

온라인 설문에서 거울 이미지를 활용한 불성실 응답 방지 방법에 대한 연구

김 우 영

서강대학교

이 태 현[†]

중앙대학교

장 재 윤[‡]

서강대학교

온라인 설문은 기존의 지필형 설문에 비해 실시의 편의성과 비용적인 측면에서 이점이 있기 때문에 일반적인 연구 장면뿐만 아니라, 고객 및 종업원 인식조사 등 실무 장면에서도 광범위하게 사용된다. 온라인 설문에서, 낮은 비율의 불성실 응답으로도 데이터의 질과 분석 결과에 미치는 부정적인 영향 및 불성실 응답의 탐지 방법에 대한 연구가 활발히 이루어지는 것에 비해, 불성실 응답을 사전에 방지하기 위한 방법에 대한 실증적인 연구는 매우 부족한 편이다. 본 연구에서는 응답자들의 불성실 응답을 방지하기 위해 사회심리학의 객관적 자기인식이론(objective self-awareness)을 적용하여 거울이미지를 배경으로 삽입하는 온라인 설문 화면 구성하였으며, 이는 매우 간단하지만 지금까지 시도되지 않았던 방법이다. 거울이미지의 불성실 응답 방지효과를 파악하기 위해 경고지시문을 활용한 조건 및 일반적인 조사 지시문을 활용한 통제 조건과 비교하였다. 비교를 위해서 기존 불성실 응답 연구들에서 도출된 지표를 활용하였으며, 연구 결과, 경고 지시문 활용이나 일반적인 지시문을 활용한 통제조건에 비해 거울이미지를 삽입한 조건에서 불성실 응답이 감소되는 경향이 있는 것으로 나타났다. 끝으로 본 연구의 의의, 제한점 및 추후 연구방향에 관하여 논의하였다.

주요어 : 객관적 자기인식 이론, 거울, 불성실 응답, 온라인 설문

† 교신저자 1: 이태현, 중앙대학교, 서울시 동작구 흑석로 84 중앙대학교 305관 918호, lee0267@cau.ac.kr

‡ 교신저자 2: 장재윤, 서강대학교, 서울시 마포구 백범로 35 서강대학교 다산관 337호, jych@sogang.ac.kr

온라인 설문은 기존의 지필형(paper-pencil) 설문에 비해서 시간, 편의성, 비용적인 측면에서 이점이 있으며, 일부 학자들은 Amazon Mechanical Turk(MTurk)와 같은 온라인조사 응답자 풀(pool)이 대표성 있는 표본이라는 긍정적인 의견을 제시했다(Buhrmester, Kwang, & Gosling, 2011; Goodman, Cryder, & Cheema, 2013). 온라인 조사 서비스나 프로그램을 활용할 경우, 기존의 지필형 설문지에 비해 더 정교하고 복잡한 연구 설계가 가능하기에 조사 분야에서 일반적인 추세로 자리 매김하였으며, 연구 장면뿐만 아니라 많은 기업과 공공조직에서는 구성원 혹은 고객들의 인식조사나 조직문화 등을 조사할 때 온라인으로 설문을 진행하고 있다. 이렇게 얻어진 데이터는 연구 결과의 도출, 인사 결정, 정책 수립 등 다양한 방면에 활용되기 때문에 기본적으로 양질의 데이터 확보가 매우 중요하다고 할 수 있다. 기존 조사연구 장면에서 사회적으로 바람직한 응답(socially desirable responding)이나 거짓응답(faking)에 대처하려는 시도들을 통해 데이터의 질을 높이기 위한 노력을 해왔으나(Ones, Viswesvaran & Reiss, 1996; Mueller-Hanson, Heggstad, & Thornton, 2003), 응답자들의 불성실 응답(careless responding)이 데이터의 질에 미치는 부정적인 영향에 대한 관심은 상대적으로 낮은 편이다. 사실상 응답자들의 불성실 응답은 측정오차(measurement error)를 유발시켜서 변인들 간의 상관을 약화시킬 수 있으며(McGrath, Mitchell, Kim, & Hough, 2010), 제 2종 오류(Type 2 error)의 발생을 증가시킬 수 있다(Huang, Liu, & Bowling, 2015). 따라서 불성실 응답의 존재는 기본적으로 조사결과의 타당도에 위협을 가할 수 있다. 하지만 불성실 응답의 정의 및 불성실 응답이 연구결과에

미치는 영향력에 대해서는 통계 또는 조사 방법론 교과서에도 거의 다루지 않고 있다(Liu, Bowling, Huang, & Kent, 2013). 불성실 응답에 대한 인식과 관련된 하나의 예시로, 254명의 미국 산업및조직심리학회(SIOP: Society for Industrial and Organizational Psychology) 회원들(과학자: 37%, 실무자 23%, 과학자 & 실무자: 40%)을 대상으로 온라인 설문의 불성실 응답에 대한 인식조사를 한 Liu 등(2013)의 연구에 따르면, '불성실 응답이 연구결과에 미치는 영향력'에 대해서는 평균 2.62점, '불성실 응답을 방지하고, 불성실 응답 케이스를 선별하기 위해 얼마나 노력을 기울이는지'에 대해서는 평균 3.13점으로 나타났다(5점 척도, "전혀 그렇지 않다" ~ "매우 그렇다"). 이러한 결과에 대해 연구자들은 불성실 응답이 데이터의 질에 미치는 실질적인 영향을 SIOP 회원들이 중요하게 생각하고 있지 않으며, 오히려 과소평가하고 있는 것이라고 언급했다. 특히 이 연구에서는 '불성실 응답이 데이터의 질에 미치는 영향에 대한 인식'과 '불성실 응답을 방지하고 걸러내기 위한 노력' 간에 유의한 부적상관이 나타났는데, 이는 인식과 노력 간에 부정적인 연쇄 순환이 나타날 수 있다는 가능성을 보여주는 결과라고 할 수 있다.

수집된 데이터의 '질'이 결과도출 및 해석에 중대한 영향을 끼칠 수 있다는 점에서, 본 연구에서는 온라인 설문 시 응답자들의 불성실 응답을 방지함으로써 양질의 자료 수집을 위한 방법에 대해 탐색하고자 한다. 이를 위해 본 연구에서는 사회심리학의 객관적 자기인식 이론(OSA, objective self-awareness theory)(Duval & Wicklund, 1972)을 적용하여 설문을 디자인한 조건과 일반적인 조사 지시문을 활용한 통제 조건을 비교하여, 새로운 설문 디

자인이 불성실 응답의 방지에 미치는 효과를 검증하고자 하였다. 이 때 불성실 응답을 판별하는데 사용되는 준거들(criteria)은 기존의 불성실 응답 연구들에서 도출된 지표들을 사용하였다.

불성실 응답의 정의

불성실 응답(careless responding)은 문항의 내용과 관련 없이 무작위로(randomly) 응답하는 것, 설문지상에 빈 칸(missing)을 남기는 것, 질문을 잘못 이해하거나 질문의 내용을 전혀 고려하지 않고 문항에 응답하는 것, 또는 내용이 서로 다른 문항들에 대해 의도적으로 동일한 반응을 반복적으로 하는 것을 특징으로 한다(Curran, Kotrba, & Denison, 2010; Johnson, 2005; Meade & Craig, 2012; Niessen, Meijer, & Tendeiro, 2016). 이러한 불성실 응답의 특징은 응답자가 의도적으로 부정확한 응답을 하는 것뿐만 아니라 많은 주의나 노력을 들이지 않고 대충 응답하는 것을 모두 포함한다고 볼 수 있으며, 응답 내용이 응답자의 진정한 특성을 반영한다고 보기 어려운 측면이 있다. 불성실 응답이라는 용어가 기존의 여러 연구들에서 많이 사용되고 있으나, IER(insufficient effort responding; 응답자가 설문에 제시된 지시문을 따르거나, 문항 내용을 제대로 해석하거나, 정확한 응답을 제공하고자 하는 동기가 없거나 낮을 때 발생하는 응답의 형태)(Bowling, Huang, Bragg, Khazon, Liu, & Blackmore, 2016; Huang, Curran, Keeney, Poposki, & DeShon, 2012)이라는 용어도 이 분야에서 널리 사용되고 있으며, 용어 자체는 다르나 핵심적인 의미는 동일하다고 볼 수 있다.

사실, 이러한 불성실 응답에 대한 논의는

아주 새로운 것은 아니며, 오래전부터 연구자들은 설문 연구에서 참가자들의 불성실 응답에 대해 우려해왔다(Buechley & Ball, 1952; Greene, 1978; Haertzen & Hill, 1963). 자기보고식 설문에서 불성실 응답을 탐색하기 위한 많은 연구들은 임상척도 도구 개발과정에서 진행되었다. 예를 들어, Personality Assessment Inventory(PAI; Morey, 1991) 혹은 Minnesota Multi-phasic Personality Inventory(MMPI; Butcher, Dahistrom, Graham, Tellegen, & Kaemmer, 1989)와 같은 도구들은 보통 수백문항으로 구성되어 있기에 참가자들로 하여금 높은 수준의 집중력 또는 주의력이 유지되는 것이 필요하다. 따라서 도구를 개발할 당시 문제가 되는 응답을 걸러내기 위한 척도들이 개발되었다. MMPI와 PAI에서 불성실 응답을 탐색하기 위해 사용되었던 비일관성 척도(inconsistency scale)는 내용이 거의 동일한 문항 쌍 여러 개가 설문의 전반과 후반부에 각각 배치되었고, 두 개의 응답을 모두 고려하여 불성실 응답의 여부를 파악하였다. 예를 들어, “I am an active person”이 전반부에 있었다면 후반부에는 “I have an active lifestyle”이 배치되었다. 두 문항 간 점수 차이의 절대 값을 여러 개의 문항 쌍에 걸쳐서 계산하였고, 높을수록 불성실한 응답으로 간주하였다. 이 척도는 실제 수집된 데이터와 무작위로 만들어진 데이터를 효과적으로 구분하는 것으로 나타났다(Bruehl, Lofland, Sherman, & Carlson, 1998; Pinesoneault, 2005).

불성실 응답의 원인

불성실 응답의 원인으로는 크게 네 가지를 들 수 있다(Meade & Craig, 2012). 첫째, 응답자의 동기(respondent motivation), 둘째, 환경의 주

의분산 요인(environmental distraction), 셋째, 연구자와의 사회적 접촉(social contact), 넷째, 설문지 길이(survey length)이다.

먼저 응답자의 동기에 대해서 살펴보면, 양질의 데이터를 얻기 위해서는 설문의 응답 과정에서 응답자의 적극적 참여나 몰입은 상당히 중요하다(Schwarz, 1999; Tourangeau, Rips, & Rasinski, 2000). 종업원 선발이나 개인에게 피드백을 주고자 하는 목적의 피드백 설문(feedback survey)과 같이 응답자가 설문의 목적과 내용에 진정한 관심을 가지고 있는 경우, 불성실 응답은 거의 나타나지 않는다. 그러나 예를 들어, 학점(course credit)을 위해 연구에 참여하는 대학생들과 같이 설문에 응답하는 것을 귀찮아하며 설문 참여 자체에 불만을 가진 사람들이 많은 경우(Schultz, 1969), 불성실 응답의 가능성도 높다.

두 번째로 환경의 주의분산 요인이다. 온라인으로 실시하는 설문의 가장 큰 단점은 연구자의 통제를 벗어난 연구 장면이라는 점이다. 이러한 상황에서는 주의 깊고 성실한 응답을 하려는 응답자들마저도 주변의 다양한 자극들로 인해 산만해지고 집중을 못하는 경우가 발생한다. 인지적 과제에 대해 주의 분산이 수행을 저하시키는 효과(Spelke, Hirst, & Neisser, 1976)를 고려할 때, 온라인 설문 참가자들이 처한 환경에서의 주의분산 요인은 불성실 응답을 유발할 수 있다.

세 번째, 연구자와의 사회적 접촉과 불성실 응답의 관련성이다. Dillman, Smyth, 및 Christian(2014)의 설문연구 설계 모델에서는 연구자와 응답자 간의 사회적 상호작용을 높이기 위한 맞춤형 소통 설계(tailored communication design)를 강조한다. 같은 맥락으로 Johnson(2005)은 연구자와 응답자의 물리적

거리가 응답자로 하여금 응답에 성실하게 임해야 한다는 책임감을 떨어뜨리고, 바람직하지 않은 응답 패턴으로 이어질 가능성이 높다고 지적했다.

