

## 자기-보고형 조사에서의 불성실 응답(IER)의 원인, 탐지, 예방: 개관 및 향후 연구 제안\*

장 재 윤                      김 우 영                      이 태 현<sup>†</sup>  
서강대학교 심리학과              SK 이노베이션              고려대학교 심리학부


본 고에서는 최근 사회조사 및 심리측정 과정에서 빈번하게 발생하는 응답자들의 불성실응답(IER: Insufficient Effort Responding)의 발생원인, 영향, 탐지 방법 및 사전적 예방법에 대한 개관을 통하여 조사 방법론을 적용하는 연구자와 실무자에게 유용한 가이드라인을 제공하고자 하였다. 우선 IER은 개인차 요인(성격 특성, 동기 수준, 인지 능력 등)과 상황 요인(조사 길이, 수집 방식, 주변 환경의 방해, 연구자와의 상호작용 부족, 문화 간 차이 등)에 의해 발생하며, 측정의 신뢰도와 타당도를 감소시키고 가설 검정의 오류를 증가시키는 등 연구나 조사 결과에 부정적인 영향을 미침을 보여주는 여러 연구를 개관하였다. 이어서 IER의 부정적 영향을 최소화하는 방안으로 IER을 사전에 방지하거나 사후에 효과적으로 탐지할 수 있는 다양한 방법들을 소개하였다. 이를 통해 데이터의 질과 연구의 신뢰성을 높일 수 있는 유용한 조언을 제시하였다. 마지막으로, 문화 간 차이를 고려한 IER 연구와 새로운 측정 도구 및 기술을 활용하여 IER 문제를 해결할 가능성을 탐색하는 향후 연구 방향을 제안하였다.

주요어 : 불성실 응답, 부주의 응답, 반응 편향, 자기 보고, 사회 조사, 심리측정

\* 이 논문은 2021년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2021S1A5B1096665).

IER 탐지 관련 웹 페이지 구성 작업을 해준 사현서(서강대 심리학과 석사과정) 군에게 감사드린다.

<sup>†</sup> 교신저자: 이태현, 고려대학교 심리학부 교수, (02841) 서울특별시 성북구 안암로 145,  
E-mail: taehun0626@korea.ac.kr

 Copyright © 2024, The Korean Psychological Association. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial Licenses(<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution and reproduction in any medium, provided the original work is properly cited.

사회 구성원들이 무엇을 생각하고 어떻게 행동하는지를 관찰하고 기록할 수 있게 하는 도구로서의 사회조사는 현대 사회과학의 근간이 되는 요소 중 하나다(한신갑, 2015). 이러한 조사에는 자신을 둘러싼 환경이나 대상에 대한 인식이나 태도와 더불어 개인적인 선호, 특성, 행동 등에 관한 질문을 포함한다. 정책이나 선거 관련 각종 여론조사, General Social Survey와 같은 사회조사, 조직 구성원 대상의 조직문화나 만족도 조사, 심리학자가 연구 목적으로 개인들을 대상으로 하는 성격검사, 기업조직에서의 인사 선발, 정신 건강 관련 평가나 포렌식 검사 등과 같이 다양한 상황에서 진술문 형태의 문항에 대한 동의 또는 일치 정도를 묻는 조사가 셀 수 없이 이루어지고 있다.

물리적 현상이나 속성의 측정과 마찬가지로, 사회 구성원들의 외부 환경 지각이나 내적 속성에 관한 신뢰롭고 타당한 측정은 사회 및 심리과학의 학문적 발전에 있어 가장 기본적인 부분이다. 그러나, 지능 검사와 같은 인지적 측정과는 다르게 자기 보고(self-report) 형식의 조사에서는 응답자들의 반응에 대한 정(正)/오(誤) 판단이 어렵다. 응답자들은 익명이거나 특별한 이해가 없으면 불성실하게 응답할 가능성이 있으며, 자신이 드러나거나 응답 결과에 따른 이해관계가 있으면 의식적(혹은 무의식적)으로 자신에게 유리하게 응답을 왜곡하기도 한다.

응답자들의 부주의하거나 왜곡된 반응은 측정을 오염시켜 측정된 변수의 신뢰도와 타당도를 낮추며, 방법 편향(method bias)의 원천으로서 요인 분석 결과나 변인 간의 관계를 왜곡시킬 수 있다(Huang et al., 2015). 즉, 타당한 측정에 기반하지 않은 계량적 분석들은 1,

2종 오류의 가능성이 크며 부정확한 결론과 잘못된 의사결정에 이르게 한다.

사회과학 영역에서 여러 형태의 변인들을 측정하거나 심리학에서 주관적인 반응을 요구하는 태도 조사나 성격을 평가할 때는 오래전부터 Likert 척도에 의한 자기 보고형 측정치가 광범위하게 사용되어왔다.<sup>1)</sup> 동시에 이런 측정에 내재된 잠재적 문제들도 인식해왔으며, 특히 심리학자들은 20세기 초부터 심리측정 과정에서 부정확하게 응답하는 응답편향(response biases)을 확인하거나 방지하는 방법들을 찾으려고 노력하였다(Marston, 1917; McGrath et al., 2010). 더불어 최근 조사나 실험 자료를 얻기 위해 MTurk이나 Prolific과 같은 온라인 플랫폼을 이용하는 연구자들이 늘면서, 이러한 플랫폼을 통해 크라우드소싱된 참가자들의 응답 자료의 질이나 표본의 대표성에 대한 논란이 계속되고 있다(Douglas et al., 2023; Reimers et al., 2022).

이러한 맥락에서 본 연구는 사회 및 심리학에서 관심을 가지는 구성개념들(constructs)의 신뢰롭고 타당한 측정을 저해하는 대표적인 반응 편향으로 조사나 검사에 대해 부주의하게 반응하는 부주의 응답(CR: careless responding) 또는 불성실 응답(IER: insufficient effort responding)에 주목하였다(앞으로 IER로 표기함). 금전적 이유로 온라인 응답에 참여하는 사람들이 늘어나면서 IER은 더욱 심각한 문제가 되고 있어, 이것의 탐지 및 방지에 대

1) 성격검사같이 심리학적 구성개념을 측정하는 경우는 복수의 문항으로 구성된 척도를 사용하지만, 여론조사와 같은 사회조사에서는 대부분 단일 문항을 사용한다. 어떤 경우이든지 불성실한 응답이나 응답 왜곡은 신뢰로운 측정을 방해하는 오차 요인이다.

한 국내외의 관심이 늘고 있으며 관련 개관 논문들도 발표되고 있다(박원우 등, 2020; Arthur et al., 2021; Ward & Meade, 2023). 본 연구는 자기 보고형 조사(측정)에서 나타나는 IER이 조사(측정) 결과에 미치는 영향, 그것들을 탐지하고 사후 조치하는 방법, 그리고 사전에 방지하는 방안과 관련된 기존 연구를 개관하였고, 마지막으로 향후 연구 제안을 하였다.

### 자기 보고 조사에서의 응답 편향

자기 보고에 의한 변인 측정은 사회과학 영역에 널리 사용된 방식이다. 특히 인터넷의 발달로 온라인 설문과 온라인 조사 응답자 풀(pool)이 널리 활용되면서 자기 보고에 의한 측정도 급증하고 있다. 온라인 조사는 기존의 지필(paper-pencil)형 설문 대비, 시간, 편의성, 비용 측면에서 이점이 있으며, 응답자가 원하는 시간에 자유롭게 응답할 수 있어 유연성이 크다. 더불어 온라인 조사 서비스나 프로그램을 통해 정교한 연구 설계가 가능하고, 다양한 멀티미디어를 활용할 수 있다. 또한, 일부 학자들은 온라인 조사 응답자 풀이 대표성 있는 표본이라는 긍정적인 의견을 제시하기도 했다(Buhrmester et al., 2011; Goodman et al., 2013). 따라서 이제는 온라인 조사가 일반적인 추세로 자리매김하였으며, 이를 통해 수집된 데이터는 연구 결과 및 시사점 도출, 정책 수립, 인사 결정 등에 다양하게 활용되고 있다.

조사의 목적과 활용 측면에서 양질의 데이터 확보는 기본적인 사항이다. 그러나 자기 보고형 측정에서의 고질적인 문제는 다음 두 가지이다. 하나는 전통적으로 주목해왔던 것으로 사회적 바람직성(social desirability) 응답

또는 인상관리(impression management)와 같은 거짓 응답(faking)이며, 다른 하나는 불성실 응답(IER)이다. 전자는 문항을 읽고 이해한 상태에서 솔직하게 응답하지 않으려는 의도나 동기가 개입된다는 점에서 후자와는 질적으로 다르다. 또한, 전자가 고부담(high-stake) 상황에서 나타날 가능성이 크다면, 후자는 주로 저부담 상황에서 나타난다 점에서도 대비된다.

기존 조사 연구들에서 사회적으로 바람직한 응답이나 거짓 응답에 대처하려는 시도들을 통해 데이터의 질을 높이기 위한 노력을 기울여왔으나(Mueller-Hanson et al., 2003; Ones et al., 1996), 응답자들의 IER이 데이터의 질에 미치는 영향에 대한 관심은 상대적으로 낮은 편이었다. 사실상 IER은 측정오차를 유발하여 변인 간 상관을 왜곡하며, 2중 오류의 가능성을 높일 수 있다(Huang et al., 2015). 기본적으로 IER의 존재는 조사나 측정의 타당도에 심각한 위협이 되에도 불구하고, 그동안 IER의 정의 및 그것이 조사나 연구 결과에 미치는 영향은 통계 또는 조사 방법론 관련 대학원 교육에서 거의 다루어지지 않았다(Liu et al., 2013).

IER에 대한 연구자들의 인식과 관련된 조사 사례로, Liu 등(2013)은 254명의 미국 산업 및 조직심리학회 회원들을 대상으로 온라인 설문에서의 IER에 대한 태도 조사를 하였다. 'IER이 연구 결과에 미치는 영향의 정도'에 대해서는 평균 2.62점, 'IER을 방지하고, 이에 해당하는 응답 사례를 선별하기 위해 얼마나 노력을 기울이는지'에 대해서는 평균 3.13점으로 나타났다('전혀 그렇지 않다' ~ '매우 그렇다'의 5점 척도). 따라서 Liu 등은 회원들이 자료의 질에 미치는 IER의 실제적인 영향을 과소평가하고 있다고 지적했다. 특히 이 조사

에서 'IER이 자료의 질에 미치는 영향에 대한 인식'과 'IER을 방지하고 걸러내기 위한 노력' 간의 유의한 부적 상관이 나타났는데, 이는 인식과 노력 간에 부정적인 순환이 나타날 가능성이 보여주는 것이다.

또한, 조직심리 및 조직행동 분야의 탐 저널인 Journal of Applied Psychology(JAP), Journal of Management(JM), 및 Academy of Management (AMJ)의 2012년부터 2014년까지 발간된 논문들(JAP: 250개 표본, JM: 104개 표본, AMJ: 107개 표본)에서 IER을 관리하는 일반적 절차를 얼마나 적용하고 있는지를 조사한 Ran 등 (2015)에 따르면, 단 세 연구에서만 IER을 탐지하는 방법을 활용하였으며, 나머지 연구들에서는 결측치가 있는 응답자를 제외하는 정도에 그친 것으로 나타났다. 이러한 조사 결과는 탐 저널의 논문들에서도 IER에 크게 관심을 두지 않음을 보여준다. 수집된 데이터의 '질'이 결과 도출 및 해석에 중대한 영향을 끼칠 수 있다는 점에서, 이러한 IER에 대한 무관심은 시급히 개선되어야 할 관행이다.

## IER의 정의

IER은 조사나 측정에서 '주의를 기울이지 않음', '임의적인 응답 패턴', '검사 요구에 순응하려는 의지가 없음'으로 정의된다(Nichols et al. 1989, p. 240).<sup>2)</sup> 즉, 문항 내용과 관련 없

이 무작위로 응답하거나, 응답하지 않거나, 문항 내용을 잘못 이해하거나 내용과는 무관하게 응답하거나, 또는 내용이 다른 문항에 대해 동일하게 응답하는 등의 특징을 보인다(Johnson, 2005; Meade & Craig, 2012). 이는 응답자가 설문에 제시된 지시문을 따르거나, 문항 내용을 제대로 해석하거나, 정확한 응답을 제공하고자 하는 동기가 없거나 낮을 때 발생하는 행동이다.

요약하면, 불성실 응답은 크게 세 가지 특성으로 구분된다. 첫째, 문항 내용과 관련 없이 무작위로(random) 응답하는 특성으로, 이런 경우 응답의 전반적인 일관성이 떨어진다. 완전히 무작위로 응답하면 문항 간 상관은 거의 0에 가까울 것이다. 둘째, 제시된 문항을 제대로 읽거나 충분히 이해하지 않고 모든 문항에 고정된(straightlining) 응답을 하는 경우나 문항 내용과는 관계없이 일정한 패턴으로 응답하는 경우이다. 마지막으로, 빠르게 대충 응답하는 특성으로, 이 경우에는 설문 완료까지 걸리는 시간이 현저히 짧은 경우이다.

온라인 조사가 늘면서 최근 IER이 주목받고 있지만, 일부 심리학자들은 임상 척도 개발 과정에서 이 문제를 오래전부터 다루어왔다(Buechley & Ball, 1952; Haertzen & Hill, 1963). 예를 들어, MMPI나 PAI(Personality Assessment Inventory) 같은 성격 평가 도구들은 수백 문항으로 이루어져 있어, 응답자들이 계속 집중력을 유지하기 쉽지 않다. 따라서 당시 도구 개발자들은 부주의한 응답을 걸러내기 위한 척도가 필요하였다. MMPI와 PAI에서 IER을 탐지하는데 사용되었던 '비일관성 척도

2) IER을 나타내는 용어들은 다양하다. 응답자가 충분한 주의를 기울이지 않고 응답하는 패턴에 중점을 둔 용어로 'inattentive'(Maniaci & Rogge, 2014), 'inconsistent'(Greene, 1978), 'variable'(Bruehl et al., 1998) responding 등이 있다. 또한, 제시된 문항 내용과는 상관없는 응답 패턴을 일컫는 용

어로 'content non-responsivity'(Nichols et al., 1989), 'content independent responding'(Evans & Dinning, 1983) 등이 있다.

(inconsistency scale)<sup>3)</sup>는 내용이 거의 동일한 문항 쌍을 설문 전, 후반에 배치하여 불성실 응답 여부를 파악하였다. 예를 들어, “I am an active person”이 전반에 있었다면 후반에는 “I have an active lifestyle”이 배치되었다. 여러 문항 쌍에 걸쳐 두 문항 간 점수 차의 절대값을 합산하여 점수가 높을수록 불성실한 응답으로 간주한 것이다. 이 척도는 실제 데이터와 무작위로 만들어진 데이터를 효과적으로 구분하는 것으로 나타났다(Bruehl et al., 1998; Pinesault, 2005).<sup>3)</sup>

### 불성실 응답의 기저율

IER은 매우 폭넓게 발생하고 있다(Bowling et al., 2016; Ward & Pond, 2015). 극단적인 예로, Piferi와 Jobe(2003)는 설문 문항 응답지에 우연히 10개의 공란을 더 추가하였는데, 159명의 참가자 중 16명(10%)이 10개의 ‘유령 문항’에 응답하는 결과를 보여주었다. 이것은 설문 문항(적어도 뒷부분)을 제대로 읽지도 않고 응답한 것임을 나타낸다.

Meade와 Craig(2012)의 잠재 프로파일 분석에 따르면, 학점을 위해 의무적으로 긴 설문에 참여한 학부생 집단에서 약 10~12% 정도

가 IER을 보였다. Maniaci와 Rogge(2014)의 연구에서는 약 9%의 응답자들이 매우 높은 수준의 IER을 보인 것으로 나타났다. 한편, 일반 성인들을 대상으로 한 Schneider 등(2018)의 연구에서 추정된 IER의 비율은 7.4%로 학부생 표본에 비해서는 낮았다. 반면 Osborne과 Blanchard(2011)은 교수법의 효과를 검증하는 연구에서 중학생 응답자의 40% 가까이가 IER을 보였으며, 이는 연구 결과 및 시사점 도출에 심각한 위협이 될 수 있다고 하였다. Brühlmann 등(2020)도 8개의 IER 지표를 사용하였을 때<sup>4)</sup> IER에 해당하는 응답자는 전체 60%(394명 중 233명)에 이르렀다. 또한, 잠재 프로파일 분석을 하였을 때는 약 46%의 참가자들이 일정 형태의 부주의 응답 행동을 보인 것으로 나타났다.

IER의 기저율(base rate)에 대한 연구는 상대적으로 많지 않지만, 앞서 보았듯이, 기저율 추정에서 기존 연구들 간에 상당한 차이가 난다. 이는 IER에 대한 개념화 및 탐지 방법 등에서의 차이 때문이다. Ward와 Meade(2023)의 최근 보고에서도 IER을 어떻게 정의하는지에 따라서 1%에서 50%에 이르기까지 매우 넓은 범위를 갖는다고 하였고, 그들은 이렇게 넓은 범위를 보이는 이유는 IER 탐지방법과 기준이 상이하기 때문이라고 하였다. 박원우 등(2020)이 국내 대기업 구성원 3,030명의 조사 자료에 대해 다섯 가지 IER 탐지 방법을 적용하였

3) MMPI-2의 타당도 척도 중 VRIN(Variable Response Inconsistency) 및 TRIN(True Response Inconsistency) 척도가 IER 탐지에 사용된다. VRIN 척도는 비일관적 반응을 나타내는 것으로, 특히 무작위 반응을 민감하게 포착한다. TRIN 척도는 묵인(Acquiescence)과 같은 고정된 반응을 탐지한다. 임상심리학자들은 비일관성이 무작위 또는 부주의 반응인지, 아니면 인지적 손상 이슈인지를 판단하기 위해 이 척도와 더불어 신경심리적 자료와 증상 타당도 척도들을 종합적으로 고려한다.

4) 이후에 살펴보겠지만, 여기서 사용된 지표는 사전 지표로 자기 보고, 가짜 문항, 지시 응답 문항, 응답 시간, 사후 지표로 롱스트링, 훔-짜 일관성, 재표집된 개인 신뢰도, 개인-전체 상관관계였다. 대체로 IER 탐지용 지표를 많이 사용할수록 IER에 해당하는 응답자의 비율은 높아진다(Jones et al., 2022).

을 때, 방법에 따라 약 0.5%~14%의 답변이 IER로 분류되는 것으로 나타났다. 구체적으로, 롱스트링은 16명을 IER로 탐지하여 가장 낮은 탐지 비율을 보였고(0.53%),<sup>5)</sup> 마할라노비스 거리 방법은 422명(13.93%)으로 가장 높은 비율을 보였다.

