

설문조사에서 불성실 응답의 탐지방법과 제거의 효과*

박원우(주저자)

서울대학교
(vwpark@snu.ac.kr)

마성혁(교신저자)

서울대학교
(ryanshmah@snu.ac.kr)

배수현(공저자)

서울대학교
(top3043@snu.ac.kr)

지선영(공저자)

서울대학교
(seonyoungji@gmail.com)

이우우(공저자)

서울대학교
(youfmc@snu.ac.kr)

김자영(공저자)

서울대학교
(marsb612@snu.ac.kr)

.....

본 연구는 설문조사에서 발생하는 불성실 응답(careless responding)을 탐지하는 방법과 이를 제거함으로써 얻을 수 있는 효과에 대해 논의한다. 우선, 본 연구는 설문조사에서 발생하는 다양한 불성실 응답의 개념과 유형을 정리하여 설명하였다. 이어서, 최근 활발히 연구가 진행되고 있는 불성실 응답의 다양한 탐지방법들을 직접적, 비개입적, 통계적 방법으로 나누어 소개하였다. 특히, 향후 연구자들이 자신의 연구 맥락에 맞추어 이를 활용할 수 있도록 각 방법이 지닌 의미와 장단점, 그리고 탐지 기준을 비교하여 안내하였다. 나아가 본 연구에서는 소개된 다수의 불성실 응답 탐지방법 및 제거의 효과를 확인하기 위해, 국내 대기업 구성원 3,030명을 대상으로 한 설문조사 결과를 활용하여 별도의 실증분석을 실시하였다. 여기에는 총 다섯 가지 불성실 응답 탐지방법이 적용되었으며, 각 방법에 의하여 탐지된 불성실 응답을 제거하기 전·후의 통계적 결과를 비교하는 방식으로 효과를 검증하였다. 실증연구를 통해 확인한 주요 결론은 다음과 같다. 첫째, 탐지에 활용된 방법에 따라, 불성실 응답으로 분류되는 응답자의 숫자와 비율이 달라졌다. 개별 탐지방법에 따라 약 0.5%~14%의 답변이 불성실 응답으로 분류되는 것으로 나타났다. 둘째, 불성실 응답의 제거가 통계적 추론 결과에 미치는 영향을 확인한 결과, 불성실 응답의 존재가 상관관계, 요인분석, 회귀분석 결과에 미치는 영향은 기존 문헌의 결과와 유사한 수준으로, 연구 가설의 결과를 왜곡시킬 정도로 크지 않은 것으로 나타났다. 셋째, 2개 이상의 탐지방법을 조합하는 다중 허들방식(multi-hurdle)을 적용하여 불성실 응답을 탐지한 예시를 보여주었다. 마지막으로 본 연구에서는 문헌고찰과 실증분석 결과를 바탕으로, 연구자들이 설문조사에서 활용할 수 있는 불성실 응답 탐지방법과 향후 관련 연구 방향을 제시하였다.

주제어: 연구방법론, 설문조사, 불성실 응답(careless responding)

.....

1. 서론

설문조사는 경영학을 포함한 사회과학 분야 전반에서 개인과 조직 특성을 측정하고 평가하기 위해 사용되는 연구 방법이다. 설문에 참여하는 응답자는 질문의 의미를 해석하고, 관련 정보를 기억에서 불러내어 통합하여 필요한 판단을 내려 적절한 답변을 선택하는 과정을 거치게 된다(Cannell, Miller, & Oksenberg, 1981). 응답자가 연구자의 의도대로 질문을 이해하여 정확하고 충실한 응답을 선택하기 위해서는 상당한 인지적 노력이 필요하기에, 간혹 설문에 참여한 응답자들은 최적의 응답을 제공하지 않을 수도 있다(Krosnick, 1991). 예를 들어, 응답자는 설문 참여 시 주어지는 인센티브를 받을 목적으로 설문을 제대로 읽지 않고 응답을 선택하여 설문을 건성으로 완료할 수 있다. 이와 같이 응답자가 설문의 내용과 상관없는 부정확한 응답을 제공하는 경우, 연구자가 이를 적절히 탐지하고 대응하지 않는다면, 연구결과가 왜곡될 위험성이 있다(Huang, Liu, & Bowling, 2015; Schmitt & Stults, 1985). 따라서 최근 연구들은 설문 응답자들이 설문에 충분한 노력과 주의를 기울이지 않아 발생하는 응답 형태를 불성실 응답(Insufficient effort responding 또는 Careless responding)으로 개념화하고 불성실 응답의 원인과 영향 및 이를 탐지하고 해결하기 위한 방안을 논의해왔다(Curran, 2016; DeSimone, DeSimone, Harms, & Wood, 2018; Meade & Craig, 2012; Oppenheimer, Meyvis, & Davidenko, 2009; Ward & Meade, 2018). 과거 연구들은 이와 같은 응답 형태가 연구 데이터의 질에 미치는 영향이 미미한 것으로 간주하였지만, 최근 연구들은 전체 데이터의 3~12%에 해당하는

