 이는 온라인조사에서는 상대적으로 소망성편향에 의한 긍정적 응답이 비교적 작기 때문이며, 면대면 조사에서는 편향의 크기가 크기 때문에 상관성도 더 부풀려진 결과임을 보여주고 있다. 사회적 신뢰와의 관계는 통계적으로 유의미하지 않지만 부호가 바뀐다.

11) 각주 9)에서 설명한 방법으로 비율의 차 검증결과 모든 하위집단에서 통계적으로 유의미한 차이를 보였다.

12) 연구목적에 의해 사회조사결과를 선별적으로 인용하였다는 비판을 피하기 위해, 이들 조사를 하위집단별로 분석해 보아도 유사한 결과를 얻는다. 2003 EAB 조사를 인용한 이유는 상관관계 분석의 문제점을 보이기 위해 필요한 조사결과이기 때문이다.

〈표 3〉 공익지향성

	참지지도	2006 갤럽옴니버스조사 <sup>a</sup>	2003 EAB <sup>b</sup>	직접설문 <sup>c</sup>
남자	0.35(0.06)***	0.57	0.52	0.44*
여자	0.20(0.06)***	0.49	0.50	0.24
20대	0.34(0.10)***	0.47	0.49	0.25
30대	0.14(0.07)*	0.50	0.49	0.25**
40대	0.29(0.07)***	0.50	0.50	0.28
50대	0.39(0.09)***	0.59	0.53	0.61*
고졸 이하	0.28(0.06)***	0.49	0.51	0.34
대재 이상	0.29(0.06)***	0.47	0.50	0.34

a) 2006년 갤럽옴니버스조사는 중간응답이 없는 4점 척도의 응답항으로 구성됨.

b) 2003 EAB의 경우도 4점 척도로 응답하게 함.

c) 통제집단의 문항 중 직접설문에 대한 답이며 \*는 온라인조사에서의 소망성편향의 유의성검증 결과이다.

주: 갤럽옴니버스조사, EAB의 모든 비율  $p < .001$  (표준오차생략)

괄호 안의 숫자는 표준오차이며, \*  $p < .05$ , \*\*  $p < .01$  \*\*\*  $p < .001$

도 불구하고 소망성편향은 통계적으로 유의미하고 하위집단별로도 상당한 차이를 보이고 있다. 특히 여성과 연령층이 높아질수록 상대적으로 큰 소망성편향을 보인다.

다시 말해서 사회적 신뢰의 문항에 대한 응답은 상당부분 규범적 차원에서 답하고 응답자 자신의 행태적 측면을 내보이고 있지 않다고 말할 수 있다. 다만 답항의 구성이 어떤 식이든 간에 그 중간응답경향에 의한 영향은 무시할만하다는 점도 알 수 있다.<sup>13)</sup> 여기서 알 수 있는 점은 면담조사에 의해 현상을 이해하는 것도 잘못될 뿐 아니라 상관성 추정도

13) 응답세트에 대한 또 하나의 중요한 연구결과는 극단적이거나 중도적인 응답성향에 관한 것이다. 기존연구에 의하면 한국·일본·중국은 중도적 응답을 하려는 경향이 강하다. 리와 그린(Lee and Green 1991)은 한국인들은 극단적 응답을 회피하고 중간 척도를 선호하는 경향이 있음을 밝히고 있다. KDI의 사회조사(김태종 외 2006)에서도 이러한 경향을 알 수 있다. 일반적으로 쓰이는 신뢰에 관한 문항뿐만 아니라, 보편적 신뢰를 세부적으로 질문하고 있다. 처음 만나는 사람, 출신지역이 다른 사람, 종교가 다른 사람, 정치적 견해가 다른 나라 사람, 세대가 다른 사람 등에 대한 신뢰로 구분하여 묻는다. 0점에서 10점까지의 11점 척도로 측정하는데 중립적 태도(5)를 취하는 응답자가 23.3%에서 35.8%에 이르는 점을 감안하면, 이들이 이분적 질문에 어떻게 응답하는가에 따라 신뢰의 측정결과나 수준은 크게 변할 가능성을 엿볼 수 있다.