또한, 표본의 특성(예, 대학생, 성인), 동기나 보상 유형, 참가자들의 개인차 요인, 연구주제나 연구에서 사용되는 과제에 따라서도 다를 수 있다(Maniaci & Rogge, 2014). 예를 들어, Kurtz와 Parish(2001)는 대학생들을 대상으로 성격 5요인 검사(NEO-PI)를 실시하였을 때 IER 비율이 10% 정도라고 하였다. 그러나 Johnson(2005)의 연구에서는 전체 응답자에서 IER에 해당하는 사람의 비율은 3.5% 정도라고 하였는데, 이 연구에 참여한 학부생들은 자진하여 연구에 참여했기 때문에 일반적인 학부생들에 비해 동기 수준이 높았다고 볼 수 있다.

한편 Jones 등(2022)은 알코올 관련 연구를 위해 크라우드소싱된 자료를 사용한 연구들을 메타 분석하여 IER이 얼마나 나타나는지를 조사하였다. 95개의 연구 중 IER 탐지를 위한 측정치를 적어도 하나 이상 사용한 연구는 51개였으며(53%), 이들 중 48개 연구가 IER로 나타난 비율을 보고하였는데, 전체 표본 중 IER 응답자는 11.7%로 나타났다.<sup>6)</sup> 크라우드소

싱된 참가자들의 자료 질에 대한 우려가 커지고 있는 상황에서도 학계와 실무에서 이런 자료의 사용은 급격하게 늘고 있다. ‘크라우드노동자’라고 불리는 이들은 보상을 목적으로 조사에 참여할 뿐, 조사 자체에 대한 내적 동기나 흥미는 거의 없다고 봐야 할 것이다.

### IER이 발생할 가능성이 높은 조건

일반적인 조사나 검사에서는 다음과 같은 조건에서 IER이 발생할 가능성이 크다(Ward & Meade, 2023). 첫째, 설문 문항이 100문항 이상으로 많은 문항이 반복적으로 제시될 때이다. 둘째, 조사가 특정한 장소에서 직접 이루어지는 것이 아니라 감시가 없는 온라인으로 실시될 때이다. 셋째, 대부분의 크라우드소싱된 온라인 조사처럼, 설문 내용이 응답자와 별 관련이 없거나 응답자가 별 관심(흥미)이 없는 내용으로 지루함을 느끼는 경우이다. 넷째, 부주의하게 응답해도 분명한 부정적인 결과(예, 보상 철회나 처벌 등)가 없는 경우이다. 다섯째, 조사나 설문에 자발적으로 참여한 것이 아니고, 모든 설문 문항에 반드시 응답해야 할 경우이다.

### IER의 원인

Palaniappan과 Kum(2019)은 조사 연구에서 나타나는 응답 편향의 근본 원인을 두 범주로 나누었다. 하나는 성격 특성, 동기 수준, 인지 능력과 같은 참가자의 개인차 요인 범주이며, 다른 하나는 조사(설문)의 길이, 자료 수집 방식(mode), 주변 환경의 방해, 연구자-참여자 상호작용, 문화 간 차이와 같은 상황 요인 범

5) 여기서는 총 51개의 설문 문항 중 23개 이상을 연속으로 동일 응답한 경우를 IER로 간주하였는데, 이는 다소 엄격한 기준으로 보인다.

6) 이 메타 분석 연구에서, IER 탐지를 위해 가장 많이 사용된 측정치는 주의 체크(attention check) 문항이었으며, 크라우드소싱을 위해 사용된 플랫폼들의 비교에서는 다른 플랫폼들보다 MTurk에서 더 IER 응답자의 비율이 높은 것으로 나타났다.

주이다.

응답 편향의 일종인 IER의 원인도 크게 이 두 범주로 구분될 수 있으며, 좀 더 세부적으로 Meade와 Craig(2012)은 IER의 원인으로 1) 응답자의 개인차 변인과 설문에 대한 낮은 동기(motivation), 2) 환경의 주의분산 요인(environmental distraction), 3) 연구자와의 사회적 접촉(social contact) 부족, 4) 설문 길이(survey length)의 네 가지를 들었다. 첫 번째가 개인차 요인이라면, 나머지 세 가지는 상황 요인이다.

첫째, 개인차 요인으로, 어떤 조사나 설문에서든지 부주의하게 응답하는 경향을 보이는 사람이 있다. Bowling 등(2016)은 IER에 대한 개인차 변인의 영향을 보여주었는데, 참가자의 지인들이 보고한 성격 5요인 중에서 성실성, 원만성, 외향성, 정서적 안정성이 IER을 탐지하는 여러 지표 및 표준화된 평균 지표와 부적인 상관을 가졌다.<sup>7)</sup> 그리고 IER은 응답자들 간의 순위(rank order)에서 13개월 동안 상당한 안정성을 보였고, 학교에서의 학업 성취(성적과 결석)와도 연관이 있었다(즉 IER의 여러 지표가 높은 수준일수록 학점이 유의하게 낮고 결석이 많았다). 특히 신중성(carefulness)이라는 특성이 담긴 성실성은 IER과 개념적으로 중복되기에 높은 부적 상관관계를 보였다. 또한, 이타적 특성을 포함하는 원만성은 연구자를 도우려는 경향성으로 나타나기에 IER과 부적인 관계를 가진 것으로 해석될 수 있다. 따라서 연구자들은 IER을 어느 정도 안정성이 있는 개인차 변인이라고 보았다. 또한, 응답자의 교육 수준이 높으면 설문 내용의 이해도가

높고 설문조사에 비교적 더 친숙하므로, IER의 경향이 덜 나타난다. 이것은 설문 응답 관련 능력도 IER에 영향을 미칠 수 있음을 보여준다.

또한, 양질의 데이터를 얻는 데 있어 응답자의 적극적 참여나 몰입은 상당히 중요하기에 응답자의 동기 수준에 따라 IER이 다르게 나타난다(Schwarz, 1999; Tourangeau et al., 2000). 응답자의 목표가 솔직하고 신중한 답변을 하는 것이 아니라, 그저 빨리 설문을 마치는 것이라면 IER은 늘어난다. 예를 들어, 학점을 위해 참여하는 대학생처럼 자발성이 낮으면 설문 응답이 귀찮고 마지못해 참여하므로 IER의 가능성도 크다. 반면 채용이나 선발 장면이거나 개인적 피드백을 주려는 목적의 조사(feedback survey)라면 응답자가 설문의 목적과 내용에 진지하게 관심을 가지므로 IER은 거의 나타나지 않을 것이다. 여기서 핵심은 참여자의 동기이지만, 참여자의 상황(맥락)이 동기에 영향을 미치기에 상황 요인일 수도 있다.

특성(trait)이 아닌 상태(state) 측면에서의 개인차도 고려해볼 수 있다. Barber 등(2013)은 불면 등의 이유로 자기 통제력(self-control)이 고갈되면 무작위로 부주의하게 반응할 가능성이 크며,<sup>8)</sup> 조사에 들이는 에너지의 양이 둘 간의 관계를 매개함을 보여주었다.

둘째, 상황 요인으로, 환경의 주의분산 또는 방해가 있는 상황이다. 이런 상황에서는 성실한 응답을 하려는 응답자들마저도 주변의 다양한 자극들(타인의 존재, TV나 기타 매체로부터의 소음, 여러 활동의 동시 수행 등)로 인해 산만해지고 집중하지 못하게 된다. 주의분

7) 여기서 사용된 IER 탐지 지표들은 개인 신뢰도, 심리측정적 동의어, 반의어, 마할라노비스 거리 등이었다.

8) 동시에 불면이 사회적 바람직성 반응(SDR)을 낮추는 긍정적 효과도 나타났다.

산이 인지적 과제의 수행을 방해하는 효과를 고려할 때, 온라인 설문 참가자들이 처한 환경에서의 주의분산 요인은 분명한 IER 유발 요인이다.

셋째, 연구자와의 사회적 접촉 여부도 IER 과 관련이 있다. Johnson(2005)은 연구자와 응답자 간의 물리적 거리가 멀어질수록 성실하게 응답해야 한다는 책임감이 떨어지고, 바람직하지 않은 응답 패턴으로 이어질 가능성이 크다고 지적했다. 온라인 설문이 바로 연구자의 직접적인 통제를 벗어난 대표적 장면으로, 응답자와 설문 진행자 간의 직접적인 상호작용이 없고 지시문과 같은 간접적인 방식을 통해 메시지가 전달된다. 직접적 소통이 없기에 주의 깊고 솔직하게 응답해야 한다는 규범이 제대로 전달되지 않을 수 있다. Gibson과 Bowling(2020)은 실제로 온라인 설문 응답과 실험실에서의 응답 간에 IER에서 유의한 차이가 남을 보여주었다. 다만, 직접 상호작용 집단과 글이나 동영상상을 통해 소통한 집단 간의 유의한 차이가 나타나지 않는다는 실험 연구(Ward & Meade, 2018)도 있는 점을 고려할 때, 이 요인이 미치는 영향은 아직 명확하지 않다.

마지막으로, 설문 길이와 IER 간의 관련성이다. 설문이 길수록 응답자들에게 더 큰 노력이 요구되며 피로 수준은 높아진다. 300개 이상의 많은 문항에 응답하는 응답자들의 응답 패턴을 살펴본 결과, 설문의 중간이나 마지막 부분에서 무작위로 반응하는 경향성이 나타났다(Baer et al., 1997). 긴 설문의 경우, 막바지로 갈수록 응답자들이 빠르게 반응하면서 변이가 줄어들고 부주의하게 반응할 가능성이 커진다. 심지어 직무지원자와 같이 고부담(high-stake) 상황의 표본들에서도 이런 경향

성이 나타나는 것으로 밝혀졌다(Berry et al., 1992).

Gibson과 Bowling(2020)은 설문 길이가 IER에 미치는 영향을 두 개의 실험에서 조사하였다. 또한, 성실하게 응답하도록 하기 위한 보상 제공이나 처벌 경고의 효과 및 이것이 설문 길이와 IER 간 관계를 조절하는지도 살펴보았다. 결과에 의하면, 설문 길이는 일부 IER 지표에 유의한 정적 영향을 미쳤으며 보상이나 처벌도 일관된 주효과를 가졌으나, 모두 효과 크기는 크지 않았다(조절 효과도 나타나지 않았다). 이후 Bowling 등(2021)은 두 실험 연구를 통해 자기 보고형 설문에서 응답자들이 설문 응답을 지속해 나가면서 점차 부주의하게 응답하는 경향성이 일관되게 나타남을 보여주었다.

설문 길이와 관련하여 조사 말미에 위치한 척도나 구성개념의 경우 응답 자료의 질(예, IER의 정도)에서 차이가 나는지는 또 다른 중요한 질문이다. 이와 관련된 기존 연구들은 혼합된 결과를 보였다. 일부 연구는 자료의 질이 저하되는 결과를 보이지만(예, Bowling et al., 2021; Galesic & Bosnjak, 2009; Herzog & Bachman, 1981), 다른 연구에서는 자료의 질적 저하가 나타나지 않았다(예, Bäckström & Björklund, 2019; Gibson & Bowling, 2020, 연구 2). 이것은 긴 설문에 따른 피로로 IER이 나타나면서 자료의 질이 저하될 수도 있고, 아니면 긴 설문이라도 피로를 유발할 정도는 아니어서 마칠 때까지 계속 주의력을 유지할 수도 있음을 의미한다. 이런 맥락에서 Bowling 등(2022)은 정교한 설계로 긴 설문의 마지막 부분에도 주의 깊은 응답을 하는지를 조사하였다. 긴 설문 문항을 포함하는 두 개의 실험(연구 1은 500 문항 이상, 연구 2는 300 문항 이



상)에서 응답자들은 특정 성격 척도 문항들이 설문문의 첫 부분에 제시되는 조건과 마지막 부분에 제시되는 조건에 무선 할당되었다. 그리고 준거로 사용하고자 지인들로부터 응답자의 성격, 태도, 및 행동에 대한 자료를 수집하였다. 두 실험 모두에서 두 조건 간 성격 척도의 구성 타당도(수렴 및 준거 관련 타당도)에서 차이가 나타나지 않았다. 이는 아주 긴 설문문이 IER 및 타당도에 미치는 영향은 그리 크지 않을 수도 있음을 시사한다.

### IER의 영향

IER을 대수롭지 않게 여기는 연구자들은 IER 자체가 잘 나타나지 않는 현상이라 자료의 질에 심각한 문제를 초래하지 않는다고도 한다.<sup>9)</sup> 그러나, 앞서 보았듯이 다양한 비율로 추정되는 IER 자료가 분석 결과에 미치는 영향도 다양하게 연구가 이루어졌다.

IER의 가장 일반적 유형 중의 하나는 무작위 반응에서 나타나는 일관성 부족인데, 이로써 측정치에 무선 오차(random error)가 발생하게 된다. 고전적 측정이론에서 측정치는 진점수와 무선 오차의 합이며, 무선 오차가 증가하면 신뢰도는 저하된다. 표본의 5~10% 정도가 불성실 응답자인 경우, 변인의 심리측정 속성이 상당히 달라지며 타당도에도 크게 영향을 미치는 것으로 나타났다(Credé, 2010;

Woods 2006). 또한, 최근 사회 및 심리과학 연구에서 크게 문제가 되는 실험 및 조사 연구에서의 재현성(replicability) 위기도 소규모 표본 및 IER이 원인인 것으로 주목받고 있다(Palaniappan & Kum, 2019). Oppenheimer 등(2009)은 이전에 보고되었던 실험 결과들이 성실한 응답자로 분류된 참가자들 자료에서는 반복검증되었으나, 불성실 응답자로 분류된 참가자들 자료에서는 반복검증이 되지 않음을 보여주었다.

IER은 여러 방향으로 연구나 조사결과를 왜곡할 수 있다. 예를 들어, IER이 구성개념의 평균을 변화시킬 수 있으며, 만약 특정 표본의 성실 응답자 평균이 척도 중간점보다 높다면, IER은 전체 평균을 낮출 것이며, 평균이 척도 중간점보다 낮다면 전체 평균을 높일 것이다. 또한, IER 자료가 포함되면 구성개념 간의 상관관계를 확대하거나 축소할 수 있으며, 예상과는 다른 요인구조를 만들어낼 수 있다(Huang et al., 2015). 리더십 설문 자료를 사용한 Goldammer 등(2020)은 IER이 구성개념의 심리측정적 속성에 미치는 영향을 조사하였다. 연구 표본에서 33%가 IER로 나타났는데, IER은 문항 변량을 부풀렸으며, 문항 평균이 중간으로 편향되게 하였으며, 구성개념 측정치(indicator)들의 잔차 변량을 증가시키며, 리더십과 같은 합의 기반 구성개념의 집단 내 일치도를 감소시켰다. 다음부터 IER이 분석 결과에 미치는 영향을 좀 더 자세히 살펴보았다.

우선 조사에 포함된 변인(구성개념) 간의 상관관계가 왜곡되는 결과를 초래한다. 정의상 무선 오차는 다른 변인들과는 아무런 상관관계가 없기에, 무선 오차가 늘면 변인 간 상관은 줄 것이다. Credé(2010)는 무작위 응답 비율을 전체 표본의 5% 정도로 가정하여 시

9) 예를 들어, 박원우 등(2020)이 IER 자료의 제거가 통계적 추론 결과에 미치는 영향을 확인한 결과, IER의 존재가 상관관계, 요인 분석, 회귀분석 결과에 미치는 영향은 기존 문헌의 결과와 유사한 수준으로, 연구 가설의 결과를 왜곡시킬 정도로 크지 않은 것으로 나타났다.

플레이션 분석을 하였는데, 상관계수가 과대 혹은 과소 추정되는 결과를 보여주었다. 그는 이것이 범위 제한이 있을 때, 신뢰도가 낮을 때, 연속변수를 임의로 이분변수로 바꾸어 분석할 때 나타나는 효과에 비견될 수 있다고 하였다. Kam(2019)은 원 표본보다 IER이 제거된 표본에서 자기 및 동료 보고 간 상관관계가 더 높게 나타남을 보여주었다.

또한, 특정 측정의 구성 타당도를 확립하는 것은 요인 분석과 같은 여러 방법으로 가능하지만, 가장 전형적인 방법은 특정 구성개념이 다른 구성개념들과 어떤 관계를 갖는지를 확인하는 수렴 및 변별타당도이다. 이 경우에도 IER에 의한 무선 오차 때문에 변인 간 관계가 왜곡되기에 구성 타당도 확립에도 부정적 영향을 미칠 수 있다(Kam, 2019; Kam & Meyer, 2015). 또한, McGrath 등(2010)도 응답자들의 비밀관적인 응답이 준거 관련 타당도에 부정적인 영향을 미침을 보여주었다.

둘째, IER로 인해 변인 간 상관이 축소되면(때로는 부풀려지면), 요인 분석 결과에도 영향을 미칠 수 있다. 문항 간 상관에 기초하여 척도를 개발할 때, IER은 오차 변량을 늘려 요인 부하량이 감소하고 내적 일관성을 낮추거나 해석 불가능한 문항 간 상관 패턴을 초래함으로써 결국 적절한 요인구조가 도출되지 못하는 결과에 이를 수 있다(Arias et al., 2020; Huang et al., 2015; Johnson, 2005; Woods, 2006). Kam과 Meyer(2015)는 직무만족과 불만족을 측정하는 문항들은 1요인 해를 만들어내지만, 높은 IER을 보인 자료와 묵인(acquiescence) 정도가 높은 자료를 포함하면 2요인 해가 나타남을 보여주었다. Kam(2019)과 Huang 등(2015)은 표본에서 IER 응답자가 제거되었을 때 요인 분석에서 정확한 요인 수가

나타남을 보여주었다. 더불어 확인적 요인 분석의 모형 적합도도 더 잘 나왔다(Arias et al., 2020).

또한, 역점수 문항에 대한 불성실한 응답은 방법 요인(method factor)의 생성으로 이어져 일반(긍정문) 문항과 역점수(부정문) 문항이 각자 다른 요인에 적재되는 결과를 초래한다(Woods, 2006). 즉, Woods(2006)의 긍정적 및 부정적 설문 문항들을 포함하는 시뮬레이션 연구에서 10% 이상의 설문 응답이 무작위로 이루어지면 예상된 1요인 해는 적합도가 좋지 않았고, 방법 요인이 추가로 나타났다.

Arias 등(2020)은 두 표본에서 IER이 네 개의 성격 척도들(외향성, 성실성, 정서적 안정성, 기질적 낙관성)의 차원, 내적 구조, 신뢰도에 미치는 영향을 분석하고자 하였으며, 의미상 양극인 문항 쌍들에 대한 응답의 비밀관성으로 IER을 포착하기 위해 요인 혼합 모형(FMM: factor mixture model)을 사용하였다. FMM은 표본 및 척도에 따라 4.4%~10%의 IER을 포착하였다. IER을 포함한 전체 표본에서 단일 차원의 구성개념(예, 성실성)에 대한 요인 분석을 실시한 결과, 1-요인 모형은 수용하기 어려운 적합도를 보였고(예, RMSEA = .108, CFI = .84), 표현(wording) 요인이 추가된 쌍 요인 모형의 적합도가 수용할 만한 수준이었다(RMSEA = .053; CFI = .97, ECV = .73).<sup>10)</sup>

10) 내향-외향처럼 심리학에서는 양극 차원을 가진 구성개념들이 많다. 이런 양극성은 균형 척도(예, '전혀 그렇지 않다~매우 그렇다'의 5점 척도)를 가진 검사에서는 긍정 및 부정 문항으로 구성된다. 그런데 균형 척도는 요인 분석에서 표현(wording) 요인(예, 부정적인 표현의 문항들로 구성된 요인)을 만들어내서(Podsakoff et al., 2003), 자료가 이론적 모형과 잘 맞지 않는(적합도가 낮은) 결과에 이르기 쉽다.