설문에서 불성실 응답과 관련한 문제가 발생한다고 보고하였다(Johnson, 2005; Maniaci & Rogge, 2014; Meade & Craig, 2012). 특히, 'Applied Psychology: International Review'는 2018년 4월호에서 집중적으로 불성실 응답과 관련한 이슈를 다루었으며, 연구자들이 불성실 응답의 문제에 대해 지속적인 주의를 기울여야 한다고 강조하였다.

불성실 응답은 일반적으로 패널 조사와 같이 반복적인 설문조사로 인해 응답자가 설문 문항에 익숙해지거나(Baer, Ballenger, Berry, & Wetter, 1997), 설문 문항 수가 많아 응답자 피로가 증가할 때 관찰된다(Liu, Bowling, Huang, & Kent, 2013). 또한 응답자의 성격적 특성(Ward, Meade, Allred, Pappalardo, & Stoughton, 2017), 동기부여(Bowling, Huang, Bragg, Khazon, Liu, & Blackmore, 2016; Huang, Curran, Keeney, Poposki, & DeShon, 2012), 설문의 구성형태(Deutskens, De Ruyter, Wetzels, & Oosterveld, 2004), 설문참여에 대한 보상의 지급(Akbulut, 2015; Meade & Craig, 2012) 등과 같은 여러 요인에 의해서 발생할 수 있으며 나타나는 양상 또한 다양하다.

불성실 응답이 연구 결과에 미치는 영향은 불성실 응답의 형태에 따라 달라질 수 있지만, 일반적으로 오차 분산(error variance), 변수 간 상관관계(correlations), 내적 일치도 계수(internal consistency validity coefficient), 요인 적재치(factor loading)와 관련한 다양한 통계적 수치에 영향을 미친다(Meade & Craig, 2012). 특히, 적은 비율의 불성실 응답이라 할지라도 연구 결과에 부정적 영향을 미칠 수 있다는 것은 이미 여러 연구에서 보고되었다(Credé, 2010; Huang, Liu et al., 2015; Maniaci & Rogge, 2014; Woods, 2006). Credé(2010)은 전체 데이터의 5%에 해당하는 불성실 응답으로

도 변수들의 상관관계가 강해지거나 약해질 수 있음을 확인하였다. Woods(2006)의 연구는 역코딩 문항이 혼재된 상황에서 불성실 응답이 발생할 경우, 단일 개념을 측정하고자 함에도 불구하고 긍정적 문항과 역코딩 문항이 별개의 요인으로 적재되는 현상이 발생한다고 설명하였다. 연구 결과에 따르면, 불성실 응답이 10%만 존재하더라도 이러한 요인의 분리현상이 발생하여 단일차원을 측정하는 척도가 요인분석에서 단일요인 모형(one-factor model)으로서 적합성을 지니지 못하는 것으로 나타났다. Schmitt & Stults(1985)의 연구 역시 가상의 불성실 응답이 10% 포함될 경우 부정적으로 표현된 항목들이 별도의 요인을 구성하는 것으로 보고하였다. 또한, 불성실 응답은 변수 간 상관관계를 약화시키거나 (McGrath, Mitchell, Kim, & Hough, 2010), 강화시키는 방향(Huang, Liu et al., 2015) 모두로 작용할 수 있다. 이는 실증연구의 가설검정에 있어 불성실 응답의 존재가 실제 변수 간 상관관계가 없음에도 불구하고 이를 기각하지 못하는 1종 오류(type I error)와 상관관계가 존재하지만 이를 기각시키는 2종 오류(type II error)의 가능성 모두를 증가시킬 수 있음을 의미한다. DeSimone et al.(2018)은 불성실 응답의 상이한 형태에 따라 불성실 응답이 통계수치를 확대시키거나 축소시키는 경향이 다르게 나타남을 확인한 바 있다. 이처럼 불성실 응답의 형태와 그에 따른 영향이 다양하기에 이에 대처하기 위하여 여러 가지 불성실 응답의 탐지방법이 활용되고 있다. 또한, 최근의 연구들은 불성실 응답을 탐지하는 다양한 방법을 비교함으로써 효과적으로 불성실 응답의 문제를 해결하는 데 상당한 관심을 가지고 있다.