잘못될 가능성이 크다는 것이다. 예를 들어 사회조사의 결과만을 보면 사회적 신뢰의 수준이 남녀 간에 큰 차이가 없는 것처럼 보일는지 모르지만 실제로는 여성의 소망성편향의 크기가 크기 때문이다. 뿐만 아니라 나이가 들수록 사회적 신뢰의 수준도 높아지는 것처럼 보이지만 실제로는 그 수준이 낮아지고 소망성편향만 커지는 사실을 확인할 수 있다. 이는 상관성추론에도 문제를 야기한다. 만약 소망성편향이 모든 하위집단에서 일정한 크기로 나타난다면 상관성 추론에는 문제가 없다. 그러나 여성이거나 나이가 들수록 소망성편향이 더 커지고 실제 수준은 더 낮다면 사회조사에 의한 상관성 추론은 크게 과장되어 때로는 음의 관계가 양의 관계로 추정될 가능성을 배제하기 어렵다.

자기희생에 기초한 공익지향에 관한 조사의 결과도 그 패턴은 사회적 신뢰의 측정에서와 유사하다. 즉 면담조사에서는 상당히 부풀려진 결과를 보이고 있어서 사회적 소망성편향이 체계적인 측정오차의 원인임을 알 수 있다. 여성과 30대나 그 이상 연령층에서 소망성에 의한 응답편향이 두드러진다. 2006년 갤럽옵니버스조사(한국갤럽조사연구소 2006)와 2003년 EAB의 조사문항을 엄격하게 비교하면 그 의미가 상당히 다른데도 불구하고 유사한 결과를 얻었다. 갤럽조사(2006)의 경우 “나는 국가를 위해서라면 나 자신의 이익을 희생할 수 있다”로 물음으로써 응답자의 개인의 행태적 측정에 가깝다. EAB조사(2003)의 경우, “개인은 사회전체를 위해 자신의 이익을 희생할 각오가 되어 있어야 한다”고 물음으로써 응답자들이 규범적 가치로서 응답하려는 경향을 보일 수 있다. 그렇지만 응답의 결과는 유사하다.

결론적으로 문항이 조금씩 바뀌어도 측정결과에 큰 영향을 미치지 않지만 사회적 규범 또는 시민적 가치로서 사회적 신뢰나 공익지향성과 같은 개념의 측정에는 사회적 소망성이 개입하고 있음을 알 수 있다. 사회적 규범의 측정에서 사회적 소망성편향이 체계적 오차로 작용하고 있음을 알 수 있다. 따라서 이렇게 체계적으로 부풀려진 측정결과를 바탕으로 현상을 이해하는 것과 상관성을 가늠하는 것 모두 잘못된 추론이 될 가능성이 높다.

사회조사에서 두 가지 규범적 가치의 구성개념 타당성을 의심하지 않을 수 없다. 구성개념의 타당성이 확보되지 못한다는 지적은 개념의 측정과정에서 어떤 심각한 오염에 의해 체계적 측정오차가 발생할 경우에 생기는 문제를 일컫는다(Chan 2009). 사회적 소망성이 체계적으로 개입하게 되면 구성개념에 의한 진정한 변이에 못지 않거나 그보다 큰 소망성에 의한 변이가 커짐으로써 측정의 타당성을 확보할 수 없다. 이러한 문제점은 측정대상이 되는 변수나 개념과 사회적 소망성이 어떤 관계를 보이는가에 달렸는데 이상의 조사결과에서는 하위집단별로 체계적인 소망성편향을 파악할 수 있었다.