반면, IER이 제거된 표본에서는 1-요인 모형의 적합도가 양호한 수준이었고(RMSEA = .057, CFI = .96), 쌍 요인 모형에서 표현 요인의 설명 변량은 매우 낮아져 사실상 단일 차원임을 지지하였다(ECV = .94). 또한, IER이 제거된 표본에서의 특성 추정치들은 전체 표본에서보다 4.5%~11.8%까지 더 정확하였다. 이런 결과는 IER 자료가 이론적 모형의 적합도를 낮추며, 허위 변량(spurious variance)을 만들어내며, 차원과 내적 구조가 분명하게 드러나지 않게 하고, 특성 추정치들의 신뢰도를 낮춘다는 것을 보여준다.

셋째, 지금까지의 논의에서 이미 짐작할 수 있듯이, 가설 검정에도 IER이 영향을 미칠 수 있다. 불성실 응답은 허위의 집단 내 변산성을 부풀리며, 신뢰도를 낮추고, 상관관계에 영향을 미침으로써, 2중 오류를 발생시킬 가능성이 커진다(Clark et al., 2003; Liu et al., 2013). Maniaci와 Rogge(2014)는 표본에서 IER 응답자를 제거하는 것이 가설 검정에 미치는 효과를 조사하는 일련의 연구를 수행하였다. 그들의 시뮬레이션 연구에 따르면, 표본에서 IER 비율이 전혀 없을 때 보다 5%일 때 검정력(power)이 약 10% 포인트 낮아졌으며, IER이 10%인 경우에는 검정력이 18% 포인트까지 낮아졌다. 반대로, 개념적으로 구분되는 구성개념 간의 관계는 과다 추정되어 1중 오류를 유발할 수 있다(Huang et al., 2015). 또한, 원 데이터 대비 IER을 제거한 데이터로 회귀 분석하였을 때 검정력이 상당히 증가하는 것으로 나타났다. 즉 회귀분석처럼 상관관계를 포함하는 분석에서의 가설 검정은 IER 자료를 제거하는 것이 매우 효과적이다.

그러나 가설 검정의 통계적 검정력은 표본의 크기에도 달려있다. 극단적인 IER을 제거

하면 오차의 근원을 제거하는 것이기에 검정력이 높아질 수도 있지만, 표본 수 감소로 저하될 수도 있기에 ‘약간 부주의한 응답자들마저도 제거’하는 것이 어떤 효과를 보일지는 분명하지 않다. 이런 이유로, 가설 검정의 경우, 가장 부주의한 반응을 보인 표본만을 제한적으로 제거하는 것이 유리할 수 있다. 그래서 특정 응답자의 불성실을 확인하는 데 있어 서로 다른 IER 지표들이 일치하지 않을 때, 때로는 표본 크기와 검정력을 유지 또는 늘리기 위해 일부 응답자는 그냥 표본에 두는 것이 나올 수 있다. 이는 IER 탐지에 있어 단일 지표에만 의존하는 것은 추천하지 않는다는 것을 의미한다. 그런 상황에서는 IER 자료를 넣거나 빼고 분석하면서 결과에 차이가 있는지를 살펴볼 필요도 있을 것이다(예, Wang & Hau, 2024).

넷째, 지금까지의 논의는 무선 오차가 비체계적인 변량을 만들어내면서 상관을 축소하고 2중 오류의 위험을 높이는 효과에 초점을 두었지만, 일정한 조건에서는 IER이 상관을 부풀리게 하는 체계적 오차(systematic error)도 초래할 수 있다. Huang 등(2015)은 IER이 어떤 경우에는 평균의 편향된 추정을 유발할 수 있어 변인 간의 허위 상관이 나타날 수 있음을 보여주었다. 이것은 대부분의 성실한 응답자들이 두 구성개념에 있어 평균이 척도 중간점보다 높거나 낮은 경우에 특히 그러하다. 예를 들어, 만약 성실한 응답자들이 두 구성개념에서 평균적으로 척도 중간점보다 더 높은 수준이고 부주의한 응답자들이 중간점 근처 혹은 그보다 더 아래라면, IER 자료는 두 구성개념 간의 상관관계를 부풀리는 이상치(outlier)로 작용할 수 있으며, 그로 인해 1중 오류를 유발할 수 있다. 그래서 특정 연구에

서 IER이 어떤 영향을 미칠지의 사전 예측이 쉽지 않다.

마지막으로, DeSimone 등(2018)은 IER을 연속되는 문항에 동일 응답을 계속하는 직선 응답(SL: straightlining)과 무작위 응답(RR: random responding)의 두 가지 형태로 구분하고, 시뮬레이션 방법을 통해 두 형태가 자료에 서로 다른 영향을 미치는지를 조사하였다. 구체적으로 그들은 세 연구에서 극단적인 SL 및 RR 조건, RR과 SL이 부분적으로 포함된 조건<sup>11)</sup>이 문항 간 상관, 알파 계수, 주성분(principal component) 구조와 같은 척도 특성에 미치는 영향을 조사하였다. 또한, 다양한 IER 반응 분포(예: Normal, Inverted-Normal, V-shaped 등)가 이러한 통계치에 미치는 영향도 조사하였다. 분석 결과, RR과 SL은 통계적 결과에 서로 다른 영향을 미쳤다. 즉, 문항 간 상관, 알파 계수, 첫 번째 성분의 고유치를 SL은 증가시키지만, RR은 감소시키는 경향이 나타났다. RR의 존재는 문항 간 상관 크기를 감소시키면서 낮은 알파 값과 평평한 고유치 분포를 갖는 주성분 분석 결과를 초래한다. 반면 SL의 존재는 문항 간 상관이 1에 가까워지도록 하며, 높은 알파 값과 편포된 고유치 분포를 갖는 주성분 분석 결과를 초래한다. 여기서 주목할 점으로, 변인 간의 상관관계가 높은 수준일 때는 RR이 SL보다 더 크게 영향을 미치겠지만, 사회과학의 현장 연구들에서는 높은 상관관계(예, .50 이상)를 갖는 경우가 드문 점을 감안할 때(Bosco et al., 2015), SL이 RR보다 더 우려되는 IER이라고 볼 수 있다.<sup>12)</sup> 무엇보다

통계분석을 하기 전에 IER의 존재뿐만 아니라 IER의 두 유형을 구분하는 것이 중요하다.

요약하면, IER에 의해서 구성개념의 평균, 신뢰도, 요인 분석, 구성개념 간의 상관관계가 달라질 수 있다. 여러 연구 결과들을 바탕으로 보았을 때, IER은 데이터의 분석 과정에서 오류를 발생시킴으로써 분석 결과에 대한 확신을 갖기 어렵게 한다. 그리고 IER로 유발되는 문제점들은 설문에서 얻어진 수치를 바탕으로 정책개발, 종업원 선발, 마케팅 계획 등과 같은 고부담(high-stake)의 의사결정을 내려야 할 때 심각히 우려되는 부분이다. 따라서 IER을 사전에 방지하거나 사후에 탐지함으로써 자료의 질을 높이는 것은 자료를 분석한 결과에 대한 타당도 및 양질의 의사결정을 내리기 위해 매우 중요한 작업이다.

## IER의 탐지

### IER의 형태

조사나 검사 문항에 대한 응답에서 IER 반응은 다음과 같은 세 가지 형태로 나타난다. 첫째, 불변성(invariability)이다. IER의 두 가지 구분되는 유형 중의 하나인 ‘직선 응답(SL)’에 해당하는 불변성 또는 롱스트링(longstring)이다. 이 유형은 연속적으로 동일한 반응(또는 패턴)을 보이는 것이다(예, 1, 1, 1, 1, 1, 1 또는 1, 2, 1, 2, 1, 2, 1, 2). 컴퓨터 기반의 온라인 조사나 검사에서는 빠르게 동일 응답을 계

11) 다섯 가지 조건이 있었는데, RR과 SL의 극단인 RR(100%)/SL(0%)와 RR(0%)/SL(100%)의 두 조건과 더불어 RR(25%)/SL(75%) 조건, RR(50%)/SL(50%), LL(75%)/SL(25%)으로 구분되었다.

12) 달리 말하면, 평균 문항 간 상관이 .50보다 낮은 수준인 경우, 일관되게 동일 응답을 하는 SL이 RR보다 상관관계의 크기에 더 영향을 미칠 것이다.

속함으로써 조기에 설문을 종료할 수 있다. 이런 불변의 톤스트링은 육안으로도 쉽게 탐지가 된다. 그래서 IER이 탐지되는 것을 우려하는 응답자(대개 설문 응답으로 보상을 받은 응답자)는 이렇게 응답하지 않는다.

둘째, 비일관성이다. 이론적 또는 논리적으로 추론할 때 나올 수 있는(예상되는) 응답 패턴과 일치하지 않는 패턴이 나타나는 경우이다. 두 문항이 거의 동일하거나 매우 높은 상관을 가진다면, 두 문항에 대한 응답에서도 이와 유사해야 할 것이다. 비일관적으로 반응하는 이들은 이런 예상에 부합하지 않는 사람들이다. 또한, 설문의 시작부터 끝까지 무작위로 반응하기보다는 후반부로 갈수록 주의력이 저하되는 것이 일반적이다. 예를 들어, 긴 설문문에 대해 처음에는 주의를 기울여 솔직하게 응답하지만, 조금 지나면 피곤해지고 흥미를 점차 잃으면서 주의가 산만해지면서 마지막에는 다소 부주의하게 응답하면서 끝낸다.

셋째, 너무 빠른 응답이다. 지나치게 빠른 응답이란 문항을 읽고 이해하고 정확하게 응답하기에는 거의 불가능한 시간에 응답하는 것을 말한다. 그래서 최소 시간 역치를 설정하여 제대로 응답하지 않은 사람을 확인할 수 있다. 조사에서 설문 응답의 경과 시간을 측정하면 과도하게 빠른 응답은 쉽게 탐지가 된다. 그러나 어떤 경우에는 이를 제대로 탐지하지 못할 수도 있다. 예를 들어, 응답 과정에서 주의 집중하기 어려운 조건이거나 다른 과제를 하면서 응답하는 경우, 실제 응답은 빠르지만 전체 응답 시간은 길 수 있다.

#### IER 탐지를 위한 사전(a priori) 조치

IER은 사전 예방이 최상이므로, 자료수집

이전에 할 수 있는 조치들에 대해 살펴보았다. 이들은 대부분 IER을 탐지할 수 있는 문항을 삽입하는 방식이다.

1) 사용 여부 자기 보고 문항(Use Me item). 이것은 설문 말미에 (참여에 대한 보상은 주어질 것이라고 말하면서) 설문 과정에서 얼마나 주의를 기울이고 몰입했는지를 직접 물어보는 방식이다. 대개 설문 마지막에 “귀하의 데이터를 사용해야 하나요?” 또는 좀 더 세부적으로 “마지막으로, 귀하께서는 우리가 귀하의 응답 내용을 믿고 연구를 진행해도 된다고 생각하십니까?”라고 직접 물어보거나, “나는 간혹 문항을 제대로 읽지 않고 응답하였다”와 같은 문항으로 자신의 응답이 성실했는지의 자기 보고를 요청하는 것이다(Meade & Craig, 2012).<sup>13)</sup> 이 방식은 이 문항에 대해 솔직하게 답하지 않을 수 있다는 단점이 있어 크게 추천하지 않는 연구자들도 있다(특히 돈을 받고 설문 응답하는 경우, Goldammer et al., 2020).

2) 지시 응답 문항(instructed response items). 이것은 IER을 탐지하는 가장 직접적인 방법으로, “이 문항에는 답하지 마십시오”, “자료의 질을 확인하기 위해, 이 문항에는 ‘매우 그렇지 않다’에 응답하십시오”, 또는 “문항 내용과 상관없이 이번 문항에는 4에 표시하십시오” 처

13) Meade와 Craig의 연구에서 사용한 지시문 예는 다음과 같다. “마지막으로 우리 연구에서는 주의를 집중하여 응답한 자료만 사용하는 것이 매우 중요합니다. 그렇지 않으면 여러 해 동안의 노력이 허사가 됩니다. 귀하는 어떤 경우에도 연구 참여에 대해 추가 점수를 받을 것이기에, 본 조사에 어느 정도 노력을 기울였는지를 솔직히 답해 주십시오.” “나는 이 조사(연구)에 \_\_\_\_\_.”  
1: 전혀 노력을 기울이지 않았다, 2: 별로 노력을 기울이지 않았다. 3: 약간 노력을 기울였다, 4: 제법 노력을 기울였다, 5: 많은 노력을 기울였다.

럼 구체적인 응답을 요구하는 것이다. 지시한 대로 응답하지 않으면 주의를 기울이지 않은 것이 분명하다. 다만, 이런 지시 응답 문항은 설문에서 쉽게 눈에 띄기에 영리한 응답자는 이를 바로 알아차리고 대응할 수 있다는 단점이 있다. 혹자는 지시 응답 문항에 많은 응답자가 걸려들기 때문에 비판적인 시각을 보이기도 하지만(예, Curran & Hauser, 2019), 지시 응답 문항의 내적 일관성이 높고, 다른 IER 지표들과의 수렴 타당도 및 목인 편향과의 부적 타당도 계수 등이 나타나기에 사용이 권장된다(Kam & Chan, 2018).

3) 가짜 문항(bogus items). 가짜 문항은 일반적인 상식에서 벗어나고 거의 있을 것 같지 않은 불가능한 일에 대한 진술문이다. 예를 들어, “내 친구는 모두 외계인이다” 또는 “나는 한 번도 컴퓨터를 사용해본 적이 없다”가 있다. 성실한 응답자라면, 이에 대해 당연히 “동의하지 않는다” 혹은 “그렇지 않다”라고 할 것이다. 또한, “나는 위성인 지구에서 태어났다”에 ‘아니오’라고 답한다면 이것도 IER의 신호가 된다.

가짜 문항은 지시 응답 문항보다 덜 눈에 띄면서 다른 문항들과 잘 섞일 수 있다. 크라우드소싱 웹사이트에서 보상을 위해 조사에 참여하는 사람들은 지시 응답 문항을 쉽게 포착하고 주의를 기울이지만, 가짜 문항은 눈에 잘 띄지 않고 컴퓨터로 자동 인식해내기도 어렵다. 다만, 가짜 문항은 성실한 응답자에게는 주제와 무관한 엉뚱한 질문으로 지각되어 불쾌하거나 시간 낭비라고 생각할 수 있다는 단점이 있다. 또한, ‘약간 동의하지 않는다’에 응답하는 경우에는 어떻게 처리할지와 같이 모호성도 있다. Curran과 Hauser(2019)가 가짜 문항 응답 시의 생각을 말로 얘기하도록 하는

방법(think-aloud method)을 적용하였을 때, 간혹 매우 성실한 응답자들도 가짜 문항에 동의하곤 하였다. 따라서 성실한 응답자를 부주의한 응답자로 잘못 지목할 수도 있는 것이다.<sup>14)</sup>

응답자의 불편을 초래하지 않기 위해 Meade와 Craig(2012)은 가짜 문항이나 지시 응답 문항이 최대 3개 이상을 넘지 않는 것이 좋다고 하였고, 조사나 검사의 길이에 따라 달라지겠지만, 50~100문항마다 하나씩 넣을 것을 추천하였다. 그리고 Arthur 등(2021)은 IER 응답자로 결정하기 위한 객관적인 기준은 아직 없지만, 연구자가 조사 이전에 이 기준을 설정해두어야 한다고 주장하였다.

4) 응답 시간. 이것은 응답에 걸린 총 시간을 고려하여 조사 이후에 불성실 응답자를 탐지하는 것으로 추가 문항 삽입이 없다. 앞서 설명한 IER의 형태 중 너무 빠른 응답에 근거한 것으로, 불성실 응답자들은 개별 문항에 주의를 기울이는 시간이 짧을 것이고, 결국 설문 응답에 걸리는 시간이 짧을 것이다. 응답 시간이 컴퓨터에 항상 기록, 저장되는 것은 아니기에 사전에 준비해야 한다는 점에서 사전 방법일 수 있다.

응답 시간의 경우, IER로 탐지되는 것을 피하기 어렵다. 다만, 전체 설문 응답 시간보다

14) “1년에 14개월 일 한다” 또는 “나는 매달 요정으로 부터 임금을 받고 있다”와 같은 문항에 대해 성실하게 응답하는 사람이라면 ‘전혀 그렇지 않다’라고 답해야 하지만, 그렇지 않은 경우가 나타난다. Curran과 Hauser는 이런 문항에 답하는 과정에서의 생각을 소리 내어 말하는 절차를 통해, 사람들은 자주 타당하지 않은 응답을 정당화하는 경우가 나타나며, 서로 다른 문항들에 대해 이런 행동에서의 개인 간 상당한 변이가 있음을 보여주었다.

는 설문 페이지 단위로 응답 시간을 측정하는 것이 추천된다. 중간에 잠시 쉬는 경우 응답자가 빠르게 응답하더라도 전체 응답 시간은 길게 나타나기 때문이다. 설문 응답 시간에는 상당한 개인차가 존재하지만, Huang 등(2012)은 전체 설문 시간을 계산했을 때, 문항당 평균 2초 이하로 응답한 응답자들을 IER로 간주할 수 있다고 하였다.<sup>15)</sup> Bowling 등(2016)의 연구에서는 문항당 2초 기준을 사용할 때 전체 시간보다 페이지 시간이 더 나은 결과를 보였다. 후속 연구(Bowling et al., 2023)에서도 페이지 시간 지표(page time index)가 IER 탐지에 상당한 이점이 있음을 보여주었다. 이 지표는 문항당 2초보다 더 빨리 응답한 설문의 페이지 수를 계산하는 방식이다. 세 연구에서 페이지 시간 지표의 구성 타당도를 확인하였는데, 다른 IER 지표들과의 수렴 타당도를 보였으며, 주의 깊게 응답하도록 경고가 있는 실험조건과 통제조건 간의 차이가 났으며, 응답자가 문항 내용을 인지하지 못하는 정도를 예측하였다. 더불어, 전체 시간보다 더 나은 타당도를 보였고, 문항당 2초 규칙은 다양한 단어 길이의 문항들에 대해서도 타당한 결과를 보였다.