앞서 살펴본 바와 같이, 불성실 응답은 설문조사를 통해 수집한 데이터의 질은 물론이고 연구의 신

뢰도와 내적 타당도에 주요한 영향을 미치는 문제이지만, 불성실 응답에 대한 국내 연구자들의 인식과 논의는 부족한 실정이다. 또한 설문조사를 활용하는 기업과 조직들은 상당한 비용을 들여 이를 수행하는 과정에서 불성실 응답의 영향으로 잘못된 결과를 도출할 수 있다는 사실을 간과하고 있다. 이에, 본 연구는 불성실 응답의 개념과 유형을 체계적으로 정리한 후, 탐지방법과 불성실 응답으로 인한 문제의 해결방안을 모색하고자 한다. 특히, 실제 설문조사를 통해 수집한 데이터를 분석함으로써 설문조사에서 나타나는 불성실 응답에 대한 실질적인 이해와 탐지 및 해결방안의 활용을 돕고자 한다. 구체적으로는 3,030명의 국내 대기업 구성원을 대상으로 한 그룹차원 온라인 설문조사 결과를 분석함으로써 대표적인 불성실 응답의 탐지방법 및 제거효과를 실증적으로 검증하고자 한다. 실증적 검증을 통해 본 연구가 답하고자 하는 세부 연구질문은 다음과 같다. 첫째, 불성실 응답의 탐지방법에 따라 탐지되는 불성실 응답의 비율과 대상이 달라지는가? 둘째, 불성실 응답이 의심되는 자료를 제거하는 것은 통계적 추론 결과에 어떤 영향을 미치는가? 셋째, 2개 이상의 탐지방법을 조합하는 다중 허들방식(multi-hurdle)은 어떻게 적용할 수 있는가? 본 연구의 목적은 설문조사의 신뢰도와 타당도를 위협할 수 있는 불성실 응답에 대한 연구자들의 인식을 제고하고 탐지 및 해결방안의 활용을 돕는 한편, 이에 대한 논의를 진척시키는 것이다. 또한 설문조사를 수행하는 기업체의 실무자 차원에서도 불성실 응답에 대한 관심을 촉구하고 이를 탐지하고 해결하는 방안을 안내하는 데 연구의 의의가 있다.

II. 불성실 응답의 개념과 유형

2.1 불성실 응답의 개념

불성실 응답이란 설문조사에서 응답자가 '부주의하게 설문에 응답(careless responding)하는 것'(Schmitt & Stults, 1985; Ward & Pond, 2015)과 '의도적으로 무작위하게 응답(random responding)하는 것'(Beach, 1989; Berry et al., 1992)을 의미한다. 연구자의 관점에서, 응답자가 설문 항목의 내용을 충분히 고려하지 않는 경우의 응답과 무작위 응답은 응답자가 정확하지 못한 정보를 제공할 때 발생하는 응답 오차(response error)로 해석할 수 있는데, 불성실 응답에 따른 응답 오차가 클수록 연구 데이터의 신뢰성은 떨어지게 된다(이윤석 · 이지영 · 이경택, 2008; LaRose & Tsai, 2014). 초기 연구들에서 불성실 응답은 주로 무작위 응답을 의미하였으나, 최근에는 충분한 주의를 기울이지 않고 설문에 응답하는 경우도 불성실 응답의 범위에 포함하고 있다. 구체적으로, 불성실 응답은 정확한 응답을 제공할 동기가 낮거나 없는 경우(Huang et al., 2012), 의도적으로 무작위 응답을 하는 경우(Meade & Craig, 2012), 부주의하거나 혹은 응답 시 충분한 노력을 기울이지 않는 경우(Huang, Bowling, Liu, & Li, 2015), 또는 진정한 감정을 반영하지 않는 경우(Ward & Meade, 2018) 등으로 정의된다.

이러한 불성실 응답의 다양한 정의들은 공통적으로 '적당한 응답가설(satisficing hypothesis)'의 측면에서 이해할 수 있다(Barge & Gehlbach, 2012; Warwick, Rimmer, Blandford, Gow, & Buchanan, 2009). Krosnick(1991)은 설문 응답자들이 설문조