구성개념의 타당성의 문제는 곧바로 개념 간의 상관성 추론이 타당한가의 문제로 직결된다. 상관성의 크기는 개념 간의 진정한 상관성과 측정방법에 의해 야기되는 상관성의 두 부분으로 구분될 수 있는데 후자의 크기가 측정방법에 의해 부풀려져 있기 때문에 추정된 상관성의 크기가 부풀려진다는 논리이다. 사회적 규범인 두 개념의 측정결과가 모든 하위집단에서 일정한 폭으로 과장되었다면 현상에 대한 기술적 이해(descriptive understanding)는 잘못되지만, 가족주의와 같은 전통적 가치와의 상관성 추리에 영향을 미치지 않는다. 그러나 표에서 보듯 편향의 폭은 하위집단별로 다르고 일정한 패턴을 보인다. 이는 기술적 추론과 인과적 추론 모두에 문제가 생긴다는 것을 의미한다(King et al. 1994, 156). 보다 구체적으로 독립변수와 종속변수에 이러한 편향이 어떤 방식으로 나타나는가에 따라 상관관계 분석이 여러 방향으로 잘못될 가능성이 높다(Ganster et al. 1983). 만약 양자의 측정에 공히 큰 편향이 개입된다면 이 때문에 공유된 분산에 의해 상관성이 부풀려지고, 그 결과 허위관계가 나타날 우려가 있다(spurious effect). 이러한 현상이 가장 빈번하다고 생각된다. 2003년 EAB의 조사를 기반으로 하는 가족주의와 공익추구의 상관성 분석결과(0.244)가 이러한 유형의 오류를 암시하고 있다. 논리적으로 부정적인 양자의 관계가 상당한 크기의 정의 관계인 것은 사회적 소망성편향이 두 변수에 공히 작용함으로써 소위 허위관계의 가능성을 엿볼 수 있다. 또 하나는 두 가지 개념에서 하나에서만 편향이 크고 다른 변수에는 거의 나타나지 않는다면 상관성 추정은 실제보다 작게 될 것이다(suppression effect).<sup>14)</sup>

전통적 가치와 사회적 규범 간의 관계를 다변량분석으로 이해하는 경우가 많은 점을 감안할 때 주로 사회적 규범의 측정에 소망성편향이 크게 작용하므로 억제효과가 두드러진다고 말할 수 있다. 이러한 맥락에서 사회조사를 바탕으로 하는 많은 계량적 연구에서 전통적 가치의 역할이 부정적이지 아니라는 결론은 이 억제효과에 기인할 수 있다.

## V. 논의

사회조사에 의한 자료의 성격 중 소망성편향에 대한 방법론적 논란은 오랫동안 계속되어왔다. 방법론적 논쟁의 한편에서는 개념의 측정과 이를 기반으로 하는 상관성 분석의 타

14) 이러한 두 가지 효과 이외에 좀 더 복잡한 관계는 조절효과(moderator effect)이다.

당성을 의심한다. 그렇기 때문에 다른 방식의 측정에 의해 이를 검증할 것을 요구하기도 한다. 이에 비해 다른 한편에서는 이러한 회의적 시각이 왜곡·과장되었다는 의미에서 신화 또는 도시전설로 간주하기도 한다(Chan 2009).

측정자료에 관한 비판은 크게 두 가지 점을 지적하고 있다. 하나는 체계적 측정오차에 의한 구성개념의 타당성을 확보할 수 없다는 점이고, 또 하나는 구성개념 간의 상관성의 크기를 제대로 추정할 수 없다는 점이다. 전자와 관련된 문제점은 1940년대 이래로 지적되어 온 사항이지만, 공통방법분산에 관한 논의는 캠벨과 피스크(Campbell and Fiske 1959)에서 시작되었다. 이 설명에 의하면 동일방법으로 측정된 양 개념 간의 상관성이 양자 간의 진정한 관계를 추정한다기 보다는 동일측정방법에서 유발된 체계적 측정오차에 의한 통계적 허위에 그칠 우려를 지적한다.

먼저 구성개념의 타당성이 확보되지 못한다는 비판에는 개념의 측정과정에서 어떠한 심각한 오염에 의해 체계적 측정오차가 발생할 경우에 생기는 문제이다. 사회적 소망성이 체계적으로 개입하게 되면 구성개념에 의한 진정한 변이에 못지 않거나 그보다 큰 소망성에 의한 변이가 커짐으로써 측정의 타당성이 훼손된다는 주장이다. 이러한 문제점은 측정대상이 되는 변수나 개념과 사회적 소망성이 어떤 관계를 보이는가에 달렸다. 이에 반하는 입장에서도 이러한 가능성을 인정하고 있다. 그렇지만 이러한 가능성은 특정 변수나 개념에 한정되어있고, 다양한 개인성·조직형태 연구에서 수렴적 타당성이나 차별적 타당성이 입증되었다는 사실을 반증으로 들고 있다(예: Vandewalle 1997).