마지막으로, 설문 길이와 불성실 응답의 관련성이다. 설문이 길수록 응답자들에게 더 많은 노력이 요구되며 피로는 높아진다. 300~400개 정도의 많은 문항에 응답하는 응답자들의 응답패턴을 살펴본 결과, 설문의 중간이나 끝 부분에서 문항 내용에 주의를 기울이기 보다는 무작위로 반응하는 경향성이 나타났다(Baer, Ballenger, Berry, & Wetter, 1997; Berry, Wetter, Baer, Larsen, Clark, & Monroe, 1992). 심지어 직무 지원자와 같이 고부담(high-stake) 상황의 표본들에서도 이런 경향성이 나타나는 것으로 밝혀졌다(Berry et al., 1992).

불성실 응답의 영향

문항 간 상관에 기초하여 척도를 개발할 때, 불성실 응답은 오차 분산(error variance)을 증가시켜 내적일관성을 약화시키거나 혹은 해석 불가능한 문항 간 상관 패턴을 유발함으로써 결국 적절한 요인구조를 도출하지 못하는 결과에 이르게 할 수 있다(Johnson, 2005). Crede(2010)에 따르면, 무작위 응답 비율을 전체 표본의 5% 정도로 가정하여 시뮬레이션 분석을 하였을 때, 상관계수가 과대 혹은 과소 추정 되었는데 그 정도는 범위제한이 있을 때, 연속변수를 임의로 이분변수로 만들어서 분석할 때, 또는 신뢰도가 낮은 경우에 나타나는 효과와 비견될 수 있다고 하였다. 또한 역산 문항에 대한 불성실 응답은 방법요인(method factor)의 생성에 기여하여, 일반(긍정문, 평서문) 문항과 역산(부정문) 문항이 각자

다른 요인에 적재되는 결과를 초래한다(Woods, 2006). 이 외에도 불성실 응답은 허위 집단 내 변산성(spurious within-group variability)을 유발하며, 신뢰도를 약화시키고 이는 다시 상관관계에 영향을 미침으로써, 제2종 오류를 발생시킬 가능성이 있다(Clark, Gironda, & Young, 2003; Liu et al., 2013).

Maniaci와 Rogge(2014)의 시뮬레이션 연구에 따르면, 표본에서 불성실 응답 비율이 5%일 때 불성실 응답이 전혀 없을 때 보다 검정력(power)이 약 10% 포인트 낮아졌으며, 불성실 응답이 10%인 경우에는 검정력이 18% 포인트까지 낮아지는 것으로 나타났다. 반대로, 개념적으로 구분되는 구성개념들 간의 관계를 과대 추정하여 제1종 오류를 유발시킬 수도 있다(Huang et al., 2015).

불성실 응답은 연구의 반복검증에도 부정적인 영향을 미치는 것으로 나타났다. Oppenheimer, Meyvis, 및 Davidenko(2009)는 예전에 보고되었던 실험 결과들이 성실한 응답자로 분류된 참가자들에게서는 반복검증 되었으나, 불성실 응답자로 분류된 참가자들에게는 반복 검증되지 않는 것을 발견했다. 또한 McGrath, Mitchell, Kim, 및 Hough(2010)의 연구에서는 응답자들의 비일관적인 응답(inconsistent responding)이 준거관련 타당도(criterion-related validity)에 부정적인 영향을 미치는 것을 밝혔다.

이러한 연구결과들을 바탕으로 보았을 때, 불성실 응답은 데이터의 분석 과정에서 오류를 발생시킴으로써 분석결과에 대한 확신을 떨어뜨리게 된다. 특히, 불성실 응답으로 유발되는 문제점들은 설문으로부터 얻어진 점수를 바탕으로 정책개발, 종업원 선발, 마케팅 계획 등과 같은 고부담(high-stake)의 의사결정을 내

려야 할 때 심각히 우려되는 부분이다. 따라서 불성실 응답을 사전에 방지하거나 사후에 탐지함으로써 수집된 자료의 질을 높이는 것은 자료를 분석한 결과에 대한 타당도 및 양질의 의사결정을 내리기 위해 매우 중요한 작업이다.

불성실 응답의 방지 방법 및 사후 탐지 방법

불성실 응답의 방지 방법

불성실 응답 방지를 위해 기존에 널리 사용되던 방법으로는 설문지의 지시문에 불성실 응답에 대한 경고 문구를 넣는 것이다(Meade & Craig, 2012; Huang et al., 2012; Huang et al., 2015; Ward & Pond, 2015). 대학생들을 대상으로 한 Huang 등(2012)의 연구에서 설문 응답에 많은 노력을 기울이지 않으면 학점(course credit)에 부정적 영향을 미칠 수 있다는 경고 문구를 제시함으로써 응답자들의 불성실한 응답 방지를 유도했다. 연구 결과, 일반적인 지시문을 사용한 통제조건에 비해 경고 문구를 사용한 실험조건에서 응답자들의 불성실 응답이 유의하게 감소되었다. 하지만 이러한 경고 문구는 응답자의 책임감을 높일 수는 있을지라도 응답자들에게 강압적이고 부정적인 느낌을 줄 수 있으며, 잠재적으로 응답자들의 반발심을 살 수도 있기에 최선의 방법은 아닐 수 있다(Meade & Craig, 2012).

이와 다르게 사회적 영향력(social influence) 관련 사회심리학적 이론이나 개념들을 지시문에 적용하여 성실한 응답의 동기를 높이려는 시도가 Ward와 Meade(2018)에 의해 이루어졌다. 이들의 연구 1에서, 연구자들은 온라인 설문에서 불성실 응답을 유발하는 원인 중 하나인 연구자와의 사회적 접촉 부재를 극복하기

위한 방안으로 지시문을 연구자가 직접 읽는 비디오 클립을 설문 초반에 제시하였다. 이러한 조작은 응답자가 지각하는 연구자와의 물리적 거리를 좁히고, 연구자의 권위를 더욱 부각시키는 역할을 하여, 성실한 응답을 유도하려는 방법이었다. 연구 결과, 글로 제시되어서 참가자가 직접 읽는 일반적인 지시문 조건의 참가자보다 연구자가 직접 지시문을 읽어주는 영상이 나오는 비디오 클립 조건의 참가자들이 성실한 응답을 나타내는 주관적 지표(응답의 성실성에 대한 자기보고 측정치)에서 유의하게 더 높은 점수를 보였다. 이 방법과 유사하게, Ward와 Pond(2015)의 연구에서는 그림 1에 제시된 것처럼 온라인 설문 페이지 왼쪽 상단 빈 공간에 가상 인간을 배치하여 연구자와 응답자 간 지각된 상호작용을 높이려는 시도를 하였다. 이러한 가상 인간 조건에 할당된 참가자들은 설문 응답 내내 고개를 끄덕이거나 눈을 깜빡이는 등의 행동을 취하는 가상인간의 모습을 볼 수 있었다. 하지만 가상인간의 모습 자체로는 불성실한 응답을 방지하는 효과가 없는 것으로 나타났다.

Ward와 Meade(2018)의 연구 2에서는 위선 패러다임(hypocrisy paradigm)을 적용하여 응답자의 성실한 응답을 유도하고자 했다. 위선 패러다임의 논리는 사람들은 자신이 가지고 있는 태도에 반하는 위선의 예시가 마음속에 떠오르면 기존에 지니고 있는 태도와 일관되지 않은 행동을 조정한다는 것이다(Aronson, Fried, & Stone, 1991; Stone, Aronson, Crain, Winslow, & Fried, 1994). 위선 조건에서는 설문 초반에 응답자들로 하여금 설문에 성실하게 응답해야 하는 이유에 대해서 웹페이지상의 빈 칸에 쓰도록 요구했다. 이유를 작성한 응답자들은 성실한 응답이 중요하다는 것에 대한 태도를 지

니게 될 것이며, 추후 진행되는 응답에서 불성실하게 응답하는 것은 위선 행동에 해당할 것이다. 따라서 위선 행동으로 유발되는 불편한 마음을 경험하지 않기 위해 기존에 자신이 지닌 태도와 일관되게 보다 더 성실하게 응답할 것이라고 예상하였다. 연구 결과, 위선 조건에서 불성실 응답이 유의하게 감소되었다.

또 다른 방법으로는 설문이 끝난 후에 참가자들에게 응답의 질에 대한 피드백을 주겠다는 지시문을 활용하는 것이다(Ward & Pond, 2015). 이 방법은 과업에 대한 피드백을 주게 되면, 응답자들이 과업에 더 많은 주의를 기울이게 되고(Kluger & DeNisi, 1996), 과업 수행이 향상된다는 연구(Northcraft, Schmidt, & Ashford, 2011)에 근거한 것이다. Ward와 Pond(2015)의 연구에서는 설문 조사가 완료된 후 응답자가 제공한 정보를 사용할 수 있는지의 여부 및 응답의 질에 대한 피드백을 제공한다는 문구를 활용했다. 연구 결과, 일부 지표에서는 피드백 조건의 참가자들이 더 성실하게 응답하는 것으로 나타났으나, 다른 지표에서는 반대의 결과가 나타나서 피드백 지시문이 불성실 응답에 미치는 효과는 일관되지 않은 것으로 나타났다.

지금까지 소개한 네 가지 방법들은 성실한 응답에 대한 동기유발 및 행동을 유도하기 위해 지시문을 조작하거나, 설문 화면에 특별한 영상이나 애니메이션을 배치하는 방식을 취하고 있다. 특히, Ward와 Meade(2018), Ward와 Pond(2015)의 연구에서는 기존에 널리 활용되던 '경고 지시문'을 사용한 방법보다는 응답자들의 기분을 해치지 않으면서도 성실하게 응답하려는 동기를 유발했다는 점에서 의의가 있다.

불성실 응답의 사후 탐지 방법

불성실 응답 탐지 문항 삽입

첫 번째 방법은 설문을 제작하는 단계에서 불성실 응답자를 탐지할 수 있는 문항들을 삽입하여, 그 문항에 대한 응답 패턴을 확인하여 불성실 응답의 정도를 파악하는 방법이다. 이러한 문항들은 다시 세 종류로 구분된다. 먼저 저빈도 접근(*infrequency approach*)에 해당하는 문항으로 지시 반응 문항(*instructed response items*)과 위조 문항(*bogus items*)이 있다(Meade & Craig, 2012; Huang et al., 2012). 다음으로 의미적 유의어(*semantic synonyms*) (DeSimone, Harms, & DeSimone, 2015)라고도 불리는 비일관성 문항(*inconsistency items*)이 있다. 마지막으로 자기보고식으로 측정하는 사용 여부 질의 문항(*UseMe item*; 응답자에게 본인의 응답 내용을 연구에 반영해도 되는지를 직접적으로 물어보는 문항) 및 자기보고 성실/흥미 문항(*self-report diligence/interest items*, 설문 응답에 대한 응답자의 성실성과 흥미를 측정하는 자기보고식 문항)(Meade & Craig, 2012; Huang et al., 2012; Niessen, Meijer, & Tendeiro, 2016)이 있다.

지시 반응 문항 및 위조 문항의 기본 가정은 ‘성실한 응답자라면 이 문항들에 대해 대부분 동일하게 응답할 것이다’이다. 즉, 답이 명백하게 정해져 있는 질문을 제시하고, 틀렸을 경우 불성실 응답으로 간주하는 것이다. 지시 반응 문항의 예시로는 “이번 문항에서는 문항의 내용과 상관없이 4번에 체크해주시기 바랍니다.” 라는 것을 들 수 있는데, 응답자가 문항을 성실히 읽고 응답한다면 4번에 체크할 것이고, 만약 불성실한 응답자라면 4번 이외에 다른 선택지에 응답할 것으로 본다. 그

리고 위조 문항은 일반적인 상식에서 벗어나는 반 사실적인 진술문으로 구성되며, 예시로는 “나는 한 번도 컴퓨터를 사용해본 적이 없다.”라는 것을 들 수 있다. 만약 성실한 참가자라면, 이 문항에 대해서는 당연히 “동의하지 않는다.” 혹은 “그렇지 않다.” 응답할 것이다.