사에서 적당하게 응답하는 경향(satisficing behavior)이 있다고 주장하였다. 적당한 응답(satisficing)이란 "satisfy"와 "sacrificing"이 결합된 용어로, 이는 응답자들이 설문에 대해 철저한 심사숙고와 정보 처리 과정을 거쳐 답을 하는 것이 아니라 비교적 만족할만한 수준까지만 인지적 영역에서 정보처리를 한 후 답변하는 행동을 말한다. 설문에 정확하고 충실한 답변을 하는 것은 응답자로 하여금 상당한 수준의 인지적 노력을 요구하기 때문에, 일부 응답자들은 많은 인지적 노력을 부여하지 않고 설문에서 받은 인상을 기반으로, 정확한 답변 대신 자신들이 적당하다고 생각하는 객관식 보기를 선택한다는 것이다. Krosnick(1991)의 관점은 다양한 형태의 바람직하지 않은 응답자 행동을 설명할 수 있는 이론적 기반으로 폭넓게 적용되고 있다(백영민 · 김은미 · 이준웅, 2012; 이윤석 외, 2008; Barge & Gehlbach, 2012). 설문조사에서 응답자가 보이는 적당한 응답 패턴들은 (1) 첫 번째 보기를 답으로 선택하는 경향, (2) 질문에 대해 긍정적으로 응답하는 경향, (3) 일련의 척도형 질문에서 비차별적으로 똑같이 응답하는 경향(non-differentiation), (4) '잘 모름(don't know)'을 답으로 선택하는 경향, 그리고 (5) 객관식 설문에서 무작위로 응답하는 경향으로 구분할 수 있다(이윤석 외, 2008; Krosnick, 1991). 불성실 응답의 예로, 설문 응답자가 '내 직업에 만족한다.'는 항목에 대해 '매우 동의함'을 선택하고, 앞의 응답과 거의 동일한 항목인 '내 직업으로 만족감을 느낌.'에 대해서는 부주의하게 '동의하지 않음'을 선택하는 경우를 들 수 있을 것이다. 또한, 불성실 응답은 응답자가 의도적으로 체계적인 패턴으로 응답하는 형태로 나타나기도 한다. 예를 들어, 하나의 설문조사에서 응답자는 30개의 연속적인 항목에 대해 모두 '거의 동의함'이라는 응답을 선택할 수 있다.

최근 불성실 응답은 자기보고식 설문 데이터의 신뢰도를 위협하는 원인으로 지적되고 있으며, 표본에 불성실 응답이 아주 적은 비율로 포함되는 것만으로도 척도 특성과 그에 상응하는 연구결과의 손상이 야기된다는 점이 밝혀졌다(Huang, Liu et al., 2015). 또한 불성실 응답은 기업 및 조직에서 광범위하게 활용되는 온라인 설문에서도 데이터의 질을 떨어뜨릴 수 있는 요인이라고 지적되었다(Niessen, Meijer, & Tendeiro, 2016). 예를 들어, 온라인 설문을 활용하는 기업과 조직들은 상당한 비용을 들여 기업의 중요한 의사결정과 인적자원정보시스템(HRIS: human resources information system) 분석의 효율성을 위하여 수행하는 설문 과정에서 불성실 응답의 영향으로 잘못된 결과를 도출할 수 있다. 구체적으로, 기업과 조직 데이터를 활용한 불성실 응답 관련 연구들은 개인 성격(Hough, Eaton, Dunnette, Kamp, & McCloy, 1990), 직무 분석(Green & Stutzman, 1986), 직무 평가(Calsyn & Winter, 1999), 온라인 설문(Behrend, Sharek, Meade, & Wiebe, 2011; Niessen et al., 2016) 등의 다양한 주제와 맥락에서 분석을 수행하였다.

2.2 불성실 응답의 유형

설문 응답자들의 불성실 응답은 다양한 유형으로 나타날 수 있다(DeSimone et al., 2018). 먼저, 불성실 응답이 전체 문항에서 차지하는 비중에 따라, 불성실 응답이 전체 문항에 걸쳐서 나타나는 경우와 부분적으로만 나타나는 경우로 구분할 수 있다(Meade & Craig, 2012). 또한, 불성실 응답은 설문조사의 응답자가 내용을 고려하지 않고 완전 무작위로 응답을 선택하는 '무작위 응답'과 응답자가 설문에 충분한 주의를 기울이지 않으나 일정한 패턴

으로 응답을 선택하는 '비무작위 응답(nonrandom responding)'으로도 구분이 가능하다(Bowling & Huang, 2018; Johnson, 2005; Kam & Meyer, 2015). 특히, 후자의 경우, 연속적으로 동일한 번호의 보기를 선택하는 '한 줄 응답(long string responding 또는 straight lining)'이 대표적으로 알려져 있다(Kim, McCabe, Yamasaki, Louie, & King, 2018). 설문조사의 질과 관련된 초기의 연구들은 무작위 응답의 식별에 중점을 두었는데, 최근에는 한 줄 응답이 무작위 응답과 더불어 설문조사의 질을 나타내는 결정적 지표로 주목받고 있다(DeSimone et al., 2018; Schonla & Toepoel, 2015). 기존 연구에서 제시된 불성실 응답의 세부 유형은 <표 1>에 간략하게 정리하였다.