#### IER의 사후(post hoc) 탐지 방법

사후적으로 IER을 탐지하는 방법은 매우 다양하며, 이 또한 각기 장단점이 있다. 다음에 소개되는 사후 IER 탐지 지표들은 롱스트링처럼 같은 응답을 반복적으로 하거나 동의어/반의어를 사용하는 것과 같이 수집된 자료에서

15) Goldammer 등(2020)은 이 기준이 너무 엄격하기에 연구마다 수집된 자료로부터 경험적으로 기준을 설정할 것을 제안하였다.

응답 패턴을 분석하는 경우나 통계적인 분석을 통해 도출된 수치를 확인함으로써 IER을 탐지한다.

1) 불변성 지표. 여기에는 롱스트링과 개인 응답 변량(표준편차)를 보는 방법이 있다. 앞서 보았듯이 불변성의 대표적 지표는 롱스트링인데, 이것은 여러 구성개념을 측정하는 문항들로 구성된 설문에서 동일한 선택지(척도점)에 응답을 계속 반복하는 것으로, 동일한 응답이 연속된 최대 수가 롱스트링 지수이다. 예를 들어, A라는 사람이 5점 척도로 구성된 10개의 문항에 대해 1111444333으로 응답했을 때, A의 롱스트링 지수는 '4'가 된다. 이것은 전체 설문에서 가장 긴 동일 응답 수나 설문 페이지당 롱스트링 지수의 평균치 등으로 계산될 수 있다. 이것은 앞서 언급된 '직선 응답(straightlining)'을 탐지하는 데 유용하며, 설문지에 하나 이상의 구성개념과 역점수(reverse worded) 문항이 포함된 경우에 특히 추천된다(Curran, 2016).

또 다른 불변성 지표인 개인 응답 변량(IRV: individual response variance)은 응답자가 지나치게 일관되게 반응하는지의 정도를 조사하기 위해 전체 응답에서의 표준편차를 계산하는 방법이다(Marjanovic et al., 2015). 롱스트링은 동일 반응이 연속되어 나타나는 경우만을 탐지하는 반면, 이 지표는 패턴에 기반한 불변성의 탐지도 가능하다. 즉, 동일 연속 반응뿐만 아니라 패턴화된 응답을 탐지하는데 뛰어나다(Dunn et al., 2018). 예를 들어, 짧은 동일 연속 반응(예, 4, 4, 4, 5, 5, 5, 4, 4, 4)이나 변갈아 가면서 동일 연속 반응(예, 3, 4, 3, 4, 3, 4, 3, 4)을 하는 경우도 포착할 수 있다. 변량과 표준편차는 대부분의 통계 패키지에서 쉽게 계산되는 것도 장점이다. 두 지표(특히

개인 응답 변량은 여러 구성개념이 포함된 조사에서 특히 유용할 수 있다. 만약 구성개념 간 응답 변량이 낮은 수준을 보이면, 이는 구분되는 구성개념 간 응답이 유사하다는 것으로 IER일 가능성이 크다. 다만, 리더십 설문 자료를 사용한 Goldammer 등(2020)은 롱스트링 지표와 개인 응답 변량이 IER 탐지에 그리 효과적이지 않으며, 균형 척도가 아닌 일방향(unidirectional) 척도<sup>16)</sup>가 사용되었을 때에는 사용하지 말 것을 제안하기도 하였다.

2) 마할라노비스 거리(Mahalanobis distance). 이상치(outlier)를 분석하는 통계적 방법으로, 특정 응답자의 반응 패턴이 전체 자료에서 일반적으로 나타나는 응답 패턴(normative response pattern)과 상이한 정도를 측정한다. 즉, 마할라노비스 거리는 동일 설문 문항들에 대한 특정 사람의 응답과 대다수 다른 사람들의 응답 평균 간의 다변량 거리로 정의된다. 높은 마할라노비스 거리를 보이는 응답자는 전체 표본과는 다른 반응 패턴을 보이기 때문에 다변량 이상치(multivariate outlier)이자 IER로 간주될 수 있다. IER 탐지에 있어 상대적으로 덜 사용되는 마할라노비스 거리는 대부분 통계 패키지에서 손쉽게 계산되며, 여러 유형의 IER을 탐지하는데 뛰어나다. 예를 들어, 대부분 응답자가 척도의 어느 한쪽 끝에 응답하는데, 누군가가 척도 중간점에 주로 응답한다면 이는 이상치가 된다. 그리고 다른 지표들과 달리 설문의 모든 문항에 대한 응답

을 사용한다는 점에서 가용한 자료를 더 잘 이용하는 셈이다. 다만, 이것은 자료가 정규분포되어 있다는 가정을 요구하는데 이 가정이 충족되지 않는 경우가 종종 있다. 자료의 비정규성은 IER 탐지 효과성을 낮추므로 이 방법의 사용에 있어 제약 요소이다.

3) 일관성(consistency) 지표. IER 지표로 가장 널리 사용되는 것으로, 의미가 동일한 문장이나 반대인 문장을 두 군데에 배치하여 일치 정도를 보는 방법이다. 예를 들어, 의미상 거의 동일한 두 문항(예: “나는 매우 열정적인 사람이다”, “나는 매사 열정이 넘치는 편이다”)을 설문의 전반부와 후반부에 각각 배치하고, 두 문항에 유사하게 응답할수록 성실한 응답으로 간주할 수 있다. 이런 일관성 지표들은 한 명의 응답에 근거한 것이기에 개인 내 상관관계를 산출한다. 예를 들어, 이 지표를 만들려면, 문항 내용의 유사성에 근거하여 여러 문항 쌍들을 만들어내야 한다. 만약 10개 문항 쌍이 만들어졌다면, 쌍의 한 문항은 A군, 다른 문항은 B군에 할당하면 두 군에 10문항씩 있게 된다. 그리고 두 군 간의 상관관계를 계산할 수 있다.

일관성 지표는 두 가지 유형으로 구분된다. 첫째, 의미적 동의어/반의어(semantic synonyms/antonyms) 지표이다. 이것은 이론, 논리, 또는 의미에 기반하여 문항들을 쌍으로 묶어 유사한 응답을 하는지를 보는 것이다. 만약 기존의 잘 타당화된 척도가 특정 구성개념을 측정하기 위해 사용된다면, 그 척도의 문항들에 대한 응답은 매우 높은 상관관계를 가질 것으로 기대할 수 있다. 그래서 동일 특성을 측정하는 문항 쌍들에 대한 반응이 유사한 정도, 즉 일관성을 볼 수 있는 개인 내 상관관계를 계산한다. 의미적 반의어는 반대되

16) 성격검사 등에서는 주로 중간점(midpoint)이 있는 균형 척도(balanced scale)가 사용되지만, 리더십과 같은 구성개념을 측정할 때에는 중간점 없이 특정 방향으로의 정도를 표시하는 척도를 사용하는 경우가 많다(예, 1=전혀, 2=아주 간혹, 3=가끔, 4=자주, 5=항상).



는 응답이 기대되는 문항 쌍에 근거하며, 특정 문항을 역점수화하여 상관을 낸다. 다만, 반대 의미 문장들의 경우는 다소 복잡하다. 즉, 반대 의미의 문장들에 대해 똑같이 ‘예’ 또는 ‘아니오’라고 답하는 경우, 이것이 변증법적 사고를 나타내는 것일 수도 있고, 맥락이나 상황에서의 차이로 똑같이 답할 수도 있다(예를 들어, 옷을 사는 데에는 매우 검약하지만, 여행하는 데에는 아끼지 않을 수 있다).

둘째, 심리측정적 동의어/반의어(psychometric synonyms/antonyms) 지표이다. 이것은 전체 표본에서 나타나는 문항 간 혹은 변인 간 상관관계의 패턴이 개인 내에서도 일관되게 나타나는지를 살펴보는 것으로, 전체와 일관되게 나타날수록 성실한 응답으로 간주된다. 이론이나 의미에 기반하여 의미적 동의어/반의어를 구성하는 것과 달리, 이것은 관찰된 표본 상관관계에 근거하여 문항 쌍을 구성한다는 점에서 심리측정적 접근이다. 전자는 문항 내용에 기반하지만, 후자는 특정 표본에서 문항 간의 상관관계를 분석함으로써 가능하다. 여기서 일반적 기준은 적어도 .60의 상관을 보이는 문항 쌍을 동의어로 간주하는 것이다(Meade & Craig, 2012).<sup>17)</sup> 문항 쌍들에서의 개인 내 상관관계가 높으면 주의 깊게 응답했음을 나타내며, 이 지표의 장점은 겉으로는 서로 다른 것으로 보이는 문항들도 간혹 매우 높게 상관될 수 있으며, 그런 문항 쌍들을 많이 확보함으로써 지표의 신뢰도를 높일 수 있다는 점이다. 그리고 계산의 전 과정을 컴퓨터 코드로 만들면 추가 노력 없이 자동으로

계산할 수 있는 것도 장점이다. 다만, 이 방법에서 유의할 점은 표본의 다수가 성실한 응답자들로 구성되어 있어야 한다는 조건이다(반면, 의미적 동의어/반의어의 경우에는 대부분 응답자가 부주의하더라도 사용할 수 있다).

4) 홀-짝 일관성(even-odd consistency) 지표와 개인적 신뢰도. 홀-짝 일관성 지표는 조사에 포함된 기존의 확립된 척도 문항들을 두 하위 척도로 나누고(홀짝으로 구분하여 두 하위 척도로 구분하여) 두 하위척도 간의 개인 내 상관을 계산하는 것이다. 이것은 확립된 척도의 문항들은 경험적 타당화 과정을 거치므로 문항 간 상관이 높을 것이며, 한 개인의 응답에서도 상당히 높은 상관을 보여야 한다는 가정에 기반한다.

일반적으로 동의어, 반의어, 홀-짝 일관성 간에는 높은 수렴 타당도를 보이므로, 이것들은 유사한 IER 행동을 포착하며 셋 중 하나만 사용해도 무방하다(Kam & Chan, 2018). 또한, 이 세 가지 IER 탐지 방법은 잘 확립된 척도를 사용하는 연구 목적의 조사에서 추천된다(Arthur et al., 2021).

앞서 다룬 지표들에서는 문항이나 하위척도의 쌍을 이루어 상관계수를 계산하는 것으로 개인 내 일치도를 보는 것이다. 그런데 컴퓨터의 계산 능력이 급격하게 확대되면서 동일한 구성개념을 측정하는 문항들은 개인 내에서 정적인 상관을 가질 것이라는 가정에 기반하여 여러 문항들이 서로 다른 하위집합에 할당되는 과정을 반복하고 그 결과를 평균 내면 더 강건한 추정치를 얻을 수 있을 것이다(Curran, 2016). 이것이 개인적 신뢰도(personal reliability) 개념으로, 홀-짝 일관성은 문항들에서 하나의 쌍만을 활용하는 개인적 신뢰도의 한 특수한 사례라고 할 수 있다. 개인적 신뢰

17) .60의 상관관계가 역치로 설정되지만(Meade & Craig, 2012), 이 수치는 자료에서 관찰된 상관관계에 따라 다르게 조정될 수도 있다. 역치를 낮추면, 불합리하게 IER 응답자를 늘릴 수도 있다.

도 지표의 사용에 있어 조사에 잘 확립된 여러 척도가 포함되어 있는 경우 더 유리하다.

5) 재표집(resampling). 일반적으로 반분 신뢰도보다 내적 일치도 신뢰도가 더 강건한 추정을 할 수 있는 것은 가능한 모든 반분 방식을 고려하여 평균한 값이기 때문이다. 이는 재표집 과정을 통해 무선 오차를 줄일 수 있다는 개념과 일맥 상통한다. 실제로 이러한 재표집 방식이 IER 탐지에 자주 사용되지는 않지만, 더 나은 IER 추정치를 얻을 수 있다. 어떻게 하위척도나 문항들이 쌍으로 만들어지던 간에 동일한 일치도를 평가한다는 점에서 이것은 재표집과 부트스트래핑을 개인적 신뢰도 개념에 적용한 것이라 할 수 있다. 이러한 새로운 방법인 재표집된 개인적 신뢰도(RIR: resampled individual reliability)는 단일차원의 구성개념을 측정하는 문항들에 대한 개인의 응답에서의 일치도(일관성)와 신뢰도에 대한 강건한 추정을 가능하게 한다. 문항들의 쌍을 어떻게 만들든지 재표집 기법은 무선 오차가 서로 상쇄되도록 한다. 따라서 RIR은 개인적 신뢰도 개념을 더 엄격하게 실행하는 방법일 수 있다 (Curran, 2016).

6) 개인-전체 상관관계(person-total correlation). 이것은 심리측정에서 많이 사용되는 문항-전체(item-total) 상관관계의 역이라고 할 수 있으며, 조사 문항들에 대한 개인의 반응과 표본 내 다른 응답자들의 평균 반응 간의 상관계수를 계산하는 것으로, 한 사람의 반응과 다른 사람들의 반응 간의 유사성 측정치이다 (Curran, 2016). 여기서 표본 크기는 분석에 포함된 문항의 수가 된다. 일반적인 신뢰도 분석에서 문항-전체 상관관계는 특정 문항이 다른 모든 문항을 잘 예측하는 정도를 나타내는 것처럼, 개인-전체 상관관계는 특정인의 응답

이 다른 모든 사람의 응답을 잘 예측하는 정도이다. 특정인의 개인-전체 상관관계가 낮으면 IER일 가능성이 큰 것으로 간주된다.

7) 다분 거트만 오차(polytomous Guttman errors). 거트만 오차 개념은 이분(dichotomous) 검사들에서 나온 것이다. 예를 들어, 거트만 오차는 어려운 문항은 정확하게 답하지만, 상대적으로 쉬운 문항은 틀리게 답할 때 나타난다. 다분 조사 자료에 이것을 적용하면, 극단적으로 진보적 태도를 나타내는 문항에 약하게 동의하였다면, 논리적으로 그보다 완화된 문항에 대해서는 강하게(또는 적어도 약하게) 동의해야 할 것이다. 특정 응답자가 (전체 표본의 평균 반응에 근거할 때) 극단적인 문항에 대해서는 ‘강한 동의’를 보였지만, 이보다 완화된 문항에 대해서는 ‘약한 동의’를 하는 경우 다분 거트만 오차가 발생한다(Curran, 2016). 그리고 이것은 IER에 의한 것일 수 있다.

8) 개인-적합 문항 반응 이론 모형(person-fit item response theory models). 개인-적합 문항 반응 이론(IRT) 모형은 다분 거트만 오차를 훨씬 더 정교한 방법으로 구현한 것이다. 복잡한 모형에 근거하는 이 방법은 자료의 성질이나 특성에 대해 이미 알려진 것들이 있다면, 관찰된 응답에서도 그러한 패턴이 나타나는 정도를 조사하는 것이다. 그래서, 거트만 오차에서처럼, 한 응답자가 상당히 극단적인 문항에 대해 동의하면서도 유사하거나 덜 극단적인 문항들에 동의하지 않는 경우 개인-적합이 낮게 된다.

Patten 등(2019)은 문항 반응 이론 모형을 기반으로 한 개인-적합 통계량(person-fit statistics, Glas & Dagohoy, 2007)을 활용하여 추정된 응답자의 능력(ability)이나 잠재 특성(latent trait)

의 수준에 비추어 볼 때 기대하기 어렵거나 전형적이지 않은 응답 패턴을 식별할 수 있음을 제시하였다. 이들은 1) 주어진 데이터 세트에서 비전형적 응답 패턴을 찾아내어 제거하고, 2) 정화된 데이터를 바탕으로 문항 반응 모형의 파라미터와 잠재 특성 수준을 재추정한 다음, 3) 새롭게 추정된 잠재 특성 수준을 고려하여 비전형적 응답 패턴을 다시 식별하고 제거하는 과정을 반복하였다. 이러한 반복적인 정화 절차(iterative cleansing procedure)를 통해, 자료의 질을 높이고, 문항 반응 모형의 파라미터와 잠재 특성 추정치의 정확도를 높일 수 있음을 보였다.

9) 잠재 프로파일 분석과 요인 혼합 모델링. 이것은 불성실 응답자들 간에 서로 다른 유형이 있는지를 확인하는 방법이다(Meade & Craig, 2012). 잠재 프로파일 분석(LPA: latent profile analysis)은 응답자의 반응 패턴을 기반으로 관찰된 데이터에 잠재적 하위 집단 혹은 프로파일을 식별하는 혼합 모형 접근법 중 하나이다(McLachlan & Peel, 2004). 이 방법은 응답자가 여러 하위 집단 중 어디에 속하는지 모델링하는 데 사용되며, IER을 보이는 하위 집단을 찾는 데에도 유용하게 사용될 수 있다. 예를 들어, Meade와 Craig(2012)는 다양한 IER 탐지 지표를 사용한 잠재 프로파일 분석을 통해 표본의 11% 정도가 IER로 나타남을 보여주었다. 이후 Shulka와 Konold(2018)는 응답 일관성(RI: response inconsistency) 지표(구체적인 공식은 원문의 공식 (2) 참조)를 개발하고 이를 잠재 프로파일 분석에 적용하여 전체 표본 중 약 6.88%에 해당하는 부주의 응답 집단을 식별하였다. 또한, Kupffer 등(2024)과 Hasselhorn 등(2023)은 각각 다차원 강제선택(multidimensional forced choice) 자료와 집중적

종단 자료(intensive longitudinal data)에서 잠재 프로파일 분석으로 부주의 응답자를 가려낼 수 있음을 보여주었다.

요인 혼합 모형(FMM: factor mixture model)은 확인적 요인분석과 잠재 계층 분석을 결합한 모형으로, 다수의 잠재 계층이 존재한다고 가정하며, 이러한 계층 간에 공통요인 모형의 파라미터(예: 요인 부하량, 요인 평균, 요인 공분산)가 서로 다를 수 있음을 인정한다. Meade와 Craig(2012)는 ‘원만성’을 측정하는 자료(n=381)를 요인 혼합 모형을 이용하여 분석한 결과, 요인 부하량이 현저하게 낮은 잠재 계층(n=45)을 식별할 수 있었고, 이 하위 집단은 IER 지표와 밀접한 관련이 있음을 확인하였다. 이러한 결과는 부주의 응답 집단이 공통요인 모형의 파라미터에서 성실한 응답 집단과 구별되며, 요인 혼합 모형을 활용하여 두 집단을 구분할 수 있음을 시사한다. 이후 연구들(Arias et al., 2020; Kam & Cheung, 2023; Steinmann et al., 2022)은 요인 혼합 모형을 활용하여 롱스트링과 무작위 반응 등을 특징으로 하는 부주의한 응답자로 구성된 하위 집단을 식별할 수 있음을 보여주었다.

10) 응답 엔트로피(RE: response entropy).<sup>18)</sup> Tawa(2021)는 RE 지표를 통해 개인의 응답 유

18) 엔트로피는 도자기 컵이 깨지면 부서지면서 결국에는 원재료로 분해되듯이, 시간이 지남에 따라 하나의 체계가 무질서라는 자연스러운 상태를 향해 간다는 아이디어이다. 측정에서도 엔트로피 공식을 통해 한 체계가 절대적인 질서에서 절대적인 무질서까지의 연속체의 어디에 위치하는지를 계량화할 수 있다. 예를 들어, 고난도 인지적 과제를 수행하는 과정에서 컴퓨터 버튼 반응이 무질서해지는 정도나 또래 집단에서 동질적인 인종 구성(질서)에서 인종 다양성이 증가(무질서)하는 것 등이다.