구성개념 타당성의 문제는 곧바로 개념 간의 상관성 추론이 타당한가의 문제로 직결된다. 상관성의 크기는 개념 간의 진정한 상관성과 측정방법에 의해 야기되는 상관성의 두 부분으로 구분될 수 있는데, 후자의 크기가 측정방법에 의해 부풀려져 있기 때문에 추정된 상관성의 크기가 부풀려진다는 논리이다. 그렇지만 이 주장을 과장된 것으로 보는 입장에서는 이러한 가능성을 부인한다. 구성개념의 측정에 있어서 무작위적 측정오차가 항상 있을 수 있고 이 때문에 완벽한 신뢰성을 전제할 수 없다면 진정한 상관관계의 크기보다 추정치는 항상 적을 것이다. 그렇기 때문에 상관성의 추정이 부풀려질지, 더 적은 것으로 보일지, 정확하게 추정될지 모른다(Chan 2001). 만약 하나의 개념에서만 소망성편향이 크다면 상관성의 크기가 실제보다 더 적게 추정될 것이다. 소망성편향이 심각하게 작용하는 경우에도 이를 통제할 다양한 방법이 있다는 것을 강조한다. 여기에는 이러한 편향을 피할 수 있는 척도의 개발(Cheung and Chan 2002), 측정과정에서 소망성편향을 직접 추정하는 노력 등을 들 수 있고, 소망성편향에 의해 부풀려진 상관성의 크기는 공통방법에 의한 변

이를 잠재변수로 보고 통제(Podsakoff et al. 2003)하는 등, 통계적 방법에 의해 수정될 수 있다고 생각한다.

이러한 논란을 종합해보면 구성개념의 타당성이나 정확한 상관관계의 추정은 상당히 경험적 문제로 볼 수 있다. 즉 사회적 소망성에 의한 문제의 심각성은 어떤 연구주제에서 어떤 변수 간의 관계를 추론하는지에 달렸고, 구체적으로는 이들 개념이나 변수의 측정에 소망성편향이 얼마나 개재되는지를 직접적으로 조사할 필요가 있다.

그러나 분명하게 바람직한 사회적 규범이나 기준이 존재하고 광범위하게 받아들여지는 경우에는 사회적으로 바람직한 응답의 가능성은 더욱 높아진다는 사실은 널리 알려져 있다. 이 연구결과에서 알 수 있는 것처럼, 전통적 가치가 사회적 규범의 수용에 어떠한 영향을 미치는지 알고자 하는 상관성 연구에서는 후자의 가치측정에 나타난 체계적인 측정오차가 정확한 이해를 방해할 것이라는 점은 분명하다. 물론 더 정확한 측정을 전제로 한 상관성 연구가 필요하지만, 전통적 가치의 영향에 관한 대부분의 연구에서 부정적 상관관계를 찾기 어려운 점은 이러한 측정오차에서 기인한다고 생각된다.

이러한 결과를 보고 고민해야 될 새로운 연구과제는 바로 이러한 편향을 제어할 수 있는 방법을 알아보는 것이다. 지금까지 연구된 사회적 소망성편향의 처리방법으로는, (1) 가치(should be)와 행태(as is) 차원의 구분 측정, (2) 선택형 측정방식(ipsative measures: Baron 1996), (3) 사회적 소망성 척도(social desirability scale: Edwards Social Desirability Scale, Marlowe-Crowne Scale 등) 활용, (4) 통계적 처리방법(common method bias 처리) 등이 개발되었다. 이중 통계적 통제방법 역시 억제효과가 작용하고 있는 상황에서는 별로 적절하지 않다. 나머지 세 가지 방법을 시도해야 할 필요가 있다고 생각되지만 이들 방법들은 모두 나름대로의 약점을 가지고 있다. 조사해야 할 설문문항이 증가하여 조사 자체를 힘들게 만들거나 통계처리가 곤란하여 실제 파악하려는 변수 간 관계를 검증하기 어렵다는 등의 문제점들을 안고 있다. 이러한 문제를 감안하여 사회과학 분야의 경험적·실증적 연구에서 사회적 소망성편향을 적절하게 처리할 수 있는 방안을 찾아야 한다.