2.2.1 무작위 응답

일반적으로 응답자가 내용을 고려하지 않고 선택한 응답에 일정한 패턴이 나타나지 않는 경우를 무작위 응답으로 구분하지만, 엄밀한 의미에서의 무작위 응답은 설문 응답자가 각 응답을 유사한 확률로 선택하는 경우를 의미한다. 설문에 참여하는 응답자들이 불성실하게 답변을 선택하면서도 주의 깊게 응답하는 것처럼 보이기 위해 다양한 응답을 선택할 때 이러한 응답 형태가 나타날 수 있다(DeSimone et al., 2018). 무작위 응답의 형태는 연구자에 따라 보다 구체적으로 구분되거나 정의된다. 특히, 시뮬레이션 연구는 무작위 응답을 세분화하거나 엄격한 정의를 제시하는 경향이 있다(Credé, 2010; Meade & Craig, 2012). Kim et al.(2018)은 참가자가 각 응답을 거의 같은 확률로 선택하는 경우를 균일 무작위 응답(uniformly random responding)이라 정의하였다. Meade & Craig(2012)는 척도

의 선택지에 균등하게 무작위로 응답을 하는 균등분포 무작위 모형(uniformly random response pattern)과 척도의 중앙점을 기준으로 하여 정규분포 패턴을 보이는 무작위 모형(normally distributed random pattern)으로 무작위 응답의 형태를 세분화하였다. 한편, Credé(2010)는 무작위 응답을 진정한 무작위 응답(truly random responding)과 효과적인 무작위 응답(effectively random responding)으로 구분하였는데, 진정한 무작위 응

답은 응답자가 설문 문항에 대해 각 응답을 선택할 확률이 동일한 경우를 말한다. 반면, 효과적인 무작위 응답이란 설문에 참여한 응답자가 설문의 내용에 주의를 기울이지 않고 응답을 선택하지만 각 응답을 선택하는 확률이 같지 않은 경우를 의미한다. 그러나 Credé(2010)가 정의하고 있는 효과적인 무작위 응답은 엄격한 의미에서의 무작위 응답이라고 할 수는 없으며, 오히려 무작위 응답과 대조되는 의미로 보아야 한다.

〈표 1〉 불성실 응답의 유형별 분류, 의미 및 예시

구분기준	유형	의미	예시
불성실 응답의 '비중'	완전 불성실 응답 (totally careless responding)	<ul style="list-style-type: none"> 불성실 응답이 전체 문항에 걸쳐서 나타남 	<ul style="list-style-type: none"> Meade & Craig(2012): 불성실 응답을 전체 설문에서 차지하는 비중에 따라 완전 불성실 응답과 부분 불성실 응답으로 구분
	부분 불성실 응답 (partially careless responding)	<ul style="list-style-type: none"> 불성실 응답이 부분적으로만 나타남 	
불성실 응답의 '형태'	무작위 응답 (random responding)	<ul style="list-style-type: none"> 척도의 선택지에 균등하게 무작위로 응답 	<ul style="list-style-type: none"> Kim et al.(2018): 균일 무작위 응답 (uniformly random responding) Credé(2010): 진정한 무작위 응답 (truly random responding) Meade & Craig(2012): 균등분포 무작위 패턴(uniformly random response pattern) Schneider, May, & Stone(2018): 완전 무작위 응답(completely random response)
		<ul style="list-style-type: none"> 척도의 중앙점을 기준으로 하여 정규분포 패턴을 보이는 무작위 응답 	<ul style="list-style-type: none"> Meade & Craig(2012): 정규분포 무작위 패턴 (normally distributed random pattern)
	비무작위 응답 (nonrandom responding)	<ul style="list-style-type: none"> 내용에 반응하지 않고 같은 방법 또는 방향으로 응답하는 일정한 패턴을 보이는 응답으로, 한 줄 응답이 대표적 	<ul style="list-style-type: none"> Schneider et al.(2018): 균일 반복 응답 (uniformly repeated responses) Credé(2010): 효과적인 무작위 응답 (effectively random responding) Kim et al.(2018): 긴 문자열 응답(long string responding), 비무작위 패턴 응답(nonrandom patterned responding) Schonla & Toepoel(2015), DeSimone et al.(2018): 직선형 응답(straight lining)