비일관성 문항은 의미적으로 유사한 문항 쌍으로 구성되어 있으며(예: “나는 매우 열정적인 사람이다.”, “나는 열정이 넘친다.”), 하나의 문항은 설문의 전반부에 다른 하나는 설문의 후반부에 배치한다(Maniaci & Rogge, 2014). 따라서 두 개 문항에 유사하게 응답할수록 성실한 응답으로 간주할 수 있다.

사용 여부 질의 문항(*UseMe item*)의 경우, 설문의 맨 마지막에 “마지막으로, 귀하께서 생각하시기에 저희가 귀하의 응답 내용을 믿고 연구를 진행해도 될까요?” 라고 직접 물어봄으로써, 참가자 스스로에게 자신의 응답이 성실했는지 아닌지에 대한 판단을 맡기는 것이라고 볼 수 있다(Meade & Craig, 2012). 자기보고 성실/흥미 문항의 경우, 설문의 말미에 응답자에게 직접적으로 얼마나 이 설문에서 성실하게 응답하였는지, 설문에서 응답하는 과정을 얼마나 즐겼는지에 대해 직접 물어보는 방식이다(Huang et al., 2015; Meade & Criag, 2012).

이처럼, 제시된 문항들에 대한 응답자들의 점수 패턴을 살펴본 이후에 응답자들의 불성실 정도를 가려낼 수 있고, 불성실하다고 판단되는 경우 해당 응답자의 응답 내용을 삭제하고 분석을 진행할 수 있다.

통계적인 방법

사후적인 방법의 두 번째로는 통계적인 분석을 통해 도출된 수치를 확인함으로써 불성실 응답을 탐지하는 방법이다. 이 방법은 설

문에 특정한 문항을 삽입하지 않아도 되며, 단지 설문 완료 후 수집된 데이터를 바탕으로 분석하는 방법이다. 이것도 크게는 (비)일관성 접근법((in)consistency approach), 응답패턴 접근법(response pattern approach), 응답시간 접근법(response time approach)으로 구분해볼 수 있다(Meade & Craig, 2012; Huang et al., 2012).

먼저 (비)일관성 접근법에 해당하는 방법으로, 심리측정적 동의어/반의어 지표(Psychometric synonyms/antonyms index), 홀-짝 일관성 지표(even-odd consistency index)가 있다. 이러한 지표들은 이름에서도 알 수 있듯이 문항들에 대해서 응답자가 얼마나 (비)일관되게 응답하는지를 파악하는 방법이다. 이들 지표의 기본 논리는, 데이터 전반에 걸쳐서 나타나는 문항 간 혹은 변인 간 상관관계의 패턴이 개인 내에서도 일관되게 나타나는지를 살펴보는 것이다. 따라서 일관되게 나타날수록 성실한 응답으로 간주된다.

다음으로 응답패턴 접근법에 해당하는 것으로, 마할라노비스 거리(Mahalanobis distance)와 최대 롱스트링(maximum longstring)이 있다(Meade & Craig, 2012; Huang et al., 2012). 마할라노비스 거리는 특정 응답자의 반응 패턴이 전체 자료에서 일반적으로 나타나는 응답 패턴(normative response pattern)에서 떨어져 있는 정도를 측정하며, 다변량 극단치(multivariate outlier)를 찾는 방법으로 자주 사용된다. Ehlers 등(2009)은 마할라노비스 거리가 클수록 불성실 응답을 나타내는 것과 관련이 있으며, 이 지표가 응답에 주의를 많이 기울이지 않는 응답자를 탐지하는 데 효과적이라는 것을 보였다. 최대 롱스트링은 문항의 내용과 상관없이 동일한 선택지에 반복적으로 응답하는 것을 의미하며, 전체 설문 응답과정에서 응답자가

동일한 선택지에 응답한 최대의 개수가 최대 롱스트링 지수라고 할 수 있다. 예를 들어, A라는 사람이 5점 척도로 구성된 10개의 문항에 대해서 1111444333으로 응답했을 때, A라는 사람의 최대 롱스트링 지수는 '4'가 된다. 그리고 이 값이 클수록 불성실 응답으로 간주할 수 있는데, 절대적인 cut-off 점수를 설정하기 어렵다는 단점이 있다. Costa와 McCrae(2008)는 NEO-PI-R 성격검사에 기반을 두어 최대 롱스트링의 개수에 대한 가이드라인을 제시한 바, 문항의 종류와 수에 따라서 6 ~ 14개를 기준으로 제안했다.

응답시간 접근법은 응답에 걸린 총 시간을 고려하여 불성실 응답자를 탐지하는 것이다. 앞서 설명한 불성실 응답의 특징에 근거하면, 불성실 응답자들은 성실한 응답자들에 비해서 개별 문항에 주의를 기울이는 시간이 짧을 것이고, 결국 설문에 걸리는 시간이 짧을 것이다. Huang 등(2012)에 따르면, 전체 설문 시간을 계산했을 때, 한 문항 당 평균 2초 이하로 응답한 응답자들을 불성실 응답으로 간주할 수 있다고 하였다.

이렇게 불성실 응답을 사후적으로 탐지하는 방법들은 상당히 많이 제안되어 왔으나, 선제적으로 방지하는 방법들은 여전히 부족한 편이다. 사후적인 방법들은 통계적 수치를 바탕으로 불성실 응답을 찾아내고, 이를 분석에서 제외하는 것이기 때문에 자료수집의 비용적인 측면에서 손실이 유발된다. 또한 줄어든 표본 크기는 검정력을 약화시키는 결과로 이어지고, 표본 분포를 왜곡시켜서 표본의 대표성을 약화시킬 수 있으며, 이는 연구결과의 외적타당도를 저하시킬 수 있다(Meade & Craig, 2012) 따라서 본 연구에서는 불성실 응답에 대한 사전 방지법에 초점을 두었으며, 새로운 방법

한 가지를 제안하고자 한다.

불성실 응답 방지 방법 제안

인간의 공격성에 대한 사회심리학의 고전적 연구인 Berkowitz와 LePage(1967)의 실험에서는 연구 참가자가 상대방(실험협조자)에게 공격성을 표출할 수 있는 기회가 주어진다. 그리고 이때 공격성을 표출하는 장치로 전기 충격기를 사용하는데, 한 조건에서는 전기충격기를 놓인 바로 옆 선반에 배드민턴 라켓과 같이 일반적인 사물을 배치하였고, 다른 조건에서는 라이프/리볼버 등의 총기류를 배치하였다. 실험 결과, 주변에 리볼버와 같은 무기류가 있는 조건에서 실험 참가자의 공격성(전기충격 주기)이 더 높게 나타났다. 실험 직후 이어진 인터뷰에서 참가자들은 자신들의 옆에 놓인 사물(무기류)이 공격성에 미치는 영향력을 인지하지 못했고, 심지어 그것이 옆에 놓여있었다는 사실도 지각하지 못하는 것으로 나타났다. 이 연구를 통해 공격성의 표출과 연관이 되어 있는 단서(cue)가 주위에 있을 경우 사람들은 무의식적으로 그 단서의 영향을 받아 공격 행동이 나타날 수 있다는 단서촉발이론이 제안되었다.

이러한 연구결과를 토대로 온라인 설문 응답자들의 주의 깊고 성실한 응답을 이끌어내기 위해 응답자 주위에 어떤 사물을 배치하여 행동을 촉발할 수 있을 지에 대해서는 Beaman, Klentz, Diener, 및 Svanum(1979)의 연구에서 시사점을 얻을 수 있었다. 이 연구는 미국의 할로윈데이 시즌에 아이들을 대상으로 한 실험이었다. 연구자는 집 앞에 대기하고 있다가 사탕을 가지러 온 아이들에게 사탕을 하나씩만 가져가라고 요청을 한 다음 아이들

이 연구자의 말을 잘 지키는지 관찰했다. 두 조건이 있었는데, 한 조건은 사탕 바구니 앞에 거울이 배치되었고, 다른 조건에서는 거울이 없었다. 연구 결과, 거울이 있는 조건에서 연구자의 말을 더 잘 따르는 것으로 나타났는데, 그 이유는 사탕을 짚어갈 때 아이들은 거울을 통해 자신의 얼굴을 보게 되고, 이로 인해 객관적 자기인식(objective self-awareness)이 유발되었기 때문이라는 것이다.

인간은 자신의 주의를 상황으로부터 자기 자신 내부로 혹은 자기 내부에서부터 상황으로 전환할 수 있는 능력을 가지는데(Duval & Wicklund, 1972; Mead, 1934), 객관적 자기인식이란 개인의 주위가 자기 내부로 향하는 상태를 의미한다. 객관적 자기 인식을 유발하는 방법들로는 녹음된 자신의 목소리를 듣는 것(Ickes, Wicklund, & Ferris, 1973), 자서전을 작성하게 하는 것(Duval, Duval, Neely, 1979), 거울을 설치하여 자신을 보게 하는 것(Duval & Wicklund, 1972) 등이 있다. 특히, 거울에 비친 자신의 모습을 보는 것은 자기반성(self reflection)을 증가시키므로, 객관적 자기인식을 조작하는 데 많이 사용되어 왔다(Beaman 등, 1979; Diener & Wallbom, 1976; Lawson, Stedmon, Zhang, Eubanks, & Frumkin, 2013).

객관적 자기인식 이론에 따르면, 객관적 자기 인식 상태에서 사람들은 환경에 적합한 행동 기준과 자신의 행동 간의 일관성에 초점을 맞출 가능성이 높다(Duval & Wicklund, 1972). 만약 자기의 행동과 환경이 요구하는 기준 간에 부조화(dissonance)가 발생하면 개인은 불편한 감정을 느끼고, 부조화를 줄이려고 노력한다. 자기인식이 개인의 행동에 미치는 영향력에 대한 설명은 Carver와 Scheier(1981)의 사이버네틱스 모형(cybernetic model)의 자기조절

(self-regulation) 개념으로도 가능하다. 이 모델에 따르면, 자기조절은 자기의 현재 상태와 사회적으로 주어진 기준 혹은 내적으로 지니고 있는 기준의 비교를 통해 자신의 행동을 조절하는 지속적인 순환과정이라고 보았으며, 이러한 과정을 Test, Operate, Test, and Exit(TOTE)라고 언급했다. 즉, 자신과 기준을 비교하여 불일치를 확인하고(Test), 불일치를 줄이기 위해 행동한 후에(Operate) 다시 자신과 기준을 비교하며(Test), 기준에 도달하면 순환 과정을 벗어나는 것이다(Exit). 자기인식은 사이버네틱스 모형의 비교과정과 맞닿아 있으며, 사람들은 현재 자신의 상태와 기준을 비교하여 불일치가 발생할 경우, 자신의 행동을 기준에 부합하게 조절하도록 동기화된다. 기준을 충족하기 위해 노력하는 것은 객관적 자기인식 이론의 가장 기본적인 효과이며, 이러한 효과는 수많은 실험에서 검증되었다(Silvia, & Duval, 2001). 또한 subliminal first-name priming (실험 참가자의 이름을 의식적으로 지각하지 못할 정도로 모니터 상에 빠르게 제시하여 자기인식을 점화하는 방법)을 통해서 자기 내부로 주의를 집중시키는 조작 방법을 사용한 Silvia와 Phillips(2013)의 실험 연구에 따르면, 참가자들이 의식적으로 지각하지 않더라도 자기와 기준을 비교하는 과정이 자동적으로 일어났다.