형(response type)의 변산성을 계량화하는 새로운 IER 탐지 방법을 제안하였다. 이것은 리커트 척도를 사용한 조사에서 척도점에 표시한 응답 유형의 변동성을 이용하여 계산한다. 이 지표는 개인 응답 변량(IRV)과 같은 불변성 지표처럼, 성실한 응답자의 일반적인 반응 패턴은 응답자의 관심 변인 관련 특성에 따라 약간의 변동을 보일 뿐이며, 너무 큰 변동이나 너무 적은 변동은 IER로 간주될 수 있다는 가정에 근거한다.

RE 지표는 설문 문항에 대한 응답에서의 무질서 정도를 나타내고자 엔트로피 개념을 적용한 것으로, RE 지표 점수는 문항들에 대한 응답 유형이 최대로 다양할 때 최대값을 가진다(계산 공식은 Tawa, 2021, pp. 303 참조). 예를 들어, 5점 척도의 20 문항이 있다고 할 때, 1, 2, 3, 4, 5라는 다섯 척도점에 균등하게 네 번씩 응답하면 RE는 최대값을, 모든 문항에 대하여 한 척도점에만 응답하면 최소값을 가진다. 조사 문항에 걸쳐 무작위로 반응하거나 하나의 척도점에만 응답 유형이 제약되는 경우는 모두 IER을 드러내는 것으로서, 앞서 살펴본 IER의 불변성(또는 변산성) 지표들은 IER의 두 유형 중 하나(예, 롱스트링은 불변성, 개인 응답 변량은 변산성)만을 평가하지만, RE 지표는 이 연속선의 양극을 동시에 평가한다. 또한, 개인 응답 변량(IRV)은 응답 유형의 변산성과 응답 값의 변산성을 혼동하는데, 1과 5에만 응답하는 사람은 2와 4에만 응답하는 사람보다 훨씬 더 높은 IRV 지수를 가지지만, 두 사람은 응답 유형에서는 동일한 수준(즉, 두 개의 선택지에만 응답함)을 보이는 것이다. RE는 IRV와 달리 반응의 값(value)이 아니라 유형(type)의 다양성을 반영한다. 예를 들면 20문항에 대하여 3과 4의 두 유형으

로만 응답한 경우와 1과 5의 두 유형으로만 응답한 경우 동일한 RE를 갖는다.

Tawa(2021)는 RE 지표가 IER 탐지 지표로서 효과적인지를 조사하고자 세 유형의 가짜 데이터 세트를 만들었다. 그것은 컴퓨터로 완전 무작위로 생성된 데이터(n=100), 컴퓨터로 정규분포하도록 생성된 무작위 데이터(n=100), 참가자가 의도적으로 부주의하게 반응하는 데이터였다(n=100). 이 가짜 데이터 세트를 깨끗한(IER이 최소화된) 사회과학 데이터(n=500)에 포함하여 다항 로지스틱 회귀 분석을 하였을 때, RE 지표는 가짜 데이터 세트를 예측하는 데 있어 다른 지표들(마할라노비스 거리와 심리측정적 동의어)과는 독립적으로 유의한 기여를 하는 것으로 확인되었다. 이는 RE 지표가 IER 탐지에 있어 다른 IER 지표들과는 다른 양상을 포착할 수 있는 능력이 있음을 시사한다.

지금까지 살펴본 것처럼 IER을 탐지하기 위한 다양한 방법들이 있어왔고, 계속 새로운 시도들이 이루어지고 있다. 참고로, IER 탐지 지표들을 자동적으로 계산하는데 필요한 R package 도구들을 모은 Yentes와 Wilhelm (2023) 웹사이트와 저자들이 만든 웹사이트 (<https://ier.notion.site/IER-R-8acbd066f3a042ca82b8253d28b39e30>)가 있다.

## IER 탐지 지표 간 관계 및 비교

IER 지표들의 효과성을 결정하기 위한 시뮬레이션 연구를 수행한 Meade와 Craig(2012)은 첫 번째 연구에서 IER이 무작위 및 비무작위 응답과 같이 두 개의 구분되는 응답 패턴이 나타났으며, 이 두 응답 패턴을 탐지하는 데에는 다른 지표들이 필요함을 보여주었다. 즉

롱스트링은 불변 반응을 탐지하는 데 효과적이었고, 마할라노비스 거리와 흠-짜 일관성은 무작위 반응을 탐지하는 데 가장 뛰어났다. 두 번째 연구에서는 시물레이션 자료를 사용하여 IER 탐지 지표들의 효과성을 비교하였는데, 자료의 성질에 따라 지표의 효과성이 달라졌지만, 양질의 자료를 담보하기 위해서 사전에 지시 응답 문항을 사용하고, 사후적으로 일관성 지표들과 마할라노비스 거리와 같은 이상치 분석을 추천하였다. 그리고 그들은 복수의 IER 지표들을 고려하는 요인 혼합 모델링 분석에서 최상의 결과를 얻었다. 더불어 그들의 연구는 IER 지표들이 세 요인으로 경험적으로 구분됨을 보여주었다. 첫 번째 요인은 회귀(regression) 유형의 지표인 심리측정적 동의어와 반의어, 흠-짜 일관성, 마할라노비스 거리, 가짜 문항이었다. 두 번째 요인은 주의나 노력을 기울인 정도에 대한 네 개의 자기 보고 문항들이었으며,<sup>19)</sup> 세 번째 요인은 두 개의 롱스트링 지표(10개의 웹 페이지별 롱스트링 지수의 평균, 10개 중 최대 롱스트링)였다. 유사하게 Grau 등(2019)과 Huang 등(2012)도 롱스트링 지표는 회귀 유형의 지표들과는 구분되는 요인임을 보여주었다. 이것은 IER이 유의한 요인구조를 가진 다차원의 구성개념임을 나타낸다.

Goldammer 등(2020, 연구 1)은 IER을 탐지하는 방법(지표)들의 정확성을 조사하였다. 그들

19) 네 가지는 다음과 같다. 첫째 Use Me는 자신의 응답이 분석에 사용되어야 할지에 대한 ‘예/아니오’의 자기 보고 문항이다. 둘째, Diligence는 자기 보고형 근면성(diligence) 척도 문항들이다. 셋째, Attention은 연구에 대해 주의를 기울인 정도에 대한 자기 보고형 단일 문항이다. 마지막으로 Effort는 연구에 들인 노력에 대한 자기 보고형 단일 문항이다.

은 일부 참여자들에게 부주의하게 응답하도록 지시하는 색다른 방식을 사용하여 세 가지 실험 조건(리더십 관련 6개 척도의 모든 문항에 신중하게 응답, 각 척도의 50% 문항에 대해 무작위로 응답, 각 척도의 50% 문항에 대해 반대로 응답하는 조건)을 만들었고, IER 탐지의 조건 간 7개 지표의 효과성을 조사하였다. 문항당 평균 반응 시간, 개인적 신뢰도, 심리측정적 동의어와 반의어, 마할라노비스 거리는 IER 탐지에 효과적이었으나, 앞서 언급했듯이, 선행 연구(Dunn et al., 2018; Huang et al., 2012; Meade & Craig, 2012)와 달리 롱스트링과 개인 응답 변량은 그리 효과적이지 않았다. 연구자들은 이러한 결과를 불변성 지표들은 척도 성질에 따라 효과성이 달라질 수 있음을 시사하는 것으로 보았다. 즉, 성격 평가처럼 균형 척도를 사용하는 경우에는 효과적이지만, 변혁적 리더십 척도처럼 일방향으로 채점되는 척도에서는 효과적이지 않을 수 있다고 하였다. 이것은 후속 연구에서 다시 검증되어야 할 이슈이다.

## 불성실 응답의 방지 방법

앞서 언급했듯이, IER을 탐지하기 전에 그것의 발생을 예방하는 것이 더 우선적이다. 그동안 IER을 예방하는 방안이 다양하게 시도되었고, 일부는 어느 정도 효과가 있으나, 기대와 달리 별 효과가 없는 것으로 나타나기도 하였다.

### 경고

불성실 응답 방지를 위해 이미 널리 사용되

는 방법은 지시문에 불성실 응답에 대한 경고 문구를 넣는 것이다(Huang et al, 2012; Meade & Craig, 2012; Ward & Pond, 2015). 예를 들어, 대학생들을 대상으로 한 Huang 등(2012)의 연구에서는 ‘설문 응답에 노력을 기울이지 않으면 학점에 부정적 영향을 미칠 수 있다’는 경고 문구를 제시함으로써 응답자들의 불성실한 응답을 줄이고자 하였다. 연구 결과, 일반적인 지시문을 사용한 통제조건에 비해 경고 문구 조건에서 응답자들의 불성실 응답이 유의하게 감소하였다. 이처럼 불성실한 응답이 통계적으로 확인되어 가려낼 것이라는 경고나 보상이 지급되지 않을 것이라는 경고는 일정 효과가 있다. 하지만 경고 문구는 응답자의 책임감을 높이기도 하지만 강압적인 느낌을 주면서 설문 자체에 대한 내적 흥미를 떨어뜨리고, 반발심을 살 수도 있기에 좋은 방법이 아닐 수 있다(Meade & Craig, 2012). 대체로 대학생 표본에서는 경고의 부정적 효과가 크지 않을 수 있지만, 성인(직장인이나 종업원)이나 장기간 긍정적 관계를 유지해야 하는 응답자(고객이나 소비자)의 경우에 경고는 얻는 것보다 잃는 것이 더 많은 방안일 수 있다.

Bowling 등(2021)은 두 실험 연구를 통해 설문 응답에서의 IER에 대해 처벌받을 것이라고 경고를 받은 응답자들의 부주의 정도가 감소하는지는 조사하였을 때 혼합된 결과가 나타났다. 감시가 없는 온라인 조사와 감시가 있는 실험실 연구의 비교에서도 혼합된 결과가 나타났다.

IER 탐지 방법을 외현적 방법과 암묵적 방법으로 구분한<sup>20)</sup> Toich 등(2022)은 긍정적 경

고(주의 깊은 응답이 연구에 중요하며, 그렇지 않으면 오랜 연구가 허사로 돌아감, 성실 답변에 감사) 조건, 부정적 경고(부주의 응답 탐지 가능하며, 탐지되면 학점을 받지 못함) 조건, 그리고 무경고 조건에서 IER이 나타나는 정도를 MTurk 참여자와 대학생의 두 표본을 가지고 비교하였다. 또한, 암묵적 및 외현적 탐지의 두 방법은 IER의 다른 양상을 포착하는 것으로 보고, 이 둘의 탐지 효과성을 비교하였다. 분석 결과는 IER을 감소시키는 데 경고 사용은 별 효용이 없었다. MTurk 참여자보다 대학생 표본에서 5개의 IER 지표(지시 응답 문항, 자기 보고, 가짜 문항, 심리측정적 동의어, 개인 응답 변량) 중 2개에서 더 높게 탐지되었지만, 나머지 세 개 지표에서는 두 표본 간에 차이가 없었다. 그리고 암묵적 지표와 외현적 지표에서도 차이가 없었다. 이런 결과는 경고의 효과성에 대해 다시금 의구심을 품게 한다.

또 다른 방법은 설문이 끝난 후에 참가자들에게 응답의 질에 대한 피드백을 주겠다는 지시문을 활용하는 것이다(Ward & Pond, 2015). 이 방법은 과업에 대한 피드백이 있을 것을 예상하면, 응답자들이 과업에 더 많은 주의를 기울이게 되고(Kluger & DeNisi, 1996), 과업 수행이 향상된다는 연구(Northcraft et al., 2011)에 근거한 것이다. Ward와 Pond(2015)의 연구에서는 설문 조사가 완료된 후 응답자가 제공한 정보의 사용 여부 및 응답의 질에 대한 피드백을 줄 것이라는 문구를 적용했다. 연구 결과, 일부 지표에서는 피드백 조건의 참가자들이 더 성실하게 응답하는 것으로 나타났으나,

속한다. 반면, 암묵적 지표는 참여자들이 전혀 알지 못하며, 사후 통계적 분석으로 IER이 탐지되는 방법들이다.

20) 외현적 지표는 정오의 답이 분명하며, 틀리면 IER일 가능성이 크다. 가짜 문항, 지시 응답 문항, 사용 여부에 대한 자기 보고 문항 등이 이에

다른 지표에서는 반대의 결과가 나타나서 피드백 지시문이 IER에 미치는 효과는 일관되지 않은 것으로 나타났다.

### 사회심리학 기반 방안들

그동안 IER 예방 방안을 찾는 데 있어 사회심리학적 이론이나 개념들을 활용하는 경우가 많았다. 대표적인 예로, 사회적 영향력(social influence) 관련 이론을 지시문에 적용하여 성실한 응답의 동기를 높이려는 시도가 Ward와 동료들(Ward & Meade, 2018; Ward & Pond, 2015)에 의해 이루어졌다.

Ward와 Meade(2018)는 응답자들이 주의 깊게 응답하려는 동기를 높일 수 있는 세 가지 방법을 사회심리학적 이론에 근거하여 제시하면서 그 효과를 검증하였다. 연구 1에서는 온라인 설문에서 IER을 유발하는 원인 중 하나인 연구자와의 직접적인 접촉의 부재를 극복하는 방안으로 연구자가 지시문을 직접 읽는 비디오 클립을 설문 초반에 제시하였다. 이러한 조작은 응답자가 지각하는 연구자와의 물리적 거리를 줄이고, 연구자의 존재 또는 권위가 두드러져 사회적 영향력이 강화될 것이기에, 성실한 응답이 유도될 것으로 보였다. 실험조건은 1) 익명이 보장되며 솔직하게 응답하도록 요구하는 일반적인 지시문이 담긴 통제조건, 2) 연구자 소개, 실험 참여의 의미, 감사 글이 기술된 텍스트를 읽는 스크립트 조건, 3) 스크립트 조건의 글을 연구자가 말로 하는 동영상을 보는 비디오 지시조건이었다. 분석 결과, 통제조건의 참여자들이 스크립트와 비디오 조건보다 더 IER의 주관적 측정치(응답의 성실성에 대한 자기 보고 측정치)에서 낮게 나타났지만, 통제조건 대비 스크립트

조건은 설문에 대한 흥미가 낮았다. IER의 객관적 지표들에서는 세 집단 간에 아무런 차이가 없어, 전체적으로 실험 1의 조작은 주목할 만한 효과가 없었다. 이 방법과 유사하게, Ward와 Pond(2015)는 온라인 설문 페이지의 왼쪽 위 빈 곳에 가상 인간을 배치하여 연구자와 응답자 간 지각된 상호작용을 높이는 조건을 만들었다. 가상 인간 조건에 할당된 참가자들은 설문 응답 내내 고개를 끄덕이거나 눈을 깜빡이는 등의 행동을 취하는 가상 인간의 모습을 볼 수 있었다. 하지만 가상 인간의 모습 자체로는 IER을 방지하는 효과가 없는 것으로 나타났다.

Ward와 Meade(2018)의 연구 2에서는 부조화에 기반한 위선(hypocrisy) 패러다임을 적용하여 응답자의 성실한 응답을 유도하고자 하였다. 위선 패러다임의 논리는 사람들이 자신이 가지고 있는 태도에 반하는 위선의 예시가 마음속에 떠오르면 기존 태도와 일치하는 행동이 강화된다는 것이다(Aronson et al., 1991). 여기서는 통제조건과 더불어 두 개의 위선 조건이 추가되었다. 하나는 인지 부조화를 유도한 지시문 조건이었고,<sup>21)</sup> 다른 하나는 위선의 느낌을 유발하는 지시문 조건이었다.<sup>22)</sup> 이 두

21) 참가자들은 심리학 연구 수행에 들어가는 노력이나 작업의 양에 관한 내용이 담긴 지시문을 읽었고, 같은 웹 페이지에서 ‘이 연구를 마치는 데 한 시간이 걸리는 것을 인지하였다’에 전자 서명하였고, 응답자가 해야 할 일에 대해 읽고, 주의 깊게 읽고 솔직하게 응답할 것임에 대해서도 서명하였다.

22) 지시문을 통해 참가자들은 자신을 모범 학생으로 보도록 점화되었고, 조사에 주의 깊게 응답하도록 설득하는 메시지를 만드는 다른 연구자들을 돕도록 요청하는 지시문을 읽었다. 이후 그들은 웹 페이지에 제공된 공간에 주의 깊은 응답이 중

조건에서 응답자들은 성실한 응답이 중요하다는 태도를 지니게 될 것이며, 추후 설문에서 불성실하게 응답하는 것은 부조화 경험을 초래하거나 위선 행동에 해당하게 된다. 따라서 부조화나 위선 행동으로 유발되는 불편한 마음을 경험하지 않기 위해 기존 태도와 일치되게 더 성실하게 응답할 것이라고 예상할 수 있다. 연구 결과, 불성실 응답이 유의하게 줄어들었다. 부조화 집단에서 반응의 논리적 일관성과 조사에 대한 흥미가 증진되었다. 위선 집단은 지시 응답 문항에의 정확성을 높였다. 구체적으로 잠재적인 부조화 또는 위선의 유도가 IER의 10개 지표 중 3개에서 긍정적인 효과를 보였다. 이것은 IER 방지에 있어 연구 1보다는 훨씬 주목할 만한 결과이지만, 여전히 제한적인 효과만을 갖는 것이다. 연구자들은 더 강력한 효과를 얻고자 연구 3을 진행하였다. 연구 3에서는 조사 지시문에 사회교환 이론을 적용하였다. 즉 설문 조사 참여에 대해 보상(선물)을 주는 호혜성 조건과 설문 참여하면 자선단체에 기부하는 효과가 있는 집단 이득 조건을 통제 집단과 비교하였다. 분석 결과, 연구 표본인 학생들의 경우에는 이런 조작이 IER에 아무런 효과가 없었다.

한편 김우영 등(2019)은 온라인 설문에서 불성실 응답을 방지하기 위해 객관적 자기 인식 이론(objective self-awareness)을 적용하여<sup>23)</sup> 거울

요한 이유를 적어야 했다. 이어지는 웹 페이지에 참가자들은 지난 해 조사에 주의 깊게 응답하지 않는 횟수를 기입하도록 하였다.