2.2.2 비무작위 응답

설문 응답자가 동일한 응답을 반복적으로 선택하거나 내용과 상관없이 일정한 패턴으로 응답을 선택함으로써 설문조사에 불성실한 답변을 하는 경우가 있다(DeSimone et al., 2018; Huang et al., 2012). 예를 들어, 각 문항 당 1번부터 5번까지의 응답 중 선택하도록 설계된 설문이 참가자에게 제공되었을 때, 응답자가 설문에 주의를 기울이지 않고 1번과 5번 응답을 번갈아 가며 선택하는 경우나 1, 2, 3, 4, 5번을 차례로 선택하는 등의 경우가 있을 수 있다. 이러한 경우 참여자의 응답은 특정한 패턴을 나타내는데 응답자가 설문조사에 충분한 주의를 기울이지 않고 답을 선택하였기 때문에, 이 역시 불성실 응답이 동반하는 다양한 문제점을 야기한다(Johnson, 2005; Meade & Craig, 2012). 특히, 한 줄 응답은 비무작위 응답의 대표적인 유형으로, 응답자가 설문의 내용에 충분한 주의를 기울이지 않고 연속적으로 동일한 응답을 선택하는 경우이다. 먼저 제시된 예와 같이, 각 문항 당 1번부터 5번까지의 응답 중 선택하도록 설계된 설문이 제공되었을 때, 응답자가 각 문항을 제대로 이해하고 충실한 사고과정을 거쳐 적절한 선택을 하는 것이 아니라, 1번을 반사적으로 여러 문항에 걸쳐 연속하여 선택하는 경우, 복수의 문항을 하나의 화면으로 보았을 때 설문참여자의 응답이 한 줄의 직선이 그려진 것처럼 패턴을 형성하게 되는 것이다.

III. 불성실 응답의 탐지

불성실 응답의 탐지방법은 탐지하고자 하는 응답

의 유형 및 특성에 따라 다양하며, 각 방법이 가지는 한계점을 고려하여 복합적으로 조합하여 활용되고 있다. 본 연구에서는 여러 탐지방법을 <표 2>와 같이 크게 (1)직접적 방법(direct methods), (2)비개입적 방법(unobtrusive methods), (3)통계적 방법(statistical methods)으로 구분하여 설명하고자 한다. 직접적 방법은 설문 실시 전에 직접적으로 불성실 여부를 탐지할 수 있는 문항을 작성하여 반영한다는 의미에서 사전적(pre-test) 방법에 속하며, 나머지 두 방법은 이미 도출된 설문 결과를 가지고 불성실 응답을 탐지할 수 있다는 점에서 사후적(post-test) 방법으로 분류하기도 한다.

3.1 직접적 방법

다양한 탐지방법 중 응답자에게 직접적으로 설문 문항을 통하여 불성실 응답여부를 탐지하는 방법을 직접적 탐지방법(direct methods)이라 한다. 직접적 탐지방법으로는 지시적 조작 점검 문항(Oppenheimer et al., 2009) 및 가짜문항(Huang, Bowling et al., 2015; Meade & Craig, 2012), 자기보고식 노력 측정 문항(Curran, 2016; Meade & Craig, 2012; DeSimone & Harms, 2018) 등이 활용되고 있다.

3.1.1 지시적 조작 점검 문항

지시적 조작 점검 문항(instructional manipulation check item)은 응답자에게 특정 답변을 선택할 것을 지시하는 방식으로, 지시된 문항(instructed question) 또는 주의력이 걸렸던 응답을 걸러내기 위한 문항(attention filtering question)이라고도 불린다(DeSimone & Harms, 2018). 이는 설문

중 일부에 포함된 지시적 조작 점검 문항에서 명시된 지시에도 불구하고 이를 따르지 않고 다른 답을 고른 응답자들을 불성실 응답자로 간주하는 것이다. 예를 들어, 긴 제시문 뒤에 하나의 핵심 문장을 넣는데, 이 핵심 문장을 통해 앞서 나온 제시문을 무시하고 특정 보기를 선택하라고 지시하거나, 혹은 제시문 없이 “이 항목에 대해서 무조건 동의함.” 이라고 체크를 지시하는 방법(DeSimone & Harms, 2018) 등이 있다. 따라서 주의를 기울이지 않고 응답하던 불성실 응답자들은 이 지시에 따른 정답을 선택하지 못할 것이고, 연구자들은 이 문항에 대한 응답을 바탕으로 불성실 응답을 탐지할 수 있다. 그 밖의 지시적 조작 점검 문항의 예로는 “나는 설문지의 지시사항을 읽었습니다.” 등의 문장을 직접 설문에 쓰게 하는 방법(Oppenheimer et al., 2009)도 있다. 이러한 지시적 조작 점검 문항의 경우, 다른 문항과 비슷한 형식으로 두드러지지 않게 제시되어야 탐지의 효과성을 발휘할 수 있다(Huang, Bowling et al., 2015; Meade & Craig, 2012). 지시적 조작 점검 문항이 두드러지는 경우, 전체 설문에 불성실하게 답한 응답자들도 해당 문항에 응답하는 순간에는 집중하게 되어 지시사항을 따라 맞은 답을 선택하게 될 것이기 때문이다.