시험 상황의 부정행위(cheating)와 자기인식 간의 관계를 다룬 Diener와 Wallbom(1976)의 연구에서는 참가자들을 거울 앞에 앉아서 녹음된 자신의 목소리를 듣는 자기인식 조건과 거울 옆에 앉아서 다른 사람의 목소리를 듣는 통제 조건으로 나누어서 참가자들의 행동이 조건 간 어떻게 다른지를 살펴보았다. 이 연구에서 인지능력 검사를 위해 참가자들에게

주어진 시간은 5분이었으며, 연구자가 밖으로 나가면 검사가 시작되는데, 연구자는 10분이 지나서도 돌아오지 못할 수 있다고 얘기하여 참가자들이 부정행위를 할 수 있을 법한 상황을 연출하였다. 다만, 5분 뒤 알람시계가 울리면 문제 풀이를 중단해야 하는 기준을 제시하였고, 시간이 다 된 후에도 응답하면 부정행위로 간주하였다. 실험 결과, 자기인식 조건에서는 14명 중에서 1명이 부정행위를 한 반면, 통제 조건에서는 14명 중 10명이 부정행위를 하였다. 이 실험에서 통제 조건과 실험조건 모두에서 실험자로부터 사회적으로 현저한 행동 기준이 주어졌으나, 거울이 있는 조건의 참가자들이 기준에 더 부합하는 행동을 한 것이다. 따라서 객관적 자기인식 상태의 개인들은 자기조절을 더 하게 되며(Kanfer, 1970), 상황이 요구하는 기준이 현저(salient)할 때 이러한 기준에 자신의 행동을 더 맞추고자 한다. 또한 앞서 언급한 Beaman 등(1979)의 연구에서 사회적으로 주어진 현저한 기준은 연구자가 아이들에게 사탕을 하나만 가져가라고 직접 언급한 것이었으며, 사탕 바구니 앞에 놓인 거울을 봄으로써 객관적 자기인식이 유발된 아이들은 연구자가 언급한 기준에 자신의 행동을 더 맞추고자 했다고 해석된다.

앞서 제시한 두 연구(Beaman et al., 1979; Berkowitz & LePage, 1967) 및 객관적 자기인식 이론을 토대로 온라인 설문에서 거울을 활용하는 방법을 적용하고자 하였다. 하지만 온라인 조사라는 점을 고려해보았을 때, 응답자들이 하여금 실제로 거울 앞에 앉게 하는 것이 불가능하며, 거울 앞에 앉아서 설문을 하라는 지시문을 넣는다고 할지라도 실제로 응답자들이 거울 앞에 앉아서 설문했다고 확인할 수가 없다. 따라서 실제 거울이 아닌 거울이미지를

활용하기로 하였는데, 온라인 설문 상에 거울의 이미지가 워터마크나 기본 배경화면처럼 깔려 있도록 하여 설문을 하는 내내 응답자들에게 거울이 노출되는 방법을 사용하였다. 물론 참가자들이 실제로 거울을 통해 자신의 모습을 보는 것은 아닐지라도 거울의 이미지가 점화(priming) 되면 거울을 보는 효과가 어느 정도 나타날 것이며, 이를 통해 객관적 자기인식 상태가 유발되면 자신의 행동을 주어진 기준에 더 맞추게 될 것이라고 예상했다. 온라인 설문에서는 지시문을 통해 성실한 응답에 대한 요구를 중요한 기준으로 제시할 수 있으며, 거울이미지가 삽입된 조건과 그렇지 않은 조건 간에 불성실 응답을 측정하는 지표들 간 유의한 차이가 나타나지지를 확인하고자 했다. 추가적으로, 기존 연구에서 경고 문구가 불성실 응답을 일부 방지하는 효과가 나타났기에(Breitsohl & Steidelmuller, 2017; Huang 등, 2012; Huang 등, 2015) 국내 표본들을 대상으로도 경고 문구의 불성실 응답 방지 효과가 나타나지지를 재차 확인할 필요가 있으며, 거울이미지가 삽입된 조건과의 상대적 비교도 필요하다고 판단하였다. 따라서 본 연구에서는 일반적인 지시문을 사용하는 통제조건, 경고 문구를 사용하는 경고조건, 일반적인 지시문에 거울이미지를 삽입한 거울조건의 세 조

건 간 불성실 응답 지표에서 차이가 나는지를 고찰하고자 하였다.

방 법

연구 설계

본 연구는 조사 회사를 통해 온라인으로 실시되었으며, 응답자들은 지시문을 읽고 주어진 문항에 응답하는 과정으로 설문이 진행되었다. 설문과정에서 응답하지 않은 문항이 있는 경우, 다음 화면으로 넘어갈 수 없도록 설정하였기 때문에 기본적으로 결측 값은 존재하지 않았다. 통제, 경고, 거울 조건별로 약 100여 명의 참가자가 무선 할당(random assignment)되었으며, 각 조건별로 지시문 및 설문의 디자인이 상이했다. 먼저 통제조건의 경우 일반적으로 사용되는 지시문을 사용하였으며, 그림 1의 왼쪽에 제시되어 있다. 두 번째 조건은 경고조건에 해당하며, Huang 등 (2012)에서 사용된 경고 문구를 참조하여 지시문에 경고 문구를 넣었다. 다만 기존 연구에서는 ‘설문 응답에 많은 노력을 기울이지 않으면 학점에 영향을 미칠 수도 있다.’는 내용이었으나, 조사 회사를 통해 실시되는 온라인

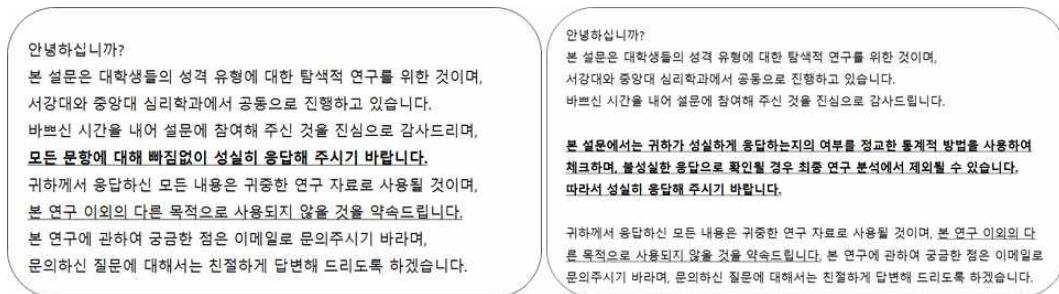


그림 1. 통제조건(왼쪽) 및 경고조건(오른쪽)의 지시문

Q1. 다음 항목들에 대해 자신과 가장 일치하는 번호에 체크해 주십시오.

		전혀 그렇지 않다	별로 그렇지 않다	보통 이다	조금 그렇다	매우 그렇다
나는 각종 모임에 다니는 것을 즐긴다.	→	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
나는 여러 사람들 가운데 있어도 편하다.	→	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
나는 말이 적은 편이다.	→	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
나는 나서지 않는 편이다.	→	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
나는 다른 사람들에 대해 관심이 있다.	→	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
나는 다른 사람들의 감정을 잘 공감한다.	→	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
나는 다른 사람들에 대해 별로 관심이 없다.	→	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
나는 다른 사람들에게 무례한 언행을 사용할 때가 자주 있다.	→	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
나는 무슨 일이든 항상 준비를 하는 편이다.	→	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
나는 세무사항도 잘 챙기려고 이에 신경을 쓰는 편이다.	→	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

그림 2. 거울조건의 설문화면 예시

설문임을 감안하여 점수나 보상을 받을 수 없다는 문구 없이 불성실한 응답은 사후 확인이 되며, 그럴 경우 제외된다는 내용이 포함되었다. 경고조건의 지시문은 그림 1의 오른쪽에 제시되어 있다. 마지막으로 거울조건의 지시문은 통제조건의 지시문과 동일하지만, 설문화면에 거울 이미지가 워터마크처럼 깔려 있어서 참가자들이 응답을 하는 내내 거울 이미지에 노출되도록 하였다. 구글 이미지 검색을 통해 3개의 거울이미지를 선택한 후 사회 및 조직심리 대학원생 20명을 대상으로 거울의 원형에 가장 적합한 것을 판단하는 설문 조사를 통해 가장 적합한 이미지를 최종적으로 선택했다. 샘플 화면은 그림 2와 같다.

연구 참가자 및 자료수집 절차

본 연구를 위한 조사는 국내 온라인 조사전문회사의 패널에서 선정된 308명을 대상으로 진행되었다. 기존 연구들에서 대부분 대학생들을 대상으로 하였기에 본 연구에서도 일관되게 2-4년제 전국 대학생들을 대상으로 설문을 실시하였으며 통제, 거울, 경고조건에 각각 100명, 101명, 107명이었다. 각 조건별 남녀의 비율은 통제, 거울, 경고조건에 각각 46 : 54, 50.5 : 49.5, 50.5 : 49.5 이었으며, 평균 나이는 각 22.22(sd=2.38), 22.29(sd=2.26), 21.79 (sd=1.75)세였다.

측정도구

성격 5요인

성격 5요인을 측정하기 위해 Goldberg(1999)의 IPIP(the International Personality Item Pool) 100 문항을 한국어로 번역한 척도(Guay, Oh, Choi, Mitchell, Mount, & Shin, 2013)를 사용하였으며, 각 요인별로 20문항이다.

우울

우울을 측정하기 위해 배성우와 신원식(2005)에서 사용된 우울척도(CES-D) 20문항을 사용하였다.

감정불능척도

개인이 경험한 감정이나 느낌을 적절하게 표현하지 못하는 개념을 나타내는 감정불능을 측정하기 위해 이양현, 임효덕, 및 이종영(1996)이 번안한 감정불능척도(TAS, Toronto alexithymia scale) 20문항을 사용하였다.

불안

불안을 측정하기 위해 김병수와 김정현(2013)에서 사용된 STAI(state-trait anxiety inventory)-2형 20문항을 사용하였다.

사회적 바람직성(social desirability)

Paulhus(1998)의 2요인 모형을 기초로 제작된 PDS(Paulhus Deception Scale)와 BIDR-7(The Balanced Inventory of Desirable Responding)을 한국대학생을 대상으로 타당화 과정을 거친 최보라(2007)의 사회적 바람직성 척도 20문항을 사용하였다.

위조문항(bogus item)

불성실 응답을 측정하기 위한 위조문항은 Breitsohl과 Steidelmüller(2017) 및 Meade와 Craig(2012)에서 사용된 문항들 중 4문항을 사용하였으며, 사용된 문항은 ‘내 친구들은 모두 외계인이다.’, ‘나는 세계의 모든 나라에 가보았다.’, ‘나는 양치질을 한 번도 해 본 적이 없다.’, ‘나는 컴퓨터를 한 번도 사용해 본 적이 없다.’이다. Meade와 Craig(2012)에서 설문 문항 50~100개마다 위조문항 1개 ~ 3개 정도 사용하는 것을 추천하였기에 본 연구에서는 위조문항의 개수를 4개로 설정하였다. 4개 문항을 합산하여 분석에 활용하였으며, 값이 클수록 불성실 응답을 의미한다.

자기보고식 성실/흥미

불성실 응답을 측정하기 위한 자기보고식 성실/흥미 문항은 Meade와 Craig(2012)에서 사용된 문항을 사용하였으며, 성실 문항은 9문항, 흥미 문항은 6문항이었다. 성실 문항의 예시는 ‘나는 모든 문항을 주의 깊게 읽었다.’, ‘나는 이 설문에 최선을 다해 응답했다.’ 등이며, 흥미 문항의 예시는 ‘이 설문에서 내가 응답을 제대로 했는지에 관심이 있다.’, ‘나는 설문 도중 지루함을 느꼈다.’(R) 등이 있으며, 값이 클수록 성실한 응답을 나타낸다.

사용여부 질의 문항(UseMe item)

불성실 응답을 측정하기 위한 사용여부 질의 문항은 Meade와 Craig(2012)에서 사용된 한 문항을 사용하였으며, 다음 문항에 대해 ‘예/아니요’로 응답하도록 하였다. ‘마지막으로, 귀하께서 생각하시기에 저희가 귀하의 응답 내용을 믿고 연구를 진행해도 될까요?’ 이 문항에 대해 예/아니오를 각각 0, 1로 코딩하였다.

기타

본 설문에 있는 내용과 유사한 문항에 응답한 경험이 있을 경우 더 빠르게 혹은 대충대충 응답할 가능성이 있다고 판단되었기에 다음 문항에 대해 '예/아니요'로 응답하도록 하였다. '본 설문에 나온 문항과 비슷한 문항에 응답해본 경험이 있다.' 또한 다른 조건에 비해 거울 조건에서는 이미지를 삽입하였기 때문에 삽입된 이미지가 문항을 읽고 응답하는 과정에서 방해 효과를 미쳤는지 확인하기 위해 다음 문항을 추가하였다. '문항을 읽고 의미를 파악하는 것에 어려움을 느꼈다.'