23) Beaman 등(1979)은 할로윈데이 시즌에 사탕을 가지러 온 아이들에게 사탕을 하나씩만 가져가라고 요청을 한 다음, 아이들이 연구자의 말을 따르는지 관찰했다. 두 조건 중 한 조건은 사탕 바구니 앞에 거울이 배치되었고, 다른 조건에서는 거울이 없었다. 연구 결과, 거울이 있는 조건에서

이미지를 배경으로 삽입하는 온라인 설문 화면을 구성하였고, 이것의 불성실 응답 방지 효과를 파악하기 위해 경고 지시문을 활용한 조건 및 일반적인 조사 지시문을 활용한 통제 조건과 비교하였다. 연구 결과, 경고 지시문 활용이나 일반적인 지시문을 활용한 통제조건 대비 거울 이미지를 삽입한 조건에서 불성실 응답이 감소하는 경향이 있는 것으로 나타났다. 즉, 분석에 사용된 총 10개의 IER 지표 중 3개의 지표(심리측정적 반의어, 응답 시간, 가짜 문항)에서 통제조건 대비 거울 조건에서 불성실 응답이 감소하였다. 또한, 불성실 응답자의 비율을 조건 간 비교하기 위한 추가 분석에서는 통제조건 대비 거울 조건의 불성실 응답 비율이 유의하게 낮게 나타났기 때문에 거울 이미지의 활용이 불성실 응답을 어느 정도 방지할 수 있음을 보여주었다. 이러한 결과는 개인의 분명한 의도나 의식적인 지각없이 단순히 특정 사물에 노출되는 것만으로도 그 사물과 관련된 태도나 고정관념이 자동적으로 활성화되고 실제 행동이나 생각에 영향을 미칠 수 있다는 연구들(Bargh et al., 2001; Perdue & Gurtman, 1990)과 일치하는 것이다.<sup>24)</sup> 거울 조건의 응답자들이 실제로 거울을 본 것은 아니지만 배경에 제시된 거울 이미지에 계속 노출되면서 거울과 관련된 의미체계(예:

연구자의 말을 더 잘 따르는 것으로 나타났는데, 이는 사탕을 집어갈 때 거울을 통해 자신의 얼굴을 보게 되고, 이로 인해 객관적 자기인식(objective self-awareness)이 유발되었기 때문이다. 객관적 자기인식이란 개인의 주의를 자기 내부로 향하는 상태를 의미한다(Duval & Wicklund, 1972).

24) 추가로 118명의 대학생 대상으로 거울 조건의 설문을 진행한 결과, 대다수 대학생(76%)이 설문 화면에 아무것도 없었다고 응답하였으며, 17%만이 거울의 존재를 인식하였다.



진실), 지식(예: 자신을 비출 수 있는 도구), 혹은 거울을 보았던 경험들이 활성화되었을 것이다. 이를 통해 객관적 자기인식이 유발되고, 지시문에서 제시한 ‘성실한 응답 요구’라는 규범에 더 부합하는 행동을 한 것으로 추측된다.

김우영(2020)은 후속 연구에서 거울 이미지 삽입의 불성실 응답 방지 효과를 반복검증하고, 거울 이미지 외에 자기인식을 높이는 다른 방법들(얼굴 형태 이미지선택, 개인적 일화 회상<sup>25)</sup>)의 효과를 탐색하였다. 연구 결과, 거울 조건의 응답자들은 훌쩍 일관성 지표와 시간 지표에서 불성실 응답이 감소하는 경향성이 나타났으며, 이미지선택 조건의 참가자들은 시간 지표에서 불성실 응답의 감소가 나타났으나, 통제조건과 회상 조건 간에는 아무런 차이가 없었다. 다소 약하지만, 여전히 자기인식 증진이 불성실 응답을 방지할 가능성을 보여주는 결과이다. 하지만 조작 체크를 위해 삽입한 자기인식 척도 점수에서 조건 간 차이가 나타나지 않았으므로 조건 간에 나타난 차이를 자기인식 증진의 효과로 설명하기 어려운 제한점이 있었다.

지금까지 소개한 IER 방지 방법들은 성실한 응답에의 동기유발 및 행동을 유도하기 위한 것으로 일부 긍정적인 효과를 보이기도 하지만, 어느 한 가지 방법도 일관되게 뚜렷한 효과를 보인 경우는 드물다. 따라서 선제적으로

IER을 예방하는 실효성 있는 방안들은 여전히 부족한 편으로 더 많은 연구가 필요하다.

## IER 예방 방안 추천

IER의 효율적인 탐지보다는 사전 예방이 더 선호되기에, 제한된 연구에 기반하지만, 다음과 같은 방법들을 강구할 수 있다. 그리고 지금까지의 연구에 의하면, IER이 발생하기 전에 미리 예방하는 것이 어느 정도는 가능하다.

Goldammer 등(2020)은 IER에 대한 사전 대비 및 예방을 위하여 다음의 다섯 가지를 제안하였다. 첫째, 응답자들에게 인센티브를 제공함으로써 설문에 응답하는 동안 동기 수준을 높일 필요가 있다. 둘째, 연구자가 직접 응답자들에게 조사 관련 지시문을 제시하면서 응답자들의 연구(조사 및 설문)에 대한 몰입 수준을 높인다. 셋째, 연구 목적에 필요한 문항들만을 설문에 포함하여 가급적 짧은 설문지를 만든다. 넷째, 설문의 앞부분에 주요 문항들을 배치하는 것이 좋다. 마지막으로, IER로 응답자 수가 줄어든 것을 예상하여 전체 표본 수를 넉넉하게 잡는다.

Ward와 Meade(2023)도 다음과 같은 방안들을 고려할 것을 제안하였다. 첫째, 조사나 설문의 지시문에 부주의하게 응답하면 어떤 결과가 나오는지 응답자에게 분명하게 주지시키면 일정 수준 IER을 예방할 수 있다(물론 불편하거나 다소 불쾌한 느낌을 줄 위험도 있다). 또한, 민감한 이슈가 아니라면 익명이 아니라 나중에 확인이 된다는 언급도 효과가 있지만(Meade & Craig, 2012), 이는 여전히 솔직한 응답을 저해할 가능성이 있다. Ward와 Meade(2018)의 연구에서 보았듯이, 성실한 응

25) 이미지선택 조건에서는 지시문 바로 다음 페이지에서 자신의 얼굴과 유사한 형태의 이미지를 선택하는 문항과 자신의 머리카락 색깔, 얼굴에 있는 점의 개수를 물어보는 문항이 삽입되었다. 회상 조건에서는 지시문 바로 다음 페이지에서 최근 자신이 한 달 동안 겪은 일화를 떠올리고, 그것을 직접 기술해 달라고 요구하였다.

답 관련 진술문에 자신의 이름(이니셜)을 입력하도록 하였을 때 응답 일관성, 정확성, 조사에의 흥미도가 높아졌다. 또한, 부주의하게 응답하는 것이 부조화나 위선적인 것으로 느껴지도록 하면 비밀관성과 부정확성이 감소될 수 있다. 성실한 응답에 대해 금전적 보상을 주는 것은 일정 효과가 있으나(Gibson & Bowling, 2020), 정중하게 성실한 응답을 요청하는 것이나 연구자의 사회적 영향력을 높이는 방식(Ward & Meade, 2018)은 아직 효과성이 충분히 검증되지 않았다.

둘째, 감시(monitoring)의 수준을 높이는 것이다. 실제로 감시하거나 감시를 한다는 시각적 표식들은 일부 형태의 IER을 효과적으로 예방하는 것으로 보인다. Francavilla 등(2019)은 서로 다른 수준의 감시(즉, 온라인 비감시 조건, 온라인 가상 감시 조건, 교실에서의 직접 감시)가 IER의 세부 지표들에 영향을 준다는 것을 보여주었다. 비감시 조건과 비교하여 가상의 감시 조건에서 가짜 문항에서 IER이 줄었고, 직접 감시 조건에서는 가짜 문항에서 IER이 줄어든 것과 더불어 성실하게 응답하였다는 자기 보고 점수도 더 높았다. 환경의 방해는 직접 감시와 성실 응답의 자기 보고 간 관계를 완전 매개하였는데, 이는 응답자의 주의 집중을 위해서는 환경의 방해가 없어야 함을 시사한다. 앞서 김우영 등(2019)의 연구에서 거울을 웹 페이지 바탕에 제시하는 것도 일종의 감시로서 IER을 어느 정도 줄이는 효과를 보였다. 다만, 자기인식 수준의 매개 효과는 나타나지 않아 후속 연구가 필요하다.

셋째, 온라인 조사 웹페이지의 디자인 특성도 IER을 줄이는 효과를 볼 수 있다. Ward와 Pond(2015)는 응답자에게 가상의 인간을 드러냄으로써 IER을 줄였다. 응답의 질이 모니터

링될 것이라는 경고가 있었기에, 가상의 인물이 응답자가 하는 행동에 주의를 기울이는 것처럼 보였을 것이다. 다만 이런 해석은 아직 경험적으로 검증되지는 않았다.

또한, 온라인 조사가 늘어나면서 자주 사용되는 감시 방법 중으로, 설문 응답 과정에서 주의 체크 문항에 제대로 답하지 못할 때 즉각적으로(알고리즘에 의해 자동적으로) 주의를 기울이도록 경고하는 라이브 조사(living survey) 방식도 있다. 이 방식이 미치는 효과에 대해서는 아직 충분한 연구가 이루어지지 않았다.

### IER 응답자 제외 관련 의사결정

Ward와 Meade(2023)는 앞서 언급한 IER이 발생할 확률이 높은 조건 중 하나라도 존재할 때에는, 다음에서 소개하듯이, IER 반응 탐지의 최소한의 기준을 적용할 것을 강하게 주장하였으며, 필요에 따라서는 해당 IER 자료를 제거하는 방안을 고려해야 한다고 하였다.

특정 응답자가 IER을 보여 표본에서 제외할지의 결정 기준은 사용한 탐지 방법에 따라 다를 것이다. 지시 응답 문항은 지시한 대로 제대로 응답했는지를 보면 되기에 간단하지만, 이 경우에도 지시 응답 문항이 두 개 이상이라면 몇 번 잘못 응답해야 IER로 간주할지 판단이 필요하다. 또한, 앞서 언급했듯이, 페이지 단위로 반응 시간을 측정한 경우, 문항당 2초라는 기준이 추천되지만, 여전히 기준(역치)에 대해서는 추가 고려의 여지가 있다. 기준을 엄격하게 정하여 IER 탐지의 가능성을 높이면 검정력 손실이 나타날 수 있기에 신중히 판단해야 한다.

롱스트링과 같은 방법은 기준을 정하는 것이 더 복잡하다. 6개에서 14개까지 구체적인 수치를 사용할 것을 제안하는 연구자도 있지만(Huang et al., 2012), 조사되는 구성개념의 성질에 따라 연속되는 동일 반응의 수가 어느 정도 영향을 받기에 IER 판정도 이에 따라 달라질 수 있다. 동일하거나 매우 상관이 높은 구성개념들을 측정하는 문항들이 포함되어 있다면, 연속되는 동일 반응이 나타날 확률이 높다. 반면, 문항들이 역접수 문항이나 서로 독립적인 구성개념들을 측정하는 것이라면 동일 반응이 계속되는 것은 매우 예외적인 일 것이다. 다른 많은 지표처럼, 롱스트링도 연구자의 자료에 부합하는 경험적으로 추출된 컷오프 점수를 적용해야 할 것이다. 한편 Yentes (2020)은 일부 IER 지표에 대한 컷오프 점수를 확인하는 방법들을 평가하기 위해 시뮬레이션 연구를 수행하였다. 그 결과, 롱스트링의 경우 표본 평균보다 0.4 표준편차 높게 컷오프 점수로 설정하는 것이, 마할라노비스 거리의 경우 표본 평균의 0.5 표준편차 높게 컷오프 점수를 설정하는 것이 적정하다(이 경우 상당수 성실한 응답자의 손실도 초래한다)는 결과를 얻었다. 마지막으로 홀-짝 일관성의 경우에는 일반화할 수 있는 컷오프 점수를 찾기 어려웠다.

심리측정적 동의어와 같은 지표도 응답 간 상관 정도에 대한 결정(예, .60 또는 .70의 컷오프)과 더불어 문항 쌍의 수를 얼마나 해야 하는지의 결정도 필요하다. 신뢰로운 지표를 만들기 위해 충분한 문항 쌍을 가지는 것이 좋을 것이지만, 여기서 추가로 고려해야 할 사항은 다수 응답자가 성실하게 응답한다는 암묵적 가정이다. 예를 들어, 다수가 부주의하게 응답한다면, 유사 문항 간의 상관관계는

높지 않을 것이다. 그러면 심리측정적 동의어/반의어와 같은 지표들을 계산하기 쉽지 않을 것이다. 이상치 분석이나 개인-전체 상관관계도 전체 표본 중 부주의하게 응답하는 사람이 많지 않다는 전제가 충족되어야 한다.

### IER 응답자 제거의 가이드라인

가용한 여러 지표를 통해 IER을 찾아내 제거하는 경우에 대한 명확한 지침이 최근까지 존재하지 않았으나, Ward와 Meade(2023)는 이에 대해 다음과 같은 의견을 제시하였다. 먼저 연구자는 IER 여부 결정 및 해당 응답자의 제거 여부가 갖는 찬과 반 중의 어느 쪽에 무게를 둘지 고려한다. 예를 들어, 평균과 표준편차의 정확한 추정이 요구되고 표본 크기가 매우 크다면 충실한 자료를 위해 응답자들을 많이 제거할 수 있을 것이다. 반면, 표본 크기가 크지 않고 분석이 상관관계처럼 약간 더 강건한(robust) 것이라면, 아주 분명하게 부주의가 드러나는 응답자만을 제거할 수도 있다.<sup>26)</sup>

또한, 응답 시간이 길거나 롱스트링 지수가 낮을수록 더 성실한 응답을 의미하는 것은 아님을 인식해야 한다. 즉 IER 지표들은 자료의 질에 대한 연속(continuous) 지표가 아니라는 것이다. 예를 들어, 동일한 반응이 30개 이상 계속된 롱스트링 지수는 분명 IER을 나타내지만, 4의 롱스트링 지수가 6의 롱스트링 지수보다 더 나은 자료 질을 의미하는 것은 아니다. 그보다는 임상 척도들처럼, 특정 지표가

26) 그러나 소규모 표본에서 일부 사례의 제거가 상관관계수에 상당한 영향을 미칠 수도 있음을 고려할 때 이 의견은 조심스럽게 받아들여야 할 것이다.

역치를 넘어서면 문제가 있는 것으로 보는 것이 더 적절하다. 따라서 대다수 IER 지표들은 연속 변인이 아니라 이분 변인으로 봐야 한다. 이 점은 IER과 연관된 상관 변인이나 예측 변인(예, 성격)을 연구할 때 특히 유의해야 할 사항이다.

### IER 자료의 제거의 세 수준.

IER이 확인되면 표본에서 제외하는 것이 기본이다. 그러나 지금까지 표본에서 자료를 제외하는 명확한 기준을 확립한 연구는 없다<sup>27)</sup>. 그럼에도 Arthur 등(2021)은 IER 응답자를 제거하는 기준은 분석 결과를 도출하기 전에 결정되어 있어야 하며, 그렇지 않으면 연구 윤리 문제가 생길 수 있다고 경고하였다.

IER 탐지 지표 중에서 엄격한 지표를 사용하는 것이 항상 최상의 선택은 아니다. 많은 경우, 그렇게 하면 가용 표본 크기의 50% 이상이 손실될 수도 있다. 적정 표본 크기를 유지하기 위해서는 간혹 부주의한 반응을 보이는 것으로 보이는 응답자들은 보유하는 것이 나올 수 있다. 예를 들어, Huang 등(2012)은 심리측정적 반의어에서 개인 내 정적 상관관계를 보이고 심리측정적 동의어에서 부적 상관관계를 보이는 응답자들만을 제거하는 매우 보수적인 접근을 취할 것을 추천하였다. 또한, 일부 연구가 보여주듯이(Bowling et al., 2016;

Ward et al., 2017), 엄격한 제거 기준을 갖는 것은 관심 있는 변인의 체계적 변산을 제거하는 위험이 뒤따른다. Ward와 Meade(2023)은 다음과 같이 세 수준에서의 IER 응답자 제거에 대해 제안하였다(표 1 참조).

1) 최소한(minimal)의 제거. 사전 조치가 가능하다면, 지시 응답 문항과 페이지 당 응답 시간을 가장 추천한다. 전자는 설문에 쉽게 추가할 수 있으며, 후자도 일반적인 소프트웨어 도구들에서 가용하다(페이지 간 경과 시간을 계산하는 것이 약간 더 어렵지만). 이 두 가지 방법으로 아주 심한 IER을 최소한의 노력으로 탐지하고 제외할 수 있다. 단점은 보상을 위해 설문에 참여하는 영리한 응답자들에게는 취약하다는 것이다. 연구자가 IER 탐지 방안을 사전에 몰랐거나 적용하지 않았다면 자료 분석을 하기 전에 IER 탐지 및 제거를 위한 최소한 조치라도 취해야 한다. 사후 조치의 최소한의 방법은 마할라노비스 거리와 같은 이상치 분석과 롱스트링 지수 또는 개인 응답 변량으로 응답 불변성을 찾아내는 것이다. 불변 응답자는 전체 IER 중 소규모 비율로 나타나는 경향이 있지만, 자료를 가장 편향시키는 것일 수 있다.

2) 적정 수준(moderate)의 제거. 여기서도 사전 조치로 지시 응답 문항과 페이지당 응답 시간이 추천된다. 보상을 위해 설문에 참여하는 영리한 응답자들이 많을 것으로 보이면, 가짜 문항을 추가하는 것이 좋다. 이때 가짜 문항이 단순히 재미거리가 되지 않도록 유의해야 한다. 학생 표본의 경우 가짜 문항을 단순히 재미로 동의할 가능성이 있기 때문이다(Meade & Craig, 2012). 사후 분석에서는 앞서 추천한 불변성 분석 및 다변량 이상치 분석에 더하여 최소한 두 가지의 일관성 지표가 추천

27) Wang과 Hau(2024)는 준거 변수와의 상관 계수 변화를 기준으로 IER 응답자를 어느 정도까지 제거할 것인지를 결정하는 방안을 제시하였다. 다만, 시뮬레이션 기반 방법으로서 응용연구자들이 즉각 활용할 수 있는 가이드라인을 제공할 만한 결과는 아닌 것으로 판단하여 본문에 포함하지는 않았다.

표 1. IER 제거의 세 수준별 탐지 방법 추천(Ward & Meade, 2023, p. 591)

제거 수준	사전 조치	사후 조치
최소	○ 지시 응답 문항 ○ 페이지당 응답 시간	○ 불변성 분석(예, 롱스트링이나 개인 응답 변량 지표) ○ 다변량 이상치 분석(예, 마할라노비스 거리)
적정	○ 지시 응답 문항 ○ 페이지당 응답 시간	○ 불변성 분석(예, 롱스트링이나 개인 응답 변량 지표) ○ 다변량 이상치 분석(예, 마할라노비스 거리) ○ 두 가지 일관성 지표(예, 심리측정적 동의어, 흘-짜 지표)
최대	○ 지시 응답 문항 ○ 페이지당 응답 시간	○ 불변성 분석(예, 롱스트링이나 개인 응답 변량 지표) ○ 다변량 이상치 분석(예, 마할라노비스 거리) ○ 세 가지 일관성 지표 ○ 개인-적합 지표 ○ 재표집 통계치 ○ 순차적인 제거 절차 적용(예, 불변성 기반 제거 후, 다른 지표들 계산) ○ 잠재 계층 분석 또는 요인 혼합 모델

되는데, 복수의 문항들로 구성된 여러 구성개념 척도들을 사용한 경우라면 흘-짜 지표와 심리측정적 동의어가 좋은 대안이다.