## VI. 결론

지금까지 아시아적 가치, 한국의 민주주의, 정치문화, 행정문화, 공공봉사동기(Public Service Motivation) 등에 대한 솔한 여론조사가 이루어졌고 이를 기반으로 비교적 시각에

서 많은 저술이 쏟아져 나왔다. 특히 전통적 가치와 시민적 가치 또는 사회적 규범 간의 상관성 연구에서는 대체적으로 계량적 연구가 긍정적 평가(예: Dalton and Shin 2006)를 하는 반면, 질적 연구는 부정적 평가를 하고 이러한 두 주장의 간극은 좀처럼 좁혀지지 않고 있다. 이 연구의 결과가 사회적 소망성편향에 의한 체계적 측정오차가 있다는 것을 밝힘으로써 앞으로의 여론조사나 이에 대한 양적 분석과 비교가 더 조심스럽게 진행되어야 된다는 사실을 보여주었다.

우리는 전통적 가치에 정서적으로 동의하지만 사회적 또는 시민적 가치를 묻는 문항에 대해서는 사회적으로 바람직한 답변을 제시하는 경향이 있는 것처럼 보인다. 문제는 여기서 야기되는 응답편향으로 인해 사회조사에 의한 측정치들의 신뢰성과 타당성이 문제가 되므로 이를 기초로 한 계량적 분석의 신뢰성을 크게 훼손하게 된다. 이러한 현상을 감안한 보다 치밀한 사회조사와 연구가 필요하다. 특히 동일한 설문으로 여러 국가들을 조사하고 이를 바탕으로 계량적 분석을 하는 것은 보다 신중해야 할 필요성도 암시하고 있다. 면담조사를 바탕으로 한 단순한 국제비교조사에서 체면을 중시하는 동양유교권의 조사와 결과의 분석은 더욱 신중해야 한다. 따라서 국제적 비교조사 이전에 각 국가에서 사회적 소망성에 의한 응답편향의 크기를 짐작할 수 있는 비교연구가 필요하다고 생각된다.

연구수행의 주체가 달라지거나 연구시기가 달라질 뿐, 유사한 조사들이 많이 이루어지고 있다. 정치문화, 시민의식 또는 시민사회를 주제로 하는 다양한 조사들이 이루어지고 있으나 문항이나 조사방법에 대한 깊은 고민 없이 시행되는 경우도 적지 않다고 할 수 있다. 이 연구는 이러한 폐해와 낭비를 줄이고 사회조사의 수준을 제고하는 새로운 접근방법을 도모해야 할 필요성을 강조하는 것이다.

투고일 2013년 4월 15일

심사일 2013년 5월 9일

게재확정일 2013년 5월 31일

## 참고문헌

### 〈사회조사 자료〉

- 김상욱. 2005. 『한국종합사회조사, 2005』. 연구수행기관: 성균관대학교 서베이리서치센터. 자료서비스기관: 한국사회과학자료원. 자료공개년도: 2006년. 자료버전: v2. 자료번호: A1-2005-0001.
- 김태중 외. 2006. 『사회적 자본확충을 위한 기본조사 및 정책연구』. KDI국제정책대학원 보고서.
- 한국갤럽조사연구소. 2006. 『갤럽옵니버스조사, 2006(11월)』. 자료서비스기관: 한국사회과학자료원. 자료공개년도: 2009년. 자료번호: A1-2006-0099.
- 패널인사이트. 2012. <http://www.panelinsight.co.kr/> (검색일: 2013. 05. 31).
- Asia Barometer Survey(ABS). 2006. <http://www.asianbarometer.org/newenglish/Introduction/Surveytimetable.htm> (검색일: 2013. 05. 31).
- East Asia Barometer(EAB). 2003. <http://www.jdsurvey.net/eab/eab.jsp> (검색일: 2013. 03. 30).
- International Social Survey Programme. 2004. <http://www.issp.org/> (검색일: 2013. 03. 30).
- \_\_\_\_\_. 2008 <http://www.issp.org/> (검색일: 2013. 03. 30).
- \_\_\_\_\_. 2010. <http://www.issp.org/> (검색일: 2013. 03. 30).
- World Values Survey(WVS). 1995. [http://www.worldvaluessurvey.org/index\\_surveys](http://www.worldvaluessurvey.org/index_surveys) (검색일: 2013. 03. 30).
- \_\_\_\_\_. 2000. [http://www.worldvaluessurvey.org/index\\_surveys](http://www.worldvaluessurvey.org/index_surveys) (검색일: 2013. 03. 30).
- \_\_\_\_\_. 2005. [http://www.worldvaluessurvey.org/index\\_surveys](http://www.worldvaluessurvey.org/index_surveys) (검색일: 2013. 03. 30).