전체 설문은 인구 통계학적 특성을 측정하는 문항을 포함하여 총 206문항으로 구성되었다. 인구통계학 문항, 사용여부 질의 문항과 기타의 1문항(예/아니요 응답 문항), 우울과 불안(4점 척도) 및 사회적 바람직성(6점 척도)을 제외한 문항들은 Likert 5점 척도(전혀 그렇지 않다 ~ 매우 그렇다)로 측정되었다. 문항이 너무 적을 경우 불성실 응답의 분산이 나타나지 않을 가능성이 있으며, 기존 연구들에서도 보통 200-300문항(많은 것은 400문항 정도)을 사용하였기에 이를 참조하여 206문항으로 구성하였다.

불성실 응답 지표

심리측정적 동의어/반의어

심리측정적 동의어는 연구에서 측정한 모든 문항 간 상관을 낸 후, 가장 큰 정적 상관을 지닌 30문항 쌍을 선별하거나 혹은 상관계수가 .60 이상인 문항 쌍을 선별한다(Meade & Craig, 2012). 그 다음 선별된 문항 쌍에 대해 각 응답자별로 개인 내 상관(within-person

correlation)¹⁾을 구하면 개인별로 선별된 문항 쌍에 대한 상관계수가 하나씩 도출된다. 전체 샘플을 대상으로 가장 큰 정적 상관을 지닌 30쌍을 골라냈기 때문에, 성실한 응답자일수록 개인별로 구한 상관계수의 크기가 크며, 이 문항 쌍에 대해 일관되게 응답했다고 볼 수 있다. 반면, 불성실한 응답자일수록 상관계수의 크기가 작거나 심지어 마이너스 값을 지닐 것이다. 심리측정적 반의어는 가장 큰 부적상관을 지닌 문항 쌍 30개 혹은 상관계수의 절대 값이 .60 이상인 문항 쌍을 추출한 후, 선별된 문항 쌍에 대해 각 응답자별로 개인 내 상관(within-person correlation)을 구하게 되고, 계수의 절대 값이 작을수록 불성실한 응답이라고 볼 수 있다. 본 연구에서는 모든 문항들 간의 상관계수를 도출한 후, 상관계수의 절대 값이 가장 큰 순서부터 30개의 문항 쌍을 선별하여, 심리측정적 동의어 및 반의어 계산에 사용하였다.

홀-짝 일관성

어떤 변인을 측정하는 문항에서 짝수 문항과 홀수 문항을 나누어 둘 간의 상관을 보는 것이 반분신뢰도(split-half reliability)라면, 홀-짝 일관성은 반분신뢰도가 확장된 것이라고 볼 수 있다. 측정된 여러 변인들에 대해서 각 변인들의 짝수문항 평균과 홀수문항 평균을 구하여 쌍을 만들고, 이 쌍들에 대해 개인 내 상관계수를 구한다. 그러면 개인 별로 일관성 지표가 도출되고, 값이 작을수록 불성실 응답을 나타낸다. 본 연구에서는 비교적 단일요인이 분명하게 드러나는 성격 5요인에 대해서 각 요인별 홀/짝 문항을 사용하여, 개인별로

1) 개인 내 상관계수는 microsoft excel에서 CORREL 함수를 활용하여 계산하였음.

흡-착 일관성 지표를 산출하였다.

불성실 응답지표들의 기술통계 및 상관관계

마할라노비스 거리(Mahalanobis distance)

마할라노비스 거리는 성격 5요인, 불안, 우울, 사회적 바람직성을 측정하는 162개 문항을 동시에 고려하여 개인별로 산출하였으며, 마할라노비스 거리는 SPSS 회귀분석 모듈에서 제공하는 값을 사용하였다. 값이 클수록 전반적인 응답 분포에서 멀리 떨어져 있다고 볼 수 있기 때문에 불성실 응답으로 간주하였다.

최대 롱스트링

동일한 선택지에 연속해서 응답하는 값이 가장 큰 경우를 각 개인별로 산출하였다. 본 연구에서는 최대 롱스트링을 산출하기 위해서 엑셀 매크로를 활용하였으며, 인구통계학 문항을 제외한 모든 문항을 대상으로 하여 동일한 선택지에 반복적으로 응답한 빈도를 계산하였다. 자세한 방법은 각주에 있는 웹사이트를 참고할 수 있고²⁾, 값이 클수록 불성실 응답으로 간주할 수 있다.

시간

설문응답에 걸리는 총 시간은 조사회사에서 자동적으로 산출해주었으며, 앞서 언급한 것처럼 전체 설문 시간을 감안하여 문항 당 평균 2초 이하로 응답한 응답자들을 불성실 응답으로 간주하여 1로 코딩하고, 나머지를 0으로 코딩하였다.

결 과

본 연구에서 계산된 불성실 응답 지표들의 평균, 표준편차, 상관계수가 표 1에 제시되어 있다. 불성실 응답 지표들 간의 상관을 살펴보면 작거나 중간 정도의 상관관계가 나타나거나 혹은 상관관계가 없는 경우도 존재하는데, 이러한 결과는 단일한 지표로는 다양한 형태로 발현되는 불성실 응답을 온전히 파악하지 못할 수 있음을 시사한다. 예를 들어, Meade와 Craig(2012)는 최대 롱스트링은 의도적인 불성실 응답을 탐지하기 위한 방법으로 이 지표만으로는 무선응답이나 비일관적인 응답 형태를 탐지할 수 없다고 하였고, 반대로 (비)일관성 접근법에 해당하는 방법들은 의도적 불성실 응답을 탐지할 수 없기 때문에 두 가지의 지표를 보완적으로 사용할 필요성을 제기하였다. 자기보고식으로 측정하였기에 비교적 주관적인 지표에 해당하는 자기보고식 성실/흥미의 경우, 통계적인 계산을 통해 도출된 객관적 지표들과 일정 수준의 상관이 나타났다. 따라서 불성실 응답의 지표로 사용할 수는 있으나 본 연구에서 측정한 사회적 바람직성 척도의 하위 차원인 인상관리와 정적상관(성실, 흥미 각 $r = .21, r = .15, p < .01$)이 나타나는 것으로 미루어보아 해석에 충분한 주의를 기울일 필요는 있다. 사용여부 질의 문항의 경우, 시간 및 위조문항 지표와 유의한 상관이 나타났으나, 사용여부 질의 문항에 '아니요'라고 표시한 빈도가 너무 낮았기에 해석상 큰 의미를 가지기는 힘들다.

상관계수에서 다소 비일관적으로 보이는 부분이 존재하는데, 먼저 심리측정적 동의어와 최대 롱스트링 간에 나타난 정적 상관이다. 하지만 Huang 등(2012)에 따르면, 어떤 구성개

2) <http://neocademic.com/2016/12/21/calculating-longstring-excel-detect-careless-responders/>

표 1. 불성실 응답 지표 10개에 대한 평균, 표준편차, 상관계수

	1	2	3	4	5	6	7	8	9	10
1	-									
2	-.29**	-								
3	.36**	-.10	-							
4	-.41**	-.13*	-.14*	-						
5	.24**	.36**	.11	-.32**	-					
6	-.21**	.34**	-.15**	.06	.37**	-				
7	-.41**	.52**	-.30**	.07	.34**	.62**	-			
8	.03	-.19**	.06	.08	-.21**	-.16**	-.23**	-		
9	.06	-.17**	.01	.02	-.18**	-.14*	-.11*	.45**	-	
10	-.07	.11	-.15*	.05	.24**	.20**	.27**	-.02	-.01	-
평균	.74	-.34	.78	161.47	9.02	.16	5.55	3.26	3.30	.02
표준편차	.19	.33	.21	32.98	9.10	.37	2.66	.35	.47	.14

주1. 1=심리측정적 동의어, 2=심리측정적 반의어, 3=홀짝 일관성, 4=마할라노비스 거리, 5=최대롱 스트링, 6= 2초 이하 응답, 7=위조문항, 8=자기보고식 성실, 9=자기보고식 흥미, 10=사용여부 질의 문항

주2. n = 308 * $p < .05$, ** $p < .01$

념을 측정할 때, 역산 문항이 없는 경우에 한하여 같은 선택지에 동일하게 응답하는 불성실 응답자가 있을 경우, 척도의 내적 일관성이 높아질 수 있다고 하였다. 또한 DeSimone, DeSimone, Harms, 및 Wood(2018)에 따르면, 최대 롱스트링의 존재가 문항 간 상관을 증가시키는 경향이 나타났다. 특히, 심리측정적 동의어의 경우 응답의 일관성을 파악하는 지표로 동일한 번호에 반복적으로 응답하는 형태인 롱스트링에 직접적인 영향을 받을 수 있다. 다음으로 마할라노비스 거리와 최대 롱스트링 간에 나타난 부적 상관이다. 이런 관계가 나타난 것에 대해서 전체 응답에 대해 각 점수별 응답 빈도(예: 206 문항의 응답 중 1을 선택한 빈도 혹은 2를 선택한 빈도)와 최대 롱스트링 간 상관관계를 살펴보았다. 그 결과

최대 롱스트링과 2 및 4를 선택한 빈도 간에는 부적 상관이 나타났으며(각 $r = -.33$, $r = -.44$, $p < .01$), 3을 선택한 빈도 간에는 정적 상관이 나타났다($r = .47$, $p < .01$). 마할라노비스 거리가 전반적인 분포에서 멀리 떨어진 극단 값을 탐지하는 지표인 반면, 최대 롱스트링이 높은 사람일수록 비교적 분포의 중간쯤에 위치하는 3을 많이 선택했기 때문에 마할라노비스 거리와는 부적 상관이 나타날 수 있을 것이다. 마지막으로, 심리측정적 반의어와 마할라노비스거리 간에 나타난 부적 상관인데, 이 둘 간의 관계에 대해서 명확한 설명을 제시하기는 어렵다. 하지만 Dunn, Heggestad, Shanock, 및 Theilgard(2018)에 따르면, 4 개의 불성실 응답지표들(위조문항, 롱스트링, 홀짝일관성, 시간) 중 두 개씩 짝을 지어서 동

시에 두 개의 지표 모두에서 불성실 응답자로 탐지되는 비율이 전체 불성실 응답자 중 약 10% 정도라는 것을 보여주었다. 다시 말하자면, 각각의 불성실 응답 지표가 불성실 응답이라는 구성개념의 서로 다른 단면을 측정하고 있으며, 불성실 응답지표들 간의 상관관계가 반드시 유의하거나 혹은 예상했던 방향으로 나오지 않을 수도 있다는 것을 간접적으로 보여주고 있다.