3) 대폭 수준(extensive)의 제거. 사전 조치로는 적정 수준의 제거와 동일하다. 철저하게 IER을 가려내려면 사후 조치 방법을 모두 사용하는 것이 좋다. 특히 앞서 언급한 IER이 나타날 조건들을 갖추고 있고 표본이 제법 크다면 다양한 사후 조치를 고려해야 할 것이다. 첫째, IER 지표들을 더 많이 고려해야 한다. 예를 들어, 개인-적합 IRT 통계치를 포함한 여러 일관성 지표들이 유용하다. 여기서 주목할 점은 불변 응답자들을 먼저 제거한 후 심리측정적 동의어나 이상치 분석을 하는 것처럼 다중 허들(multiple hurdles)의 순차적(sequential) 과정을 거치는 것이 효과적이다(박원우 등, 2020; Curran, 2016). 앞서 언급된 Patton 등(2019)의 시뮬레이션 연구에서도 개인-적합 측정치를 사용하여 부주의 응답자들

을 확인하고 제거하는 반복적 과정이 좋은 결과를 얻었다. 유사하게 많은 부주의 응답 지표들에 대한 재표집을 사용하면 한 번의 계산으로 부주의 여부를 가리는 것보다 더 정확한 추정치를 제공할 수 있다. 마지막으로, 복잡한 수학적 모델링을 사용하면 IER을 더 포괄적으로 검토할 수 있다. 예를 들어, IER 지표들을 측정치(indicators)로 사용하는 잠재 프로파일 분석이나 요인 혼합 모델링으로 자료에 포함된 IER의 서로 다른 유형들을 파악할 수도 있다(Meade & Craig, 2012).

#### IER 자료의 처리

앞의 세 수준의 IER 자료 탐지 과정을 거쳐 IER로 최종 판정된 자료를 처리하는 방법도 세 가지로 구분할 수 있다(Arthuer et al., 2021; Ward & Meade, 2023).

첫째, IER로 확인된 응답자들을 자료에서

완전히 제거하는 것이다. 이 경우 연구 설계나 특성을 고려하면서 다양한 유형의 IER 응답자들을 걸러내기 위해 복수의 방법을 사용하는 것이 좋다. 이렇게 하면 불필요한 오차 변량 및 1, 2중 오류를 줄이고, 측정의 내적 일관성, 요인 구조, 검정력, 구성 타당도 등의 측면에서 긍정적 효과를 가질 수 있다.

둘째, IER을 보인 응답자에 대해 통계적 통제(statistical control)를 하는 것이다. 이렇게 하면 자료 손실 없이 전체 자료를 사용할 수 있고, IER이 관심 연구 질문에 미치는 영향도 경험적으로 조사할 수도 있다. 그러나 이 방법이 갖는 효과는 아직 명확하지 않다는 한계가 있어 유의가 필요하다.

셋째, 표본 크기가 작을 때의 대안으로, IER 응답 자료를 완전히 제거하지 않고 개별 문항(또는 설문의 특정 부분)에 대한 불성실 응답을 결측(missing)치로 처리하는 것이다. 이렇게 함으로써 전체 자료를 다 사용할 수 있게 된다.

#### IER 응답자 제거 관련 보고.

Ward와 Meade(2023)은 앞으로 경험적 논문에서는 IER 관련 대처 절차에 대해 보고할 필요가 있다고 하였다. 즉, 데이터 클리닝의 확장으로서, 논문의 ‘방법’ 부분에서 IER을 어떻게 예방, 탐지, 및 제거하였는지를 기술할 것을 제안하였다. 또한, 통계적 검정력을 유지하는 것이 중요하다면, IER을 제거한 경우와 제거하지 않은 경우의 분석을 별도로 할 것을 제안하였다. 이것은 IER이 자료에 미치는 영향을 결정하고, 두 분석 결과 간의 의미 있는 차이를 보고하고, IER을 보인 응답자들을 제거하지 않은 것에 대해 정당화하는 것이다.

#### 향후 연구 제안

지금까지 IER과 관련된 연구들을 기저율, 발생 원인 및 조건, 분석 결과에 미치는 영향, 효과적인 탐지 방법, 예방 방안 및 처리 등의 주제를 중심으로 포괄적으로 개관하였다. 오늘날 클라우드소싱된 온라인 조사가 늘어나면서 사회조사나 심리측정에서 이제 더는 IER 문제가 소홀하게 다룰 수 없는 우려 사항으로 떠오르고 있다. 그러나 IER과 관련 주요한 의사결정에 있어 분명한 답을 찾은 경우보다 여전히 모호하고 명확한 답을 찾지 못한 이슈들이 더 많은 것 같다. 예를 들어, IER의 발생 원인과 관련된 개인차 변인에 관한 연구, 상황이나 맥락 요소, 문화 간 차이 등에 대해서는 더 많은 연구가 필요하다. 또한, 연구 설계나 특성에 따른 최상의 IER 탐지 방법에 대한 조건(예, 가짜 문항이나 지시 응답 문항을 사용한다면 몇 문항이나 사용해야 하는지)이나 IER 응답자로 분류하기 위한 역치 또는 컷오프 수치에 대한 판단 기준(복수의 가짜 문항에서 몇 개의 잘못된 응답을 IER로 간주할지) 등은 연구마다 다양하게 제시된다. IER 발생의 기저율의 경우에도 무시해도 괜찮을 정도로 발생한다는 시각과 연구 결과를 왜곡할 정도로 많이 발생한다는 시각이 공존하고 있다. 무엇보다 IER을 사전에 방지하는 예방 조치들에 대한 연구들도 부족하다. 그간 이루어진 연구들에서 주목할 만한 효과를 보인 경우가 많지 않다는 점에서 더 많은 창의적인 시도가 요구된다.

지금까지 논의한 미래 필요한 연구들과 더불어, IER 관련 향후 연구에서는 1) IER에서의 문화 간 차이와 2) IER 관련 우려를 새로운 측정 도구나 테크놀로지를 사용하여 해결하려

는 이슈들도 다루어야 할 것이다.

첫째, 문화 간 또는 국가 간 비교연구에서 자기 보고 형태의 조사 연구를 수행할 때 IER이 어떤 양상으로 다르게 나타나는지를 살펴볼 필요가 있다. 일반적 반응 양식(response style)에서 문화 간(또는 국가 간) 차이를 보인 연구들은 많이 있다. 예를 들어, He 등(2014)은 사회적 바람직성 반응, 묵인 반응, 극단 및 중간점 반응 등의 여러 반응 양식<sup>28)</sup>이 나타난 정도는 인간 개발 지표(HDI)<sup>29)</sup>와 개인주의(individualism) 간 부적 관련성이 있음을 발견했다. 또한, 묵인 반응 양식은 권력 거리가 높은 나라에서 더 흔하게 나타났다(Smith, 2004). 일본 관리자들은 자기 보고형 심리 척도에서 미국 및 영국 관리자들보다 중간점을 선택하는 경향이 더 강하였다(Stening & Everett, 1984). 대만인들과 미국 및 캐나다인들을 비교한 연구에서도 유사한 결과가 나타났다(Lee et al., 2002). 호주와 중국 간 주관적 안녕감에서의 차이를 반응 편향에서의 차이로 귀인한 Lai 등(2013)의 연구에서도 동아시아인들은 심리평가 척도의 양극단보다는 중간점을 선택하

는 경향이 나타났다. 여러 연구에서 나타난 이러한 경향에 대해 Lu(2001)는 겸손하고 삼가는 것이 중시되는 유교, 도교, 불교 등에 기반한 동아시아 문화에 그 원인이 있는 것으로 보았다.

IER에서의 문화 간 차이와 관련하여, Maniaci와 Rogge(2014)는 설문 문항에 답할 때 비코커서스인이 덜 주의를 기울이고 덜 순종적이라고 하였다. Litman 등(2015)은 미국과 인도에서 크라우드소싱된 자료에서의 차이를 탐색하였는데, 인도에서 더 데이터 질이 낮았으며, 온라인 조사에 대한 보상이 주 수입원인 경우 더 그러하였다. Grau 등(2019)은 앞서 언급한 He 등의 문화 차 연구를 IER에 확대하였다. 그들은 34개국의 8,320명의 표본으로 단축형 Big 5 성격 설문 문항에 대한 반응에서의 IER을 조사하였다. 롱스트링, 마할라노비스 거리, 심리측정적 동의어와 반의어, 개인-전체 상관관계의 다섯 가지 IER 지표<sup>30)</sup>에서, 교육 수준과 성격 특성이 IER과 연관성이 있으며, 인간 개발 지표, 개인주의, 젠더 불평등(gender inequality), 권력 거리<sup>31)</sup>의 문화 차원들이 IER과 강한 상관관계를 보였다. IER과 반응 양식(극단 반응, 중간 반응, 묵인 반응, 사회적 바람직성 반응) 간의 비교에서는 둘이 어느 정도 중복되는 부분이 있는 것으로 나타나지만, 서로 다른 요인으로 구분되었다. 즉, IER은 극단 반응, 중간 반응, 묵인 반응과 일치하는 양

28) 반응 양식(response styles)은 응답자가 문항의 내용과는 무관하게 응답하는 것으로 정의되며 (Paulhus, 1991), 단순히 무작위 응답 과정으로 보기보다는 나이, 성별, 사회적 지위와 관련이 있는 성격 특성으로 보는 경향이 강하다. IER이 동기 부족으로 발생한다면, 반응 양식은 특정하게 자신을 드러내려는 동기를 가지고 반응하는 것일 수 있다. 그러나 두 가지 모두는 문항 내용과 상관없이 반응하는 공통점이 있다. 또한, IER, 묵인 반응, 극단(중간점) 반응은 문항을 자세히 읽지 않지만, 사회적 바람직성 반응을 위해서는 문항을 자세히 읽어야 하는 점에서 차이가 있다.

29) HDI는 기대 수명, 교육 수준(문맹), 1인당 소득의 세 차원으로 구성된 혼합 지표이다.

30) 5개 지표에 대한 요인분석에서 롱스트링 요인만이 1요인 해에서 .50 이하의 부하량을 보였다. 2요인 해에서도 롱스트링만이 두 번째 요인으로 나타났다.

31) 서로 (정적 및 부적) 상관이 높은 이 네 차원은 사회학적 용어로 ‘해방(emancipation)’요인으로 불린다(Ebbeler et al., 2017). 또한, 사회의 민주화나 개방성의 정도를 나타낸다.

상을 보였지만, 사회적 바람직성 반응과는 구분되었다. 이는 사회적 바람직성 반응만이 문항을 자세히 읽고 응답해야 하는 반응 양식이기 때문이다.

앞서 RE 지표를 소개한 Tawa(2021)는 IER 탐지 지표가 문화적 가치에 기반한 반응 양식과 부주의 응답을 혼동할 가능성을 조사하였다. 이를 위해, 다양한 인종 집단(백인 159명, 흑인 78명, 아시아 59명)을 대상으로 한 설문 자료에서 인종과 IER 지표 간 관계를 반응 양식을 매개 변수로 하여 회귀 분석하였다. 그 결과, 아시아인은 백인에 비해 묵인(acquiescent) 응답을 더 많이 하였고, 흑인 응답자들은 백인에 비해 중간점(midpoint) 응답을 덜 하는 경향이 나타났다. 이러한 차이는 마할라노비스 거리, 심리측정적 동의어, 개인 응답 변량, 짝-홀수 지표 등 IER 지표에도 유의한 영향을 미치는 것으로 나타났다. 즉, IER 지표의 인종간 차이는 반응 양식에 의해 매개되는 것으로 나타났다. 그러나 RE 지표는 다른 IER 지표들과는 달리 인종 간 차이가 관찰되지 않았으며, 이것은 RE가 문화적 혼동(cultural confound)의 가능성이 낮음을 나타낸다. 따라서 Tawa(2021)는 RE 지표가 다양한 사회문화적 집단 간 반응 양식의 문화 차이를 IER로 잘못 판단하게 되는 오류를 줄이면서 동시에 IER을 더 정확하게 식별할 수 있는 방법으로 보았다.

한편 Magraw-Mickelson 등(2022)은 자료수집의 방식(지필, 컴퓨터/웹, 스마트폰)이 참가자들의 IER 경향에 미치는 영향을 여러 나라에 걸쳐 조사하였다. 그들은 이전 연구들이 여러 자료수집 방식에서 IER이 비슷하게 나타났다고 보지만, 테크놀로지가 계속 발전하고 매우 다양한 전집으로부터 자료가 수집되고 있기에

재조사가 필요하다고 보았다. IER에 미치는 자료수집 방식의 효과를 서로 다른 세 표본을 반복 측정된 자료로 분석한 결과, 중국 직장인 표본에서 지필 방식보다 컴퓨터/웹 기반 방식에서 약간 더 부주의한 반응이 나타났다. 또한, 독일 학생 표본에서는 스마트폰 방식 대비 지필 양식에서 약간 더 부주의한 반응을 보였다. 중국 학생 표본은 방식 간 아무런 차이를 보이지 않았다. 전반적으로는 문화(국가) 및 조사 방식에 따른 IER의 차이는 크지 않은 것으로 나타났다.

이처럼 IER에서의 문화(국가) 간 차이를 일부 연구들이 상이한 관점에서 단편적으로 진행하고 있지만, 좀 더 이론적 기반 위에서 통합적으로 접근하는 연구가 필요하다.

둘째, 최근 심리학자들은 자기보고형 검사와 같은 외현적(explicit) 측정치와 더불어 암묵적인 측정(implicit measure) 방식을 함께 사용하는 것을 제안하고 있다(Lang et al., 2012). 대부분의 암묵적 측정치에는 불성실한 응답자를 바로 찾아낼 수 있는 장치(기준)가 내재되어 있기에 온라인 패널이나 학생 표본에서 IER을 신속하게 포착해 데이터에서 제외할 수 있다.

소셜 미디어 자료를 활용하는 것과 같은 새로운 도구나 자료 출처를 사용하는 방안도 고려할 수 있을 것이다. 자기 보고에 의존하지 않고, 소셜 미디어 자료와 같은 디지털 족적에서 기존 심리학적 구성개념들을 측정할 수 있음을 보여준 연구들이 많이 있다. 예를 들어, 페이스북의 “좋아요(Likes)”의 사용 패턴을 분석하여 성격을 예측하였는데, 전통적인 성격 5요인 척도인 IPIP 척도의 점수를 비교적 정확하게 예측하였다(Youyou et al., 2015). 디지털 기록이 인간 특성의 지표라고 보는 이러한 분석은 온라인 환경에서의 사람들의 행동



도 자신의 태도, 선호, 흥미, 행동 경향성을 반영한다는 원리에 근거한다(Back et al., 2010).

또한, 온라인 조사가 늘어나고 수입원으로 써 조사에 참여하는 이들이 늘어나면서 인간이 답하지 않고 로봇이나 컴퓨터 스크립트를 활용하는 경우가 늘고 있고, 한 사람이 동일 설문에 여러 번 반복해서 응답하는 경우도 나타나고 있다. 이런 경우들은 모두 IER의 가능성과 유사한 부정적 효과를 가질 것이다. 이에 대응하여 향후 인공지능(AI)이나 기계학습이 실시간 IER 탐지에 활용될 잠재력이 충분히 있다. 이는 IER 응답 패턴을 학습한 기계가 중간에 경고를 주면서 실시간 감독이나 감시 효과를 가질 것이다. 이때 훈련 데이터가 매우 중요할 것인데, 예를 들어, 고부담 상황에서 수집된 자료로 학습된 인공지능 알고리즘을 저부담 상황에서 사용하는 것은 부적절할 것이다.

### 참고문헌

김우영 (2020). 온라인 설문에서의 불성실 응답 방지 방법에 대한 연구: 객관적 자기인식이론의 적용. 서강대학교 박사학위 청구논문. UCI I804:11029-000000064932

김우영, 이태현, 장재윤 (2019). 온라인 설문에서 거울 이미지를 활용한 불성실 응답 방지 방법에 대한 연구. *한국심리학회지: 일반*, 38(4), 669-698.  
<https://doi.org/10.22257/kjp.2019.12.38.4.669>

박원우, 마성혁, 배수현, 지선영, 이유우, 김자영 (2020). 설문조사에서 불성실 응답의 탐지방법과 제거의 효과. *경영학연구*, 49(2), 331-364.

<https://doi.org/10.17287/kmr.2020.49.2.331>

한신갑 (2015). 빅 데이터와 사회과학하기: 자료기반의 변화와 분석전략의 재구상. *한국사회학*, 49(2), 161-192.

<https://hdl.handle.net/10371/94779>

Arias, V. B., Garrido, L. E., Jenaro, C., Martínez-Molina, A., & Arias, B. (2020). A little garbage in, lots of garbage out: Assessing the impact of careless responding in personality survey data. *Behavior Research Methods*, 52(6), 2489-2505.  
<https://doi.org/10.3758/s13428-020-01401-8>

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.  
<https://doi.org/10.2105/AJPH.81.12.1636>

Arthur Jr, W., Hagen, E., & George Jr, F. (2021). The lazy or dishonest respondent: Detection and prevention. *Annual Review of Organizational Psychology and Organizational Behavior*, 8, 105-137.  
<https://doi.org/10.1146/annurev-orgpsych-012420-055324>

Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological Science*, 21, 372-374.  
<https://doi.org/10.1177/0956797609360756>

Bäckström, M., & Björklund, F. (2019). Is reliability compromised towards the end of long personality inventories? *European Journal of Psychological Assessment*, 35, 14-23.