3.1.2 가짜문항

가짜문항(bogus item) 방법은 앞서 설명한 지시적 조작 점검 문항 방법과 유사한 방식으로 불성실 응답을 탐지하는데, 실제 연구내용과 전혀 무관하지만 정답이 정해져 있는 질문들을 설문제 임의로 배치한 후, 응답자가 이를 맞추지 못하면 불성실 응답으로 판별하는 방법이다. 특정 답을 고르라는 지시는 없다는 점에서 지시적 조작 점검 문항 방법과 차

이가 있다. 예를 들어, “나는 하루에 잠을 1시간도 자지 않는다.”라는 가짜 문항에 대해서는 7점 척도 중 1점(아니다) 또는 2점(매우 아니다)을 선택한 경우만 성실 응답으로 간주하고 나머지 3점~7점 선택은 불성실 응답으로 간주하는 방식으로 사용한다(Bagby, Gillis, & Rogers, 1991; Meade & Craig, 2012). 이 방식은 활용 방법이 간단하고 직관적이라는 장점이 있다. 그러나 가짜문항의 내용은 설문제에 대한 신뢰도를 낮추고 응답자들의 감정을 자극할 가능성이 있으므로, 반드시 연구지시문에 데이터 품질을 높이기 위한 가짜문항이 섞여 있음을 미리 밝히도록 권장된다(Breitsohl & Steidelmüller, 2018).

연구자들은 전체 설문문항의 길이에 따라 2~11개까지 다양한 범위의 가짜문항을 자신의 연구목적에 맞추어 변형하여 사용하며, Roivainen, Veijola, & Miettunen(2016)의 경우처럼, 별도로 타당도가 확인된 특이응답 탐지척도(예, CIS; Chapman Infrequency Scale)를 연구에 포함하여 활용하는 경우도 있다. 복수의 가짜문항이 존재하는 경우, 응답이 몇 개 이상일 때 불성실 응답으로 판단하는지의 기준은 전체 가짜문항의 개수에 따라 달라질 수 있다. 예로, DeSimone & Harms(2018)의 경우, 가짜문항 및 지시적 조작 점검 문항 총 4개 중 1개라도 틀린 경우, 해당 응답자를 불성실 응답자로 간주했으며, 전체 샘플 중 약 7.5%가 이러한 기준에 의하여 탐지되었다.

3.1.3 자기보고식 노력 측정 문항

자기보고식 노력 측정(self-reported effort item)은 응답자에게 직접 본인이 설문제에 응답하는 데 얼마나 주의와 노력을 기울였는지 평가하도록 하는 방

법이다. Meade & Craig(2012)는 이와 관련하여 응답자의 참여열의(participant engagement) 수준을 측정할 수 있는 도구를 개발하였다. 설문 응답에 있어서의 성실함(diligence)과 흥미(interest), 두 하위 척도로 구성된 이 측정도구는 불성실 응답을 탐지하는 다른 측정도구들과 높은 상관관계를 보였다. 해당 연구자는 성실함과 흥미를 측정함과 더불어, 응답자에게 활용가능 여부를 직접 묻는 문항(예, “솔직하게 판단하건대, 당신의 데이터는 우리가 분석에 사용해도 문제가 없을 데이터입니까?”)을 활용하기도 하였다. 이 방법이 지닌 단점은 불성실 응답을 실제보다 과대 측정하게 되는 1종 오류를 범할 가능성이 있다는 것이다(Curran, 2016). 사용할 만한 데이터의 기준이 응답자마다 다를뿐더러, 자기보고식 측정 문항에 솔직하게 답할 유인도 다르기 때문이다. 설문에 성실히 임한 양심적인 응답자일수록 설문이 진행됨에 따라 스스로 주의력이 감퇴하는 것을 느끼고 응답의 질을 과소평가할 가능성이 높으며, 불성실 응답자일수록 보상을 받지 못하게 될 것이 두려워 성실히 응답했다고 거짓말을 할 수도 있다. 따라서 이 방법을 단독으로 사용하기보다는 다른 탐지방법과 함께 사용하는 것이 안전하며, Curran (2016)이 권고한 바와 같이, 설문을 시작하기 전에 지시사항을 통해 해당 자기보고식 노력 측정 문항에 대해 어떠한 답변을 하더라도 보상은 주어진다는 것을 밝히고 솔직히 답변해 줄 것을 미리 부탁한다면 1종 오류를 예방하는 데 도움이 될 것이다.

3.2 비개입적 방법

비개입적 방법(unobtrusive methods)은 설문을 실행하는 과정에서 얻은 정보를 활용한 탐지방법으로, 비교적 간단한 통계 기법을 사용하여 불성실 응

답을 탐지한다. 별도의 계산이나 수치의 변환을 요구하지 않으며, 관찰된 응답자의 응답패턴을 그대로 활용하기에 ‘비개입적’이라 불린다(DeSimone & Harms, 2018). 비개입적 탐지방법으로는 응답시간(DeSimone & Harms, 2018), 한 줄 응답(Behrend et al., 2011; Huang et al., 2012; Meade & Craig, 2012), 개인 응답 분산도(Dunn, Heggstad, Shanock, & Theilgard, 2018) 등이 활용되고 있다.