불성실 응답에 대한 조건 간 차이분석

먼저, 불성실 응답에 영향을 미칠 수도 있

는 이전 응답경험과, 문항 응답의 난이도에 대한 세 조건 간 비교에서는 유의한 차이가 나타나지 않았다(표 2). 다음으로 조건 별로 불성실 응답 지표에 대한 평균과 표준편차를 표 3에 제시하였으며, 불성실 응답에 대한 조건 간 차이에 대한 분산분석을 실시하였다. 다만, 성격 5요인 중에서 우호성, 성실성이 불성실 응답지표들과 관련성이 있다는 연구들에 의거하여(Dunn, Heggestad, Shanock, & Theilgard, 2018; Maniaci & Rogge, 2014), 우호성과 성실성을 공변인으로 설정한 후에 분산분석을 실시하였고 분석결과는 표 4에 제시되어 있다. 조건 간 유의한 차이가 나는 불성실 응답지표는

표 2. 응답 경험 및 문항 응답의 난이도에 대한 조건 간 비교

	df	F	sig
본 설문에서 나온 문항과 비슷한 문항에 응답해본 경험이 있는가?	2	.71	.49
문항을 읽고 의미를 파악하는 것에 어려움을 느꼈다.	2	.47	.62

표 3. 불성실 응답 지표에 대한 조건별 평균, 표준편차

불성실 응답 지표	조 건					
	통제 (n = 100)		경고 (n = 101)		거울 (n = 107)	
	m	sd	m	sd	m	sd
심리측정적 동의어	.72	.19	.74	.18	.74	.19
심리측정적 반의어	-.30	.32	-.31	.35	-.41	.30
흔적 일관성	.79	.23	.77	.22	.78	.17
미할라노비스 거리	167.14	31.78	159.92	35.54	157.65	31.13
최대 룽스트링	9.15	9.15	9.83	10.87	8.13	6.97
시간(2초 이하 응답)	.25	.44	.15	.36	.08	.28
위조문항	5.69	2.80	5.80	2.87	5.17	2.28
자기보고식 성실	3.25	.38	3.30	.35	3.23	.33
자기보고식 흥미	3.31	.54	3.29	.42	3.29	.43
사용여부 질의 문항	.03	.17	.02	.14	.01	.10

표 4. 불성실 응답 지표 10개에 대한 조건 간 평균 비교

불성실 응답 지표	df	F	p	η^2
심리측정적 동의어	2	.90	.41	
심리측정적 반의어	2	6.49	.00	.04
흔-짝 일관성	2	.52	.60	
마할라노비스 거리	2	2.38	.10	
최대 룡스트링	2	1.27	.28	
시간(문항 당 2초 이하 응답)	2	7.48	.00	.05
위조문항	2	3.44	.03	.02
자기보고식 성실	2	.72	.49	
자기보고식 흥미	2	.13	.88	
사용여부 질의 문항	2	.75	.47	

표 5. 심리측정적 반의어의 조건 간 차이에 대한 사후검정

조건	n	조정된 평균	표준오차	F	사후검정 (Fisher's LSD)
통제	100	-.28	.03	6.49**	통제 ≍ 경고
경고	101	-.30	.03		통제 > 거울**
거울	107	-.43	.03		경고 > 거울**

주) * $p < .05$, ** $p < .01$

표 6. 문항 당 2초 이하 응답자 비율의 조건 간 차이에 대한 사후검정

조건	n	조정된 평균	표준오차	F	사후검정 (Fisher's LSD)
통제	100	.26	.04	7.48**	통제 > 경고*
경고	101	.15	.04		통제 > 거울**
거울	107	.07	.03		경고 ≍ 거울

주) * $p < .05$, ** $p < .01$

심리측정적 반의어($F(2,305) = 6.49, p < .01$), 시간($F(2,305) = 7.48, p < .01$), 및 위조문항 실 응답 지표에 대해서 각 조건 간 세부적인 ($F(2,305) = 3.44, p < .05$)이었다.

표 7. 위조문항의 조건 간 차이에 대한 사후검정

조건	n	조정된 평균	표준오차	F	사후검정 (Fisher's LSD)
통제	100	5.79	.26	3.44*	통제 ≍ 경고
경고	101	5.86	.26		통제 > 거울*
거울	107	5.02	.25		경고 > 거울*

주) * $p < .05$, ** $p < .01$

차이를 파악하기 위해 공분산을 통제하고 조정된 점수를 가지고 Fisher's LSD 사후검정을 실시하였으며, 사후검정 결과는 표 5, 표 6 및 표 7에 제시되어 있다. 표 5에서 볼 수 있듯이 심리측정적 반의어에 대해서 통제조건과 경고 조건 간에는 유의한 차이가 없었고, 거울조건은 통제조건과 경고조건에 비해 심리측정적 반의어 지표의 값이 유의하게 작은 것으로 나타났다. 표 6에서는 문항 당 2초 이하의 응답을 한 응답자의 비율이 통제조건에 비해 경고 조건 및 거울조건에서 유의하게 낮은 것으로 나타났다. 하지만 경고조건과 거울조건 간에는 유의한 차이가 나타나지 않았다. 표 7에서는 위조문항에 대해서 통제조건과 경고조건 간에는 유의한 차이가 없었고, 거울조건은 통제조건과 경고조건에 비해 위조문항에 응답한 값이 유의하게 작은 것으로 나타났다.

분산분석 결과를 종합하면, 사용된 10개의 불성실 응답지표 가운데 심리측정적 반의어와 시간 및 위조문항 세 개의 지표에서 조건 간 유의한 차이가 발생하였다. 표 1의 상관계수를 참조해보면, 이 세 개의 지표는 일관된 방향으로 불성실 응답을 탐지하는 지표라고 볼 수 있다. 그리고 세 개 지표에 대한 사후검정 결과를 통제조건과 경고조건, 통제조건과 거울조건, 경고조건과 거울조건의 비교 순서로

종합하자면, 먼저 심리측정적 반의어 및 위조문항에 대해서는 통제조건과 경고조건이 유의한 차이가 나지 않았으나, 시간 지표에서는 통제조건에 비해 경고조건에서 불성실 응답이 유의하게 낮은 것으로 나타났다. 다음으로 심리측정적 반의어와 시간지표 및 위조문항 모두에서 통제조건과 거울조건 간 유의한 차이가 낮고, 통제조건에 비해 거울조건에서 불성실 응답이 유의하게 낮은 것으로 나타났다. 마지막으로 경고조건과 거울조건 간의 비교에서 시간 지표에서는 유의한 차이가 나지 않았으나, 심리측정적 반의어 및 위조문항에 대해서는 경고조건에 비해 거울조건에서 불성실 응답이 유의하게 낮은 것으로 나타났다.

추가 분석

각각의 불성실 응답 지표의 평균값이 조건에 따라 통계적으로 유의한 차이가 나지 않더라도 다소 극단적인 불성실 응답의 비율은 조건 간 차이를 보일 수 있으며, 이러한 비율의 차이는 집단 간 평균 비교 분석에서 탐지되지 않을 수 있다. 비록 낮은 비율의 불성실 응답이라도 조사비용과 관련이 있고, 수집한 데이터의 질과 향후 분석결과에 영향을 미칠 수 있다는 점에서, 통계적인 관점뿐만 아니라 실

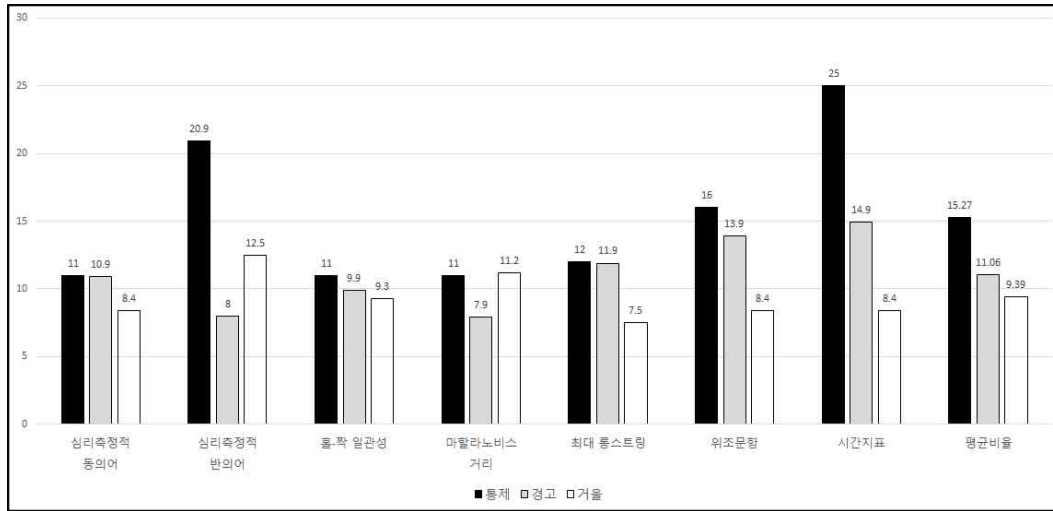


그림 3. 불성실 응답의 각 지표별 상위 10% 응답자들의 조건별 분포 비율

무적인 관점에서 각각의 불성실 응답 지표에 대해서 극단적인 불성실 응답의 비율을 조건 별로 비교해 보는 것이 중요할 수 있다.

시간 지표를 제외하고는 불성실 응답 지표의 계수를 토대로 불성실 응답의 여부를 나누는 명확한 기준 점수가 부재하기에, 각각의 불성실 응답지표 점수의 상위 10%³⁾에 해당하는 31명을 1로 코딩하고 나머지를 0으로 코딩⁴⁾하여, 1로 코딩되는 불성실 응답자의 비율

이 각 조건에 얼마나 분포되어 있는지를 비교하였다. 주관적 지표에 해당하는 자기보고식 성실/응답, 사용여부 질의 문항을 제외하고, 비교적 객관적인 정량 지표에 해당하는 지표들을 사용하였으며, 결과는 그림 3에 제시되어 있다. 마할라노비스 거리를 제외한 모든 지표에서 통제조건에 비해 경고 및 거울조건 응답자들의 비율이 낮은 것으로 나타났으며, 불성실 응답 비율의 평균은 거울 조건에서 가장 낮으며, 통제 조건에서 가장 높은 것으로 나타났다. 각각의 불성실 응답 지표에서 불성실 응답으로 분류된 비율의 평균값에 대한 조건 간 비교⁵⁾에 따르면(표 8-9), 경고조건과 거울조건 간에는 유의한 차이가 나타나지 않았으나, 통제조건에 비해 거울조건 및 경고조건의 불성실 응답자 비율의 평균값이 유의하게 낮은 것으로 나타났다.

3) Maniaci와 Rogge(2014)의 연구에서는 극단적인 불성실 응답자의 비율이 대략 9%라고 언급하였으며, Meade와 Craig(2012)에서 실시한 잠재 프로파일 분석에서는 약 11%의 응답자가 불성실 응답으로 구분되었다. 또한 Dunn 등(2018)에서는 각 불성실 응답 지표에 대해 상위 10%에 해당하는 사람들을 극단적 불성실 응답자로 간주할 것에 대한 가이드라인을 제시하였기 때문에 극단적 불성실 응답자의 기준을 10%로 삼고자 하였다.

4) 31번째 사람의 불성실 응답 지표의 계수와 동일한 사람의 범위까지 1로 코딩하였기 때문에 특정 지표에 있어서는 1로 코딩되는 사람들의 비율이 10%보다 약간 높은 경우가 있다.

5) 7개의 불성실 응답 지표에 대해 10%에 속하는 응답자들을 불성실=1로 나머지를 성실=0으로 코딩한 7개의 더미 변인들의 평균값에 대해 조건 간 비교 분석을 실시하였다.

표 8. 각 불성실 응답지표에서 불성실로 분류된 비율의 평균값에 대한 조건 간 비교

	df	F	p	η^2
불성실 응답 비율의 평균	2	6.06	.00	.04

표 9. 불성실 응답 비율 평균값의 조건 간 차이에 대한 사후검정

조건	n	평균	표준편차	F	사후검정 (Fisher's LSD)
통제	100	.15	.20	6.06**	통제 > 경고**
경고	101	.09	.14		통제 > 거울**
거울	107	.08	.13		경고 ≍ 거울

주) ** $p < .01$

논 의

본 연구에서는 온라인 설문에서 나타나는 응답자들의 불성실 응답을 방지하기 위해 사회심리학의 연구들 및 이론을 적용하여 거울 이미지를 배경화면처럼 넣은 거울조건을 아무 조작도 하지 않은 통제조건과 비교하여 실제 불성실 응답 방지의 효과가 나타나는지를 탐색적으로 살펴보았다. 또한 지시문에 경고 문구를 삽입하는 기존의 방식과 거울이미지를 넣은 방식 간의 불성실 응답 방지 효과에 있어 차이가 있는지를 검정하였다.