### 〈연구논문 및 단행본〉

- 김승현. 2010. “가족주의와 공공성: 전통적 가치와 현대적 가치의 측정에서 사회적 소망성편향.” 『한국정치학회보』 44집 3호, 53-75.
- Almond, Gabriel A. and Sidney Verba. 1963. *The Civic Culture: Political Attitudes and Democracy in Five Nations*. Newbury Park: Sage.
- \_\_\_\_\_. eds. 1980. *The Civic Culture Revisited: An Analytic Study*. Boston: Little Brown.
- Bardi, Anat and Shalom H. Schwartz. 2003. “Values and Behavior: Strength and Structure of Relations.” *Personality and Social Psychology Bulletin* 29. No.10, 1207-1220.
- Baron, Helen. 1996. “Strengths and Limitations of Ipsative Measurement.” *Journal of Occupational and Organizational Psychology* 69. No.1, 49-56.



- Blair, Graeme and Kosuke Imai. 2012. "Statistical Analysis of List Experiment." *Political Analysis* 20. No.1, 47-77.
- Campbell, Donald T. and Donald W. Fiske. 1959. "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix." *Psychological Bulletin* 56. No.2, 81-105.
- Chan, David. 2009. "So Why Ask Me? Are Self-Report Data Really That Bad?" In Charles E. Lane and Robert J. Vandenberg, eds. *Statistical and Methodological Myths and Urban Legends*, 309-336. New York: Routledge.
- \_\_\_\_\_. 2001. "Method Effects of Positive Affectivity, Negative Affectivity, and Impression Management in Self-Reports of Work Attitudes." *Human Performance* 14. No.1, 77-96.
- Cheung, Mike W.-L. and Wai Chan. 2002. "Reducing Uniform Response Bias With Ipsative Measurement in Multiple-Group Confirmatory Factor Analysis." *Structural Equation Modeling* 9. No. 1, 55-77.
- Cronbach, Lee J. 1946. "Response Sets and Test Validity." *Educational and Psychological Measurement* 6. No.3, 475-494.
- Crowne, Douglas P. and David Marlow. 1964. *The Approval Motive*. New York: John Wiley & Sons.
- Dalton, Russell J. and Doh C. Shin, eds. 2006. *Citizens, Democracy and Markets around the Pacific Rim: Congruence Theory and Political Culture*. Oxford: Oxford University Press.
- Druckman, James N., Donald P. Green, James H. Kuklinski, and Arthur Lupia. 2006. "The Growth and Development of Experimental Research in Political Science." *American Political Science Review* 100. No.4, 627-36.
- Edwards, Allen L. 1957. *The Social Desirability Variable in Personality Assessment and Research*. New York: Dryden.
- Evans, David C. et al. 2003. "In the Privacy of Their Own Homes: Using the Internet to Assess Racial Bias." *Personality and Social Psychology Bulletin* 29. No.2, 273-284.
- Ganster, Daniel C., Harry W. Hennessey, and Fred Luthans. 1983. "Social Desirability Response Effects: Three Alternative Models." *Academy of Management Journal* 26. No. 2, 321-331.
- Hitlin, Steven and Jane Allyn Piliavin. 2004. "Values: Reviving a Dormant Concept." *Annual Review of Sociology* 30, 359-93.
- Holbrook, Allyson L. and Jon A. Krosnick. 2010. "Social Desirability Bias in Voter Turnout Reports: Tests using the Item Count Technique." *Public Opinion Quarterly* 74. No.1, 37-67.
- Horiuchi, Yusaku, Kosuke Imai, and Naoko Taniguchi. 2007. "Designing and analyzing randomized experiments: Application to a Japanese election survey experiment." *American*