- <https://doi.org/10.1027/1015-5759/a000363>  
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.  
[https://doi.org/10.1207/s15327752jpa6801\\_11](https://doi.org/10.1207/s15327752jpa6801_11)
- Barber, L. K., Barnes, C. M., & Carlson, K. D. (2013). Random and systematic error effects of insomnia on survey behavior. *Organizational Research Methods*, 16(4), 616-649.  
<https://doi.org/10.1177/1094428113493120>
- 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(6), 1014-1027.  
<https://doi.org/10.1037/0022-3514.81.6.1014>
- 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.  
<https://doi.org/10.1037/0022-3514.37.10.1835>
- 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.  
<https://doi.org/10.1037/1040-3590.4.3.340>
- Bosco, F.A., Aguinis, H., Singh, K., Field, J.G., & Pierce, C.A. (2015). Correlational effect size benchmarks. *Journal of Applied Psychology*, 100, 431-449.  
<https://doi.org/10.1037/1040-3590.4.3.340>
- Bowling, N. A., Gibson, A. M., & DeSimone, J. A. (2022). Stop with the questions already! Does data quality suffer for scales positioned near the end of a lengthy questionnaire? *Journal of Business and Psychology*. Advance online publication.  
<https://doi.org/10.1007/s10869-021-09787-8>
- Bowling, N. A., Gibson, A. M., Houpt, J. W., & Brower, C. K. (2021). Will the questions ever end? Person-level increases in careless responding during questionnaire completion. *Organizational Research Methods*, 24(4), 718-738.  
<https://doi.org/10.1177/1094428120947794>
- 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.  
<https://doi.org/10.1037/pspp0000085>
- Bowling, N. A., Huang, J. L., Brower, C. K., & Bragg, C. B. (2023). The quick and the careless: The construct validity of page time as a measure of insufficient effort responding to surveys. *Organizational Research Methods*, 26(2), 323-352.  
<https://doi.org/10.1177/109442812111056520>
- Bruehl, 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, 3-9.  
<https://doi.org/10.1037/1040-3590.10.1.3>
- Brühlmann, F., Petralito, S., Aeschbach, L. F., &

- Opwis, K. (2020). The quality of data collected online: An investigation of careless responding in a crowdsourced sample. *Methods in Psychology*, 2, 1-13.  
<https://doi.org/10.1016/j.metip.2020.100022>
- 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, 3-5.  
<https://doi.org/10.1177/1745691610393980>
- Buechley, R., & Ball, H. (1952). A new test of "validity" for the group MMPI. *Journal of Consulting Psychology*, 16(4), 299-301.  
<https://doi.org/10.1037/h0053897>
- 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.  
<https://doi.org/10.1037/1040-3590.15.2.223>
- Credé, M. (2010). Random responding as a threat to the validity of effect size estimates in correlational research. *Educational and Psychological Measurement*, 70, 596-612.  
<https://doi.org/10.1177/0013164410366686>
- Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*, 66, 4-19.  
<https://doi.org/10.1016/j.jesp.2015.07.006>
- Curran, P. G., & Hauser, K. A. (2019). I'm paid biweekly, just not by leprechauns: Evaluating valid-but-incorrect response rates to attention check items. *Journal of Research in Personality*, 82, 103849.  
<https://doi.org/10.1016/j.jrp.2019.103849>
- 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, 309-338.  
<https://doi.org/10.1111/apps.12117>
- Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *PLoS ONE*, 18(3), e0279720.  
<https://doi.org/10.1371/journal.pone.0279720>
- 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, 105-121.  
<https://doi.org/10.1007/s10869-016-9479-0>
- Duval, S., & Wicklund, R. A. (1972). *A theory of objective self awareness*. New York: Academic Press.
- Ebbeler, C., Grau, I., & Banse, R. (2017). Cultural and individual factors determine physical aggression between married partners: Evidence from 34 countries. *Journal of Cross-Cultural Psychology*, 48(7), 1098-1118.  
<https://doi.org/10.1177/0022022117719497>

- Evans, R. G., & Dinning, W. D. (1983). Response consistency among high F scale scorers on the MMPI. *Journal of Clinical Psychology, 39*, 246-248.  
[https://doi.org/10.1002/1097-4679\(198303\)39:2<246::AID-JCLP2270390217>3.0.CO;2-9](https://doi.org/10.1002/1097-4679(198303)39:2<246::AID-JCLP2270390217>3.0.CO;2-9)
- Francavilla, N., Meade, A., & Young, A. (2019). Social interaction and internet-based surveys: Examining the effects of virtual and in-person proctors on careless response. *Applied Psychology, 68*(2), 223-249.  
<https://doi.org/10.1111/apps.12159>
- Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly, 73*(2), 349-360.  
<https://doi.org/10.1093/poq/nfp031>
- Gibson, A. M., & Bowling, N. A. (2020). The effects of questionnaire length and behavioral consequences on careless responding. *European Journal of Psychological Assessment, 36*(2), 410-420.  
<https://doi.org/10.1027/1015-5759/a000526>
- Glas, C. A. W., & Dagohoy, A. V. T. (2007). A person fit test for IRT models for polytomous items. *Psychometrika, 72*, 159-180.  
<https://doi.org/10.1007/s11336-003-1081-5>
- Goldammer, P., Annen, H., Stöckli, P. L., & Jonas, K. (2020). Careless responding in questionnaire measures: Detection, impact, and remedies. *The Leadership Quarterly, 31*(4), Article 101384.  
<https://doi.org/10.1016/j.leaqua.2020.101384>
- 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, 26*, 213-224.  
<https://doi.org/10.1002/bdm.1753>
- Grau, I., Ebbeler, C., & Banse, R. (2019). Cultural differences in careless responding. *Journal of Cross-Cultural Psychology, 50*(3), 336-357.  
<https://doi.org/10.1177/0022022119827379>
- Greene, R. L. (1978). An empirically derived MMPI carelessness scale. *Journal of Clinical Psychology, 34*, 407-410.  
[https://doi.org/10.1002/1097-4679\(197804\)34:2<407::AID-JCLP2270340231>3.0.CO;2-A](https://doi.org/10.1002/1097-4679(197804)34:2<407::AID-JCLP2270340231>3.0.CO;2-A)
- 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*, 407-412.  
[https://doi.org/10.1002/1097-4679\(196310\)19:4<407::AID-JCLP2270190410>3.0.CO;2-N](https://doi.org/10.1002/1097-4679(196310)19:4<407::AID-JCLP2270190410>3.0.CO;2-N)
- Hasselhorn, K., Ottenstein, C., & Lischetzke, T. (2023). Modeling careless responding in ambulatory assessment studies using multilevel latent class analysis: Factors influencing careless responding. *Psychological Methods*. Advance online publication.  
<https://doi.org/10.1037/met0000580>
- He, J., Van de Vijver, F. J. R., Dominguez Espinosa, A., & Mui, P. H. C. (2014). Toward a unification of acquiescent, extreme, and midpoint response styles: A multilevel study. *International Journal of Cross Cultural Management, 14*, 306-322.

- <https://doi.org/10.1177/1470595814541424>  
Herzog, A. R., & Bachman, J. G. (1981). Effects of questionnaire length on response quality. *Public Opinion Quarterly*, 45(4), 549-559.  
<https://doi.org/10.1086/268687>
- Huang, J. L., Bowling, N. A., Liu, M., & Li, Y. (2015). Detecting insufficient effort responding with an infrequency scale: Evaluating validity and participant reactions. *Journal of Business and Psychology*, 30, 299-311.  
<https://doi.org/10.1007/s10869-014-9357-6>
- 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.  
<https://doi.org/10.1007/s10869-011-9231-8>
- Huang, J. L., Liu, M., & Bowling, N. A. (2015). Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology*, 100, 828-845.  
<https://doi.org/10.1037/a0038510>
- Jones, A., Earnest, J., Adam, M., Clarke, R., Yates, J., & Pennington, C. R. (2022). Careless responding in crowdsourced alcohol research: A systematic review and meta-analysis of practices and prevalence. *Experimental and Clinical Psychopharmacology*, 30(4), 381-399.  
<https://doi.org/10.1037/pha0000546>
- Johnson, J. A. (2005). Ascertaining the validity of individual protocols from web-based personality inventories. *Journal of Research in Personality*, 39, 103-129.  
<https://doi.org/10.1016/j.jrp.2004.09.009>
- Kam, C. C. S. (2019). Careless responding threatens factorial analytic results and construct validity of personality measure. *Frontiers in Psychology*, 10, 1258.  
<https://doi.org/10.3389/fpsyg.2019.01258>
- Kam, C. C. S., & Meyer, J. (2015). How careless responding and acquiescence response bias can influence construct dimensionality. *Organizational Research Methods*, 18(3), 512-541.  
<https://doi.org/10.1177/1094428115571894>
- Kam, C. C. S., & Chan, G. H. H. (2018). Examination of the validity of instructed response items in identifying careless respondents. *Personality and Individual Differences*, 129, 83-87.  
<https://doi.org/10.1016/j.paid.2018.03.022>
- Kam, C. C. S., & Cheung, S. F. (2023). A constrained factor mixture model for detecting careless responses that is simple to implement. *Organizational Research Methods*, Advance online publication.  
<https://doi.org/10.1177/10944281231195298>
- 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.  
<https://doi.org/10.1037/0033-2909.119.2.254>
- Kupffer, R., Frick, S., & Wetzel, E. (2024). Detecting careless responding in multidimensional forced-choice questionnaires. *Educational and Psychological Measurement*. Advance online publication.  
<https://doi.org/10.1177/00131644231222420>
- Kurtz, J. E., & Parrish, C. L. (2001). Semantic

- response consistency and protocol validity in structured personality assessment: The case of the NEO-PI-R. *Journal of Personality Assessment*, 76, 315-332.  
[https://doi.org/10.1207/S15327752JPA7602\\_12](https://doi.org/10.1207/S15327752JPA7602_12)
- Lai, L. C., Cummins, R. A., & Lau, A. L. (2013). Cross-cultural difference in subjective wellbeing: Cultural response bias as an explanation. *Social Indicators Research*, 114, 607-619.  
<https://doi.org/10.1007/s11205-012-0164-z>
- Lang, J. W., Zettler, I., Ewen, C., & Hülshager, U. R. (2012). Implicit motives, explicit traits, and task and contextual performance at work. *Journal of Applied Psychology*, 97, 1201-1217.  
<https://doi.org/10.1037/a0029556>
- Lee, J. W., Jones, P. S., Mineyama, Y., & Zhang, X. E. (2002). Cultural differences in responses to a Likert scale. *Research in Nursing & Health*, 25(4), 295-306.  
<https://doi.org/10.1002/nur.10041>
- Litman, L., Robinson, J., & Rosenzweig, C. (2015). The relationship between motivation, monetary compensation, and data quality among US-and India-based workers on Mechanical Turk. *Behavior Research Methods*, 47(2), 519-528.  
<https://doi.org/10.3758/s13428-014-0483-x>
- 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.
- Lu, L. (2001). Understanding happiness: A look into the Chinese folk psychology. *Journal of Happiness Studies*, 2, 407-432.  
<https://doi.org/10.1023/A:1013944228205>
- Magraw-Mickelson, Z., Wang, H. H., & Gollwitzer, M. (2022). Survey mode and data quality: Careless responding across three modes in cross-cultural contexts. *International Journal of Testing*, 22(2), 121-153.  
<https://doi.org/10.1080/15305058.2021.2019747>
- 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.  
<https://doi.org/10.1016/j.jrp.2013.09.008>
- Marjanovic, Z., Holden, R., Struthers, W., Cribbie, R., & Greenglass, E. (2015). The inter-item standard deviation (ISD): An index that discriminates between conscientious and random responders. *Personality and Individual Differences*, 84, 79-83.  
<https://doi.org/10.1016/j.paid.2014.08.021>
- Marston, W. M. (1917). Systolic blood pressure symptoms of deception. *Journal of Experimental Psychology*, 2, 117-163.  
<https://doi.org/10.1037/h0073583>
- 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, 450-470. <https://doi.org/10.1037/a0019216>
- McLachlan, G., & Peel, D. (2004). *Finite mixture models*. New York, NY: John Wiley & Sons.
- Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological*

- Methods, 17, 437-455.  
<https://doi.org/10.1037/a0028085>
- 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.  
<https://doi.org/10.1037/0021-9010.88.2.348>
- Nichols, D. S., Greene, R. L., & Schmolck, P. (1989). Criteria for assessing inconsistent patterns of item endorsement on the MMPI: Rationale, development, and empirical trials. *Journal of Clinical Psychology, 45*, 239-250.  
[https://doi.org/10.1002/1097-4679\(198903\)45:2<239::AID-JCLP2270450210>3.0.CO;2-1](https://doi.org/10.1002/1097-4679(198903)45:2<239::AID-JCLP2270450210>3.0.CO;2-1)
- 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*, 1076-1086.  
<https://doi.org/10.1037/a0023221>
- 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.  
<https://doi.org/10.1037/0021-9010.81.6.660>
- Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*, 867-872.  
<https://doi.org/10.1016/j.jesp.2009.03.009>
- Osborne, J. W., & Blanchard, M. R. (2011). Random responding from participants is a threat to the validity of social science research results. *Frontiers in Psychology, 1*, 220.  
<https://doi.org/10.3389/fpsyg.2010.00220>
- Palaniappan, K., & Kum, I. Y. S. (2019). Underlying causes behind research study participants' careless and biased responses in the field of sciences. *Current Psychology, 38*(6), 1737-1747.  
<https://doi.org/10.1007/s12144-017-9733-2>
- Patton, J. M., Cheng, Y., Hong, M., & Diao, Q. (2019). Detection and treatment of careless responses to improve item parameter estimation. *Journal of Educational and Behavioral Statistics, 44*(3), 309-341.  
<https://doi.org/10.3102/1076998618825116>
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson & P. R. Shaver & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). San Diego, CA: Academic Press.  
<https://doi.org/10.1016/B978-0-12-590241-0.50006-X>
- Perdue, C. W., & Gurtman, M. B. (1990). Evidence for the automaticity of ageism. *Journal of Experimental Social Psychology, 26*(3), 199-216.  
[https://doi.org/10.1016/0022-1031\(90\)90035-K](https://doi.org/10.1016/0022-1031(90)90035-K)
- Piferi, R. L., & Jobe, R. L. (2003). Identifying careless response in questionnaire research: A serendipitous finding. *Psychological Reports, 93*, 320-322.  
<https://doi.org/10.2466/pr0.2003.93.1.320>
- Pinsonneault, T. B. (2005). Detecting random, partially random, and nonrandom Minnesota Multiphasic Personality Inventory-Adolescent

- Protocols. *Psychological Assessment*, 17, 476-480.  
<https://doi.org/10.1037/1040-3590.17.4.476>
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879-903.  
<https://doi.org/10.1037/0021-9010.88.5.879>
- Ran, S., Liu, M., Marchiondo, L. A., & Huang, J. L. (2015). Difference in response effort across sample types: Perception or reality? *Industrial and Organizational Psychology*, 8, 202-208.  
<https://doi.org/10.1017/iop.2015.26>
- Reimers, J. A., Turner, R. C., Crawford, B. L., Jozkowski, K. N., Lo, W. J., & Keiffer, E. A. (2022). Demographic comparisons on data quality measures in web-based surveys. *Personality and Individual Differences*, 193, 111612.  
<https://doi.org/10.1016/j.paid.2022.111612>
- Schneider, S., May, M., & Stone, A. A. (2018). Careless responding in internet-based quality of life assessments. *Quality of Life Research*, 27, 1077-1088.  
<https://doi.org/10.1007/s11136-017-1767-2>
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, 54, 93-105.  
<https://doi.org/10.1037/0003-066X.54.2.93>
- Shukla, K., & Konold, T. (2018). A two-step latent profile method for identifying invalid respondents in self-reported survey data. *The Journal of Experimental Education*, 86(3), 473-488.  
<https://doi.org/10.1080/00220973.2017.1315713>
- Smith, P. B. (2004). Acquiescent response bias as an aspect of cultural communication style. *Journal of Cross-Cultural Psychology*, 35, 50-61.  
<https://doi.org/10.1177/0022022103260380>
- Steinmann, I., Strietholt, R., & Braeken, J. (2022). A constrained factor mixture analysis model for consistent and inconsistent respondents to mixed-worded scales. *Psychological Methods*, 27(4), 667-702.  
<https://doi.org/10.1037/met0000392>
- Stening, B. W., & Everett, J. E. (1984). Response styles in a cross-cultural managerial study. *Journal of Social Psychology*, 122(2), 151-156.  
<https://doi.org/10.1080/00224545.1984.9713475>
- Tawa, J. (2021). The response entropy index: Comparative assessment of performance and cultural bias across indices of careless responding. *Survey Research Methods*, 15(3), 299-325.
- Toich, M. J., Schutt, E., & Fisher, D. M. (2022). Do you get what you pay for? Preventing insufficient effort responding in MTurk and student samples. *Applied Psychology*, 71(2), 640-661.  
<https://doi.org/10.1111/apps.12344>
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. New York: Cambridge University Press.  
<https://doi.org/10.1017/CBO9780511819322>
- Wang, M. D., & Hau, K. T. (2024). Flagging insufficient effort responses in surveys: Stopping rule to prevent insufficient or



- excessive removal of doubtful data. *Acta Psychologica*, 243, 104135.  
<https://doi.org/10.1016/j.actpsy.2024.104135>
- Ward, M. K., & Meade, A. W. (2018). Applying social psychology to prevent careless responding during online surveys. *Applied Psychology*, 67, 231-263.  
<https://doi.org/10.1111/apps.12118>
- Ward, M. K., & Meade, A. W. (2023). Dealing with careless responding in survey data: Prevention, identification, and recommended best practices. *Annual Review of Psychology*, 74, 577-596.  
<https://doi.org/10.1146/annurev-psych-040422-045007>
- Ward, M. K., Meade, A. W., Allred, C., Pappalardo, G., & Stoughton, J. (2017). Careless response and attrition as sources of bias in online survey assessments of personality traits and performance. *Computers in Human Behavior*, 76, 417-430.  
<https://doi.org/10.1016/j.chb.2017.06.032>
- 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.  
<https://doi.org/10.1016/j.chb.2015.01.070>
- Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment*, 28, 186-191.  
<https://doi.org/10.1007/s10862-005-9004-7>
- Yentes, R. D. (2020). In search of best practices for the identification and removal of careless responders. PhD Thesis, N.C. State Univ., Raleigh.
- Yentes, R. D., & Wilhelm, F. (2023). careless: Procedures for computing indices of careless responding. R package version 1.2.2. <https://cran.r-project.org/web/packages/careless/index.html>.
- Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, 112, 1036-1040.  
<https://doi.org/10.1073/pnas.1418680112>

1차원고접수 : 2024. 02. 23

2차원고접수 : 2024. 05. 21

최종게재결정 : 2024. 06. 06

## The Causes, Detection, and Prevention of Insufficient Effort Responding(IER) in Self-Report Measures: A Comprehensive Literature Review and Suggestions

Jae Yoon Chang

Department of Psychology  
Sogang University

Woo Young Kim

SK Innovation

Taehun Lee

School of Psychology  
Korea University

This paper offers a comprehensive and critical review of Insufficient Effort Responding (IER) in social surveys and psychological measurements, a concern that has garnered increasing attention. It aims to provide researchers and practitioners with practical guidance by delving into IER's causes, impacts, detection strategies, and preventive measures. Our review underscores that IER stems from both individual difference factors—such as personality traits, motivation levels, and cognitive abilities—and situational factors, including survey length, data collection methods, environmental distractions, lack of interaction between researchers and participants, and cultural variances. These elements significantly undermine measurement reliability and validity, leading to heightened errors in hypothesis testing and negatively impacting the outcomes of research and surveys. To counteract these adverse effects, the paper outlines several methodological strategies for the effective prevention and post hoc detection of IER, with the goal of enhancing both data quality and research reliability. The authors advocate for future research directions focused on 1) exploring the IER issues by accounting for cultural differences and 2) leveraging new measurement technologies to deepen the understanding and management of IER in research contexts.

*Key words* : *Insufficient Effort Responding, Careless Responding, Response Bias, Self-report, Social Survey, Psychological Measurement*