조사 회사를 통해 전국 소재 2년 혹은 4년제 대학에 다니고 있는 308명의 대학생들을 대상으로 온라인 설문을 실시하였으며, 수집한 데이터를 바탕으로 불성실 응답 지표에 대한 조건 간 비교분석을 실시하였다. 분석에 사용된 총 10개의 불성실 응답 지표들 중 3개의 지표(심리측정적 반의어, 시간, 위조문항)에서 통제조건에 비해 거울조건에 응답자의 불성실 응답이 감소하는 것으로 나타났다. 또한

불성실 응답자의 비율을 조건 간 비교하기 위한 추가분석에서는 통제조건에 비해 거울 조건의 불성실 응답 비율이 유의하게 낮게 나타났기 때문에 거울 이미지의 활용이 불성실 응답을 방지할 가능성을 제시했다. 이러한 결과는 개인의 분명한 의도나 의식적인 지각없이도 단순히 특정 사물에 노출되는 것만으로도 그 사물과 관련된 태도나 고정관념이 자동적으로 활성화되고(Devine, 1989; Perdue & Gurtman, 1990; Pratto & Bargh, 1991; Uleman & Bargh, 1989), 실제 행동이나 생각에 영향을 미칠 수 있다는 연구들(Bargh, Gollwitzer, Lee-Chai, Barndollar, & Trötschel, 2001; Gibson & Zielaskowski, 2013; Payne, Brown-Iannuzzi, & Loersch, 2016; Weingarten, Chen, McAdams, Yi, Hepler, & Albarracín, 2016)과 같은 맥락이다. 관련 연구의 예시로, 특정한 단어에 노출되는 것이 개인의 행동에 미치는 영향을 살펴본 Payne 등(2016)의 연구에서는 실험참가자들에게 카드 게임의 일종인 블랙잭을 하도록 지시하였다. 참가자들은 수차례 진행되는 게임에

서 점수를 따기 위한 목적의 배팅(betting)을 할 것인지 혹은 다음 게임으로 그냥 넘어갈 것인지를 결정해야 했다. 본 연구는 개인 내 설계(within-subject design) 연구로써, 몇몇의 시행에서는 참가자들이 의사결정하기 직전에 도박, 노름(gamble, wager)과 같은 단어를 점화시켰으며, 다른 몇몇의 시행에서는 게임을 포기하거나 중단하는 의미인 fold, stay와 같은 단어를 의사결정 직전에 점화시켰다. 실험 결과, fold나 stay를 점화시켰을 때에 비해 gamble과 wager를 점화시킨 경우에 배팅 빈도가 높아지는 것으로 나타났다. 이러한 결과에 대해 저자들은 도박, 노름과 같은 단어의 점화는 단어와 관련된 정보, 생각, 및 느낌에 대해 보다 접근가능하게 만들기 때문에 블랙잭 게임이라는 상황에서 할 수 있는 특정한 행동 옵션(ex: 배팅)에 영향을 미친 것이라고 해석했다. 이 연구에 비추어보았을 때, 거울조건의 응답자들이 실제로 거울을 본 것은 아니지만 배경에 제시된 거울이미지에 지속적으로 노출됨으로써 거울과 관련된 의미체계(예: 진실), 지식(예: 자신을 비출 수 있는 도구), 혹은 거울을 보았던 경험들이 활성화되었을 것이다. 이를 통해 객관적 자기인식이 유발되었고, 지시문에서 제시한 '성실한 응답 요구'라는 규범에 더욱 더 부합하는 행동을 한 것으로 추측해볼 수 있다.

사용한 10개의 불성실 응답 지표들 중에서 3개 지표에서만 유의한 차이를 보인 것으로 나타났다. 하지만 자신을 좋게 포장하기 위해 거짓응답이 가능한 자기보고식 성실도, 흥미도, 혹은 사용여부 질의 문항과 같은 주관적인 불성실응답 지표(DeSimone, Harms, & DeSimone, 2015)보다는 통계적 분석을 통해 얻은 지표인 심리측정적 반의어, 답이 명확히

있는 위조문항, 그리고 응답시간과 같은 객관적인 지표에서 유의한 차이가 나타났다는 점에서 의미가 있다. 또한 앞서 언급한 불성실 응답이 신뢰도, 검정력, 제 1종 오류 혹은 제 2종 오류에 미치는 영향력을 감안할 때, 불성실 응답이 조금이라도 방지된다면 자료의 질과 분석결과의 타당성에 긍정적인 영향을 미칠 수 있을 것이다.

경고조건과 거울조건의 비교에서, 시간 지표에서는 경고조건과 거울조건 간 성실한 응답의 정도가 차이가 없었으나, 심리측정적 반의어 및 위조문항 지표에 대해서는 경고조건에 비해 거울조건의 응답자들이 더 성실하게 응답한 것으로 나타났다. 앞서 언급한 것처럼, 응답자들에게 경고를 하는 것은 오히려 응답자들에게 강압적이고 부정적인 느낌을 줄 수 있으며, 성실한 응답에 대한 동기를 높이기보다는 저하시킬 가능성이 있다. 다른 해석의 가능성은 경고의 강도이다. Huang 등(2012)의 연구에서 사용된 경고 문구는 불성실하게 응답하는 경우 학점에 부정적인 영향을 받을 수 있다고 비교적 강하게 경고하였다. 하지만 본 연구의 온라인 설문에서는 응답자들에게 주는 보수를 실제로 조정할 수도 없었고, 실제 조정이 가능하다 할지라도 그런 내용의 경고를 받을 수 있는 상황이 아니기 때문에 경고의 강도가 조금은 약했을 수 있다. 따라서 기존의 연구들에 비해서 본 연구에서는 경고의 효과성이 다소 감소했을 것으로 추측되기도 한다. 따라서 경고조건과 통제조건 간의 차이도 시간지표 하나에서만 차이가 났을 가능성도 있다.

하지만 경고의 효과성이 실제로 어떠한가에 응답자들에게 경고를 하는 것은 다소 부정적이고 강압적인 방법으로 응답자들이 기분

을 상하게 할 염려가 있으며, 성실한 태도로 설문에 임하고자 마음먹은 응답자들의 동기도 저하시킬 가능성이 충분히 있다(Meade & Craig, 2012). 반면, 거울이미지를 활용했을 때는 일반적인 지시문을 활용해도 되며, 응답자들의 기분을 상하게 할 염려가 없다. 조사 회사를 통해 추가적으로 118명의 대학생들을 대상으로 거울조건의 설문을 진행한 결과, 118명 중 76%가량이 설문화면에 아무 것도 없었다고 응답했으며, 오직 17%의 학생들만이 거울의 존재를 인식했다. 대다수의 사람들이 거울의 존재를 인식하지 못했으며, 인식한다 하더라도 거울 이미지가 문항을 읽고 응답하는 것에 큰 무리를 주지 않는다는 것이 본 연구결과에도 나타났기 때문에 경고문구 보다는 거울 이미지를 활용하는 것이 더 효과적이라고 할 수 있다.

또한 Ward와 Meade(2018)에서 사용된 지시문을 연구자가 직접 읽는 비디오 클립을 활용하는 방식이라든가, Ward와 Pond(2015)에서 사용된 가상인간의 삽입에 비해서 단순하며 개발에 필요한 시간이나 비용이 거의 들지 않기 때문에 거울이미지를 활용한 방식이 보다 경제적이고 효율적인 방법이라고 볼 수 있다. 기존에 불성실 응답을 사후적으로 탐지하여 제거하는 방법들은 많으나, 이는 표본 분포를 왜곡시켜서 표본의 대표성을 떨어뜨리며, 결과의 외적타당도를 저하시킬 수 있다는 점(Meade & Craig, 2012)을 감안할 때, 불성실 응답을 방지할 수 있는 새로운 방법을 제안하고 그 가능성을 검증했다는 점에서 본 연구의 의의가 있다.

사회과학 연구, 조직의 실무, 일반대중 혹은 고객들을 대상으로 온라인 설문을 사용하는 것이 일반화되어 있는 상황에서, 불성실 응답

에 대한 이해의 증진은 다양한 변인들의 측정에 대한 정확성과 타당도를 향상시키는 데 기여할 수 있다. 이러한 관점에서 본 연구의 결과는 온라인 조사에서 응답의 질을 높이려는 연구자 및 실무자들의 시도에 도움이 될 수 있을 것이며, 불성실 응답을 방지할 수 있는 새로운 접근법을 제기했다고 볼 수 있다.

연구의 제한점 및 추후 연구 제안

본 연구의 한계점과 향후 연구 방향에 대한 논의는 다음과 같다.

첫째, 통제조건에 비해서 거울조건에서 상대적으로 불성실 응답방지의 효과가 나타났기 때문에 거울조건에서 객관적 자기인식이 유발되었을 것이라고 가정하였다. 하지만 객관적 자기인식의 유발이나 증가를 직접적으로 측정 한 것이 아니기 때문에, 추후 연구에서는 객관적 자기인식을 측정하는 도구를 활용하여 통제조건과 거울조건에서 실제로 객관적 자기인식의 수준에서 유의한 차이가 나는지를 검증 할 필요가 있다.

둘째, 지시문에 제시된 성실한 응답에 대한 기준을 보다 현저하게 만들 필요가 있다. 본 연구에서 성실한 응답에 대한 지시문은 단지 밑줄을 긋고 글을 진하게 만든 정도이다. 하지만 지시문을 자세히 읽지 않고 넘어가는 응답자들도 상당수 있으며, 이러한 응답자들은 성실한 응답에 대한 기준 자체를 인식하지 못했을 수 있기 때문에 거울이미지의 효과가 이들에게 제대로 나타나지 않았을 가능성도 있다. 따라서 추후연구에서는 성실한 응답과 관련된 질문을 던짐으로써(예: 다음 중 성실한 응답형태로 적절한 것은 무엇인가요?) 응답자

들이 성실한 응답에 대한 기준을 보다 명확하게 인지시키고, 그러한 기준 인식 제고가 거울의 불성실 응답방지 효과를 더 크게 만들 가능성에 대해서도 조사할 필요가 있다.

셋째, 객관적 자기인식을 유발할 수 있는 다양한 방법들 가운데 거울만을 활용했다는 점이다. 기존 연구들을 살펴보았을 때, 객관적 자기 인식을 유발하기 위해 온라인에서 가능한 방법들로는 개인적인 일화 회상하기 (Govern, Marsch, 2001), 다른 사람들에 비해 자신을 독특하게 만드는 특성 혹은 타인과 나를 구분 짓는 특성 적어보기(Silvia & Eichstaedt, 2004), SNS의 프로필 업데이트하기(Gonzales, & Hancock, 2011), 개인의 자서전적 정보에 대한 질문하기(Duval, Duval, & Neely, 1979) 등과 같이 다양하다. 따라서 이러한 방법들을 활용하여 실제 불성실 응답의 방지효과가 나타나는지, 거울이미지를 활용한 방법과는 어떤 차이가 있는지를 탐색해보는 것이 추후의 연구방향이 될 수 있다. 또한 객관적 자기 인식 이론 이외에 다른 이론이나 방법을 적용함과 동시에 설문지의 길이에 따른 불성실 응답 방지를 위한 최적의 조합을 찾는 것도 중요한 연구가 될 것이다. 가령, 다소 긴 설문에서는 거울이미지 사용과 더불어 응답자가 설문 응답에 대한 내적 동기를 느낄 수 있도록 설문지의 중간 중간 성실하게 답하고 있다는 피드백을 주는 방법을 활용하거나, 다소 짧은 설문에서는 거울이미지와 Ward와 Meade(2018)의 연구 2에서 사용한 위선 패러다임(hypocrisy paradigm) 등을 동시에 활용할 수 있을 것이다.

마지막으로, 본 연구에서는 대학생 표본만을 대상으로 연구를 진행했기에, 결과의 일반화 측면에서 다소 제약이 있다. 온라인 설문 참여자들은 매우 다양하기 때문에 표본의

특성에 따라서 거울의 효과가 다르게 나타나는 않는지 검증이 필요하다. 추후 연구에서는 다양한 표본을 대상으로 연구를 진행하여, 거울조건의 효과가 다양한 표본에 걸쳐서 일반화될 수 있는지에 대한 확인이 필요하다.

참고문헌

- 김병수, 김정현 (2013). 스트레스 척도 핸드북. 서울: 학지사
- 배성우, 신원식 (2005). CES-D 척도(The Center for Epidemiologic Studies-Depression Scale)의 요인구조 분석. *보건과 사회과학*, 18, 165-190.
- 이양현, 임효덕, 이종영 (1996). 한국판 20 항목 Toronto 감정표현불능증 척도(TAS-20K)의 개발과 타당도. *신경정신의학*, 35(4), 888-899.
- 최보라 (2007). 사회적 바람직성 척도 타당화. 이화여대석사학위논문.
- Aronson, E., Fried, C., & Stone, J. (1991). Overcoming denial and increasing the intention to use condoms through the induction of hypocrisy. *American Journal of Public Health*, 81(12), 1636-1638.
- Baer, R. A., Ballenger, J., Berry, D. T., & Wetter, M. W. (1997). Detection of random responding on the MMPI-A. *Journal of Personality Assessment*, 68, 139-151.
- Bargh, J. A., Gollwitzer, P. M., Lee-Chai, A., Barndollar, K., & Trötschel, R. (2001). The automated will: Nonconscious activation and pursuit of behavioral goals. *Journal of Personality and Social Psychology*, 81, 1014-1027.