- Journal of Political Science* 51. No.3, 669-687.
- Inglehart, Ronald. 1971. "The Silent Revolution in Europe: Intergenerational Change in Post-Industrial Societies." *American Political Science Review* 65. No.4, 991-1017.
- King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry*. Princeton: Princeton University Press.
- Kuklinski, James H., Michael D. Cobb, and Martin Gilens. 1997. "Racial attitudes and the 'New South.'" *Journal of Politics* 59. No.2, 323-349.
- Lalwani, Ashok K., Sharon Shavitt, and Timothy Johnson. 2006. "What is the Relation between Cultural Orientation and Socially Desirable Responding?" *Journal of Personality and Social Psychology* 90. No.1, 165-178.
- Lee, Chol and Robert T. Green. 1991. "Cross-Cultural Examination of the Fishbein Behavioral Intentions Model." *Journal of International Business Studies* 22. No.2, 289-305.
- Maior, Gregory R. et al. 2001. "Addressing Discrepancies between Values and Behavior: The Motivating Effect of Reasons." *Journal of Experimental Social Psychology* 37. No.2, 104-117.
- \_\_\_\_\_. 2003. "Ideologies, Values, and Behavior." In J. DeLamater, ed. *Handbook of Social Psychology*. New York: Plenum.
- Maior, Gregory R. and James M. Olson. 1998. "Values as Truism: Evidence and Implications." *Journal of Personality and Social Psychology* 74. No.2, 294-311.
- Mutz, Diana C. 2011. *Population-Based Survey Experiments*. Princeton: Princeton University Press.
- Norris, Pippa. 2002. *Democratic Phoenix: Reinventing Political Activism*. New York: Cambridge University Press.
- Paulhus, Delroy L. 1991. "Measurement and Control of Response Bias." In J. P. Robinson, P. Shaver, and L. S. Wrightsman, eds. *Measures of Personality and Social Psychological Attitudes*, 17-60. San Diego: Academic Press.
- \_\_\_\_\_. 2002. "Socially Desirable Responding: The Evolution of a Construct." In Henry I. Braun et al., eds. *The Role of Constructs in Psychological and Educational Measurement*, 51-76. Mahwah, NJ: Lawrence Erlbaum Associates.
- Podsakoff, Philip M., Scott B. MacKenzie, and J. Lee. 2003. "Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies." *Journal of Applied Psychology* 88. No.5, 879-903.
- Richman, Wendy L., Sara Kiesler, Suzanne Weisband, and Fritz Drasgow. 1999. "A Meta-

- Analytic Study of Social Desirability Distortion in Computer-Administered Questionnaires, Traditional Questionnaires, and Interviews.” *Journal of Applied Psychology* 84. No.5, 754-775.
- Rokeach, Milton. 1973. *The Nature of Human Values*. New York: Free Press.
- Rosenberg, Morris. 1956. “Misanthropy and Political Ideology.” *American Sociological Review* 21. No.6, 690-695.
- Schuluf, Boaz, John Hattie, and Robyn Dixon. 2008. “Factors Affecting Responses to Likert Type Questionnaires: Introduction of the ImpExp, a New Comprehensive Model.” *Social Psychology of Education* 11. No.1, 59-78.
- Schwartz, Shalom H. 1992. “Universals in the Content and Structure of Values: Theoretical Advances and Empirical Tests in 20 Countries.” In Mark P. Zanna, ed. *Advances in Experimental Social Psychology* 25, 1-66. New York: Academic Press.
- Schwartz, Shalom H. and Anat Bardi. 2001. “Value Hierarchies across Cultures: Taking a Similarities Perspective.” *Journal of Cross Cultural Psychology* 32. No.3, 268-290.
- Sniderman, Paul M. and Edward G. Carmines. 1997. *Reaching beyond Race*. Cambridge: Harvard University Press.
- Streb, Matthew J., Barbara Burrell, Brian Frederick, and Michael A. Genovese. 2008. “Social Desirability Effects and Support for a Female American President.” *Public Opinion Quarterly* 72. No.1, 76-89.
- Triandis, Harry C. 1994. “Cross-Cultural Industrial and Organizational Psychology.” In H.C. Triandis, M.D. Dunnette, and L.M. Hough eds. *Handbook of Industrial and Organizational Psychology* 4 (2nd ed.). Palo Alto, CA: Consulting Psychologists Press.
- Triandis, Harry C. 1995. *Individualism and Collectivism*. Boulder: Westview Press.
- Uslaner, Eric. 2004. “Trust and Social Bonds: Faith in Others and Policy Outcomes Reconsidered.” *Political Research Quarterly* 57. No.3, 501-507.
- Vandewalle, Don. 1997. “Development and Validation of a Work Domain Goal Orientation Instrument.” *Educational and Psychological Measurement* 57. No.6, 995-1015.