- Beaman, A. L., Klentz, B., Diener, E., & Svanum, S. (1979). Self-awareness and transgression in children: Two field studies. *Journal of Personality and Social Psychology, 37*, 1835-1846.
- Berkowitz, L., & LePage, A. (1967). Weapons as aggression-eliciting stimuli. *Journal of Personality and Social Psychology, 7*, 202-207.
- Berry, D. T., Wetter, M. W., Baer, R. A., Larsen, L., Clark, C., & Monroe, K. (1992). MMPI-2 random responding indices: Validation using a self-report methodology. *Psychological Assessment, 4*, 340-345.
- Bowling, N. A., Huang, J. L., Bragg, C. B., Khazon, S., Liu, M., & Blackmore, C. E. (2016). Who cares and who is careless? Insufficient effort responding as a reflection of respondent personality. *Journal of Personality and Social Psychology, 111*, 218-229.
- Breitsohl, H., & Steidelmüller, C. (2018). The impact of insufficient effort responding detection methods on substantive responses: Results from an experiment testing parameter invariance. *Applied Psychology, 67*(2), 284-308.
- Buehl, S., Lofland, K. R., Sherman, J. J., & Carlson, C. R. (1998). The Variable Responding Scale for detection of random responding on the Multidimensional Pain Inventory. *Psychological Assessment, 10*(1), 3-9.
- Buechley, R., & Ball, H. (1952). A new test of "validity" for the group MMPI. *Journal of Consulting Psychology, 1*(4), 299-301.
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*(1), 3-5.
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). Minnesota multiphasic personality inventory-2 (MMPI-2): *Manual for administration and scoring*. Minneapolis: University of Minnesota Press.
- Clark, M. E., Girona, R. J., & Young, R. W. (2003). Detection of back random responding: Effectiveness of MMPI-2 and Personality Assessment Inventory validity indices. *Psychological Assessment, 15*, 223-234.
- Carver, C. S., & Scheier, M. S. (1981). *Attention and self-regulation: A control theory approach to human behavior*. New York: Springer-Verlag.
- Costa, P. T., & McCrae, R. R. (2008). The revised neo personality inventory (neo-pi-r). *The SAGE handbook of personality theory and assessment, 2*, 179-198.
- Credé, M. (2010). Random responding as a threat to the validity of effect size estimates in correlational research. *Educational and Psychological Measurement, 70*(4), 596-612.
- Curran, P. G., Kotrba, L., & Denison, D. (2010). Careless responding in surveys: Applying traditional techniques to organizational settings. In *25th annual conference of Society for Industrial and Organizational Psychology, Atlanta, GA*.
- DeSimone, J. A., DeSimone, A. J., Harms, P. D., & Wood, D. (2018). The differential impacts of two forms of insufficient effort responding. *Applied Psychology, 67*(2), 309-338.
- DeSimone, J. A., Harms, P. D., & DeSimone, A. J. (2015). Best practice recommendations for data screening. *Journal of Organizational*

- Behavior*, 3(2), 171-181.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of personality and social psychology*, 5(1), 5-18.
- Diener, E., & Wallbom, M. (1976). Effects of self-awareness on antinormative behavior. *Journal of Research in Personality*, 10, 107-111.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: the tailored design method*. Hoboken: John Wiley & Sons.
- Dunn, A. M., Heggstad, E. D., Shanock, L. R., & Theilgard, N. (2018). Intra-individual response variability as an indicator of insufficient effort responding: Comparison to other indicators and relationships with individual differences. *Journal of Business and Psychology*, 33(1), 105-121.
- Duval, S., Duval, V. H., & Neely, R. (1979). Self-focus, felt responsibility, and helping behavior. *Journal of Personality and Social Psychology*, 37(10), 1769-1778.
- Duval, S., & Wicklund, R. A. (1972). *A theory of objective self awareness*. New York: Academic Press.
- Ehlers, C., Greene-Shortridge, T. M., Weekley, J. A., & Zajack, M. D. (2009). The exploration of statistical methods in detecting random responding. In *Annual Meeting of the Society for Industrial/Organizational Psychology*, Atlanta, GA.
- Gibson, B., & Zielaskowski, K. (2013). Subliminal priming of winning images prompts increased betting in slot machine play. *Journal of Applied Social Psychology*, 43, 106-115.
- Gonzales, A. L., & Hancock, J. T. (2011). Mirror, mirror on my Facebook wall: Effects of exposure to Facebook on self-esteem. *Cyberpsychology, Behavior, and Social Networking*, 14(1-2), 79-83.
- Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making*, 2(3), 213-224.
- Govern, J. M., & Marsch, L. A. (2001). Development and validation of the situational self-awareness scale. *Consciousness and Cognition*, 10(3), 366-378.
- Greene, R. L. (1978). An empirically derived MMPI carelessness scale. *Journal of Clinical Psychology*, 34(2), 407-410.
- Guay, R. P. Oh, I.-S. Choi, D., Mitchell, M. S., Mount, M. K., & Shin, K. (2013). The interactive effect of conscientiousness and agreeableness on job performance dimensions in South Korea. *International Journal of Selection and Assessment*, 21, 233-238.
- Haertzen, C. A., & Hill, H. E. (1963). Assessing subjective effects of drugs: An index of carelessness and confusion for use with the Addiction Research Center Inventory (ARCI). *Journal of Clinical Psychology*, 19(4), 407-412.
- Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, 1, 104-121.
- Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to

- surveys. *Journal of Business and Psychology*, 27, 99-114.
- Huang, J. L., Liu, M., & Bowling, N. A. (2015). Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology*, 100(3), 828-845.
- Ickes, W. J., Wicklund, R. A., & Ferris, C. B. (1973). Objective self awareness and self esteem. *Journal of Experimental Social Psychology*, 9(3), 202-219.
- Johnson, J. A. (2005). Ascertaining the validity of individual protocols from web-based personality inventories. *Journal of Research in Personality*, 39(1), 103-129.
- Kanfer, F. H. (1970). Self-regulation: Research, issues, and speculations. *Behavior Modification in Clinical Psychology*, 7(4), 178-220.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254-284.
- Kreuter, F., Yan, T., & Tourangeau, R. (2008). Good item or bad-can latent class analysis tell?: the utility of latent class analysis for the evaluation of survey questions. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171, 723-738.
- Lawson, G., Stedmon, A. W., Zhang, K., Eubanks, D. L., & Frumkin, L. A. (2013). The effects of self-awareness on body movement indicators of the intention to deceive. *Applied Ergonomics*, 44(5), 687-693.
- Liu, M., Bowling, N. A., Huang, J. L., & Kent, T. A. (2013). Insufficient effort responding to surveys as a threat to validity: The perceptions and practices of SIOP members. *The Industrial-Organizational Psychologist*, 51, 32-38.
- Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, 48, 61-83.
- McGrath, R. E., Mitchell, M., Kim, B. H., & Hough, L. (2010). Evidence for response bias as a source of error variance in applied assessment. *Psychological Bulletin*, 136(3), 450-470.
- Mead, G. H. (1934). *Mind, self and society* (Vol. 111). Chicago: University of Chicago Press.
- Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17, 437-455.
- Morey, L. C. (1991). *Personality assessment inventory*. Odessa, FL: Psychological Assessment Resources.
- Mueller-Hanson, R., Heggestad, E. D., & Thornton III, G. C. (2003). Faking and selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology*, 88, 348-355.
- Niessen, A. S. M., Meijer, R. R., & Tendeiro, J. N. (2016). Detecting careless respondents in web-based questionnaires: Which method to use?. *Journal of Research in Personality*, 63, 1-11.
- Northcraft, G. B., Schmidt, A. M., & Ashford, S. J. (2011). Feedback and the rationing of time and effort among competing tasks. *Journal of Applied Psychology*, 96(5), 1076-1086.

- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*, 660-679.
- Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*(4), 867-872.
- Payne, B. K., Brown-Iannuzzi, J. L., & Loersch, C. (2016). Replicable effects of primes on human behavior. *Journal of Experimental Psychology: General, 145*, 1269-1279.
- Perdue, C. W., & Gurtman, M. B. (1990). Evidence for the automaticity of ageism. *Journal of Experimental Social Psychology, 26*(3), 199-216.
- Pinsooneault, T. B. (2005). Detecting random, partially random, and nonrandom Minnesota Multiphasic Personality Inventory-Adolescent Protocols. *Psychological Assessment, 17*(4), 476-480.
- Pratto, F., & Bargh, J. A. (1991). Stereotyping based on apparently individuating information: Trait and global components of sex stereotypes under attention overload. *Journal of Experimental Social Psychology, 27*(1), 26-47.
- R Development Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Schultz, D. P. (1969). The human subject in psychological research. *Psychological Bulletin, 72*, 214-228.
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54*, 93-105.
- Silvia, P. J., & Duval, T. S. (2001). Objective self-awareness theory: Recent progress and enduring problems. *Personality and Social Psychology Review, 5*, 230-241.
- Silvia, P. J., & Eichstaedt, J. (2004). A self-novelty manipulation of self-focused attention for Internet and laboratory experiments. *Behavior Research Methods, Instruments, & Computers, 36*(2), 325-330.
- Silvia, P. J., & Phillips, A. G. (2013). Self-awareness without awareness? Implicit self-focused attention and behavioral self-regulation. *Self and Identity, 12*, 114-127.
- Spelke, E., Hirst, W., & Neisser, U. (1976). Skills of divided attention. *Cognition, 4*, 215-230.
- Stone, J., Aronson, E., Crain, A. L., Winslow, M. P., & Fried, C. B. (1994). Inducing hypocrisy as a means of encouraging young adults to use condoms. *Personality and Social Psychology Bulletin, 20*(1), 116-128.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. New York: Cambridge University Press.
- Uleman, J. S., & Bargh, J. A. (Eds.). (1989). *Unintended thought*. New York: Guilford Press.
- Ward, M. K., & Meade, A. W. (2018). Applying social psychology to prevent careless responding during online surveys. *Applied Psychology, 67*, 231-263.
- Ward, M. K., & Pond III, S. B. (2015). Using virtual presence and survey instructions to

- minimize careless responding on Internet-based surveys. *Computers in Human Behavior*, 48, 554-568.
- Weingarten, E., Chen, Q., McAdams, M., Yi, J., Hepler, J., & Albarracín, D. (2016). From primed concepts to action: A meta-analysis of the behavioral effects of incidentally presented words. *Psychological Bulletin*, 142, 472-497.
- Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment*, 28, 186-191.
- 1차원고접수 : 2019. 07. 23.
2차원고접수 : 2019. 09. 08.
3차원고접수 : 2019. 10. 25.
최종게재결정 : 2019. 12. 21.

A study on the effectiveness of the use of mirror-image as a preventive measure of careless responding in online survey research

Woo Young Kim

Sogang University

Tae Hun Lee

Chungang University

Jae Yoon Chang

Sogang University

Online surveys are widely used not only in general survey research but also in organizational research, such as customer and employee surveys, due to the ease of implementation and the considerations of cost-effectiveness compared to conventional paper-pencil surveys. There has been much research devoted to showing the negative impact of careless responding on subsequent data analysis and to evaluating statistical methods for detecting the careless respondents embedded in the data collected via online surveys. However, it would be fair to say that there has been a paucity of empirical research evaluating the effectiveness of a preventive measure against careless responding that would operate during the data collection phase of online survey. In this study, inspired by the objective self-awareness theory in social psychology, we proposed a simple but novel method of inserting a mirror-image in the background of each online survey page as a preventive measure of careless responding to the survey items. We collected data under three different conditions of control(standard instructions only), warning(standard instructions plus warnings against careless responding), and the mirror-image condition. Statistical properties of several existing indicators developed for detecting various types of careless respondents were compared among the three conditions. Our results suggested that the simple mirror-image method tended to reduce the careless responding more effectively than the warning and control condition. We concluded by discussing implications and limitations of our study.

Key words : *careless responding, mirror, objective self-awareness, online survey*