## 부록: 집단별 문항구성

### 통제집단

I. 아래 문항은 우리 사회가 당면한 문제에 대한 견해를 묻는 질문입니다. 어떤 항목에 동의하시는지는 저희가 알 필요가 없습니다. 다음 세 개의 문항에서, 응답하시는 분이 동의하거나 지지하는 내용을 담은 **문항의 개수(0, 1, 2 또는 3)**를 다음 괄호 안에 써 넣어 주시기 바랍니다. ( )

- 휘발유나 경유 값을 더 올려도 좋다.
- 나는 정부의 잘못된 교육정책이 교육현실을 더 악화시켰다고 생각한다.
- 우리나라는 미국 보다는 중국과 가깝게 지내야 된다.

II. 아래 문항들은 응답하시는 분의 견해에 대한 질문입니다.

1. 나는 가족을 위해서라면 무슨 일이든 할 수 있다.  
1) 그렇다 2) 아니다
2. 나는 모르는 사람도 신뢰하는 편이다.  
1) 그렇다 2) 아니다
3. 나는 사회를 위하여 기꺼이 개인적 희생을 감수할 수 있다.  
1) 그렇다 2) 아니다

### 실험집단 1.

아래 문항은 우리 사회가 당면한 문제나 개인적 태도에 대한 견해를 묻는 질문입니다. 어떤 항목에 동의하시는지는 저희가 알 필요가 없습니다. 단지 다음 네 개의 문항에서, 응답하시는 분이 동의하거나 지지하는 내용을 담은 **문항의 개수(0, 1, 2, 3 또는 4)**를 다음 괄호 안에 써 넣어 주시기 바랍니다. ( )

- 휘발유나 경유 값을 더 올려도 좋다.
- 나는 정부의 잘못된 교육정책이 교육현실을 더 악화시켰다고 생각한다.
- 우리나라는 미국 보다는 중국과 가깝게 지내야 된다.
- 나는 가족을 위해서라면 무슨 일이든 할 수 있다.

### 실험집단 2.

아래 문항은 우리 사회가 당면한 문제나 개인적 태도에 대한 견해를 묻는 질문입니다. 어떤 항목에 동의하시는지는 저희가 알 필요가 없습니다. 단지 다음 네 개의 문항에서, 응답하시는 분이 동의하거나 지지하는 내용을 담은 **문항의 개수(0, 1, 2, 3 또는 4)**를 다음 괄호 안에 써 넣어 주시기 바랍니다. ( )

- 휘발유나 경유 값을 더 올려도 좋다.
- 나는 정부의 잘못된 교육정책이 교육현실을 더 악화시켰다고 생각한다.
- 우리나라는 미국 보다는 중국과 가깝게 지내야 된다.
- 나는 모르는 사람도 신뢰하는 편이다.

### 실험집단 3.

아래 문항은 우리 사회가 당면한 문제나 개인적 태도에 대한 견해를 묻는 질문입니다. 어떤 항목에 동의하시는지는 저희가 알 필요가 없습니다. 단지 다음 네 개의 문항에서, 응답하시는 분이 동의하거나 지지하는 내용을 담은 **문항의 개수(0, 1, 2, 3 또는 4)**를 다음 괄호 안에 써 넣어 주시기 바랍니다. ( )

- 휘발유나 경유 값을 더 올려도 좋다.
- 나는 정부의 잘못된 교육정책이 교육현실을 더 악화시켰다고 생각한다.
- 우리나라는 미국 보다는 중국과 가깝게 지내야 된다.
- 나는 사회를 위하여 기꺼이 개인적 희생을 감수할 수 있다.

ABSTRACT

---

## Measurement of Social Norms: An Experimental Study of Response Bias

**Seung Hyun Kim** | Seoul National University of Science and Technology

Social science is heavily dependent on social survey as a data collection method. However, it is not easy to get frank answers from the respondents. They tend to respond with what is socially desirable and with neutral or extreme answers rather than what they truly feel or behave. These problems are usually termed response bias or response set. To analyze the problem of response set, this study compares such social norms as social trust and self sacrifice for public interest with traditional values such as familism. The experimental method called list experiment or item count technique is employed to figure out true support for these norms and value. The results are compared with those from the nationwide probability sampling surveys. The measurements of social norms show the systematic errors while those of familism show little bias. This means that the descriptive understanding of social norms and the correlational analysis of social norms and familism may be wrong.

**Keywords:** Response Set, Social Desirability Bias, List Experiment