3.2.1 응답시간

설문에 응답하는 데 소요된 시간(response time)이 예상보다 지나치게 짧을 경우, 불성실 응답이 이루어졌다고 의심해볼 수 있다. 응답시간은 전체 설문 문항 또는 개별 문항이나 개별 페이지에 대해 기록될 수 있는데, 설문의 특정 부분에 집중되어 있는 불성실 응답을 탐지하는 데에는 후자의 경우가 더 효과적일 것이다(DeSimone & Harms, 2018). 하지만, 응답시간과 응답의 질 간에는 비선형적인 관계가 존재하기 때문에(Meade & Craig, 2012), 응답시간만으로 불성실 응답을 탐지하여 걸러내는 방법은 1종 오류 발생 가능성을 높인다는 측면에서 효과적이지 못할 수도 있다. 실제로 Meade & Craig (2012)의 결과에 따르면, 총 응답시간이 비정상적으로 짧았던 소수의 응답자를 제외한 나머지 불성실 응답자들, 즉 응답시간 외 다른 측정도구에 의해 불성실하다고 탐지된 응답자들의 대부분은 응답시간이 짧지 않은 것으로 나타났다. 즉, 응답시간만으로는 설문에 집중하지 못하고 도중에 다른 행위를 하느라 총 응답시간이 길게 나타나는 불성실 응답자들을 탐지할 수 없다는 한계가 있다. 그러나 이러한 한계에도 불구하고 응답시간은 비교적 간단하게, 시간과

비용 측면에서 효율적으로 불성실 응답을 걸러낼 수 있는 방법으로, 여러 연구를 통해 그 효용성이 입증된 바 있다(DeSimone & Harms, 2018; Meade & Craig, 2012). 따라서 응답시간을 단독으로 사용하기보다는 다른 불성실 응답 탐지방법과 함께 보조적으로 사용하는 것이 불성실 응답을 탐지하는 데 보다 효과적일 것이다.

3.2.2 한 줄 응답

연속된 문항에 대해 응답자가 비정상적으로 동일한 응답을 함으로써 같은 응답으로 구성된 긴 줄이 형성되는 경우, 해당 응답자가 설문에 임하는 태도가 불성실하다고 의심해볼 수 있다(DeSimone & Harms, 2018; Huang et al., 2012; Johnson, 2005; Meade & Craig, 2012; Schonla & Toepoel, 2015). 불성실 응답 탐지 지표로서의 한 줄 응답(long string 또는 straight lining)은 기본적으로 연속적 문항들에 대해 개별 응답자가 동일한 응답을 한 최대 횟수(maximum long string, 이하, 최댓값 방법)를 의미한다. 예를 들어, 설문의 전체 문항에 답한 응답자가 최대 9개의 문항에 동일한 답변을 연속적으로 한 경우, 해당 응답자의 한 줄 응답의 값은 9가 된다. 한편, 전체 설문에 나타난 한 줄 응답 길이의 평균값(average long string)을 한 줄 응답의 값으로 정의할 수도 있는데, Meade & Craig(2012)는 두 방식이 불성실 응답을 걸러내는 효과에 있어서 유의한 차이가 없으므로 상대적으로 더 간편한 최댓값 방법을 활용할 것을 권고하고 있다.

한 줄 응답으로 불성실 응답을 판단하는 기준은 연구자마다 다양하다. 가장 편의성이 높은 방법은 전체 문항 중 특정 개수 이상 한 줄 응답을 보일 경우 불성실 응답으로 판단하는 것으로 기준을 정해두는

것이다. 예를 들어, DeSimone & Harms(2018)은 24개의 항목으로 이루어진 자기보고식 성격 검사(ACST: abbreviated character strengths test)에서 9개 이상의 연속된 응답이 나올 경우 불성실 응답으로 간주하였다. 이 기준을 정하는데 어려움이 있을 경우, Johnson(2005)의 방식을 참고할 수 있다. Johnson(2005)은 스크리 검사식의 방법(scree-like test)을 적용하여 한 줄 응답의 빈도가 급격히 하락하면서 현저하게 값이 작아지기 시작하는 값을 불성실 응답을 탐지하는 기준치로 적용하였다. 한편 보다 복잡한 기준 설정 방식으로는 Costa & McCrae(2008)의 방식이 있다. 이들의 연구에서는 사람들이 설문에서 좌우 양단의 극단적인 응답과 부정적 응답을 잘 선택하지 않는 경향이 있음을 고려하여, 5점 리커트(Likert) 척도의 각 항목별로 서로 다른 기준을 적용하였다. 예를 들어, “전혀 동의하지 않는다.”는 6개, “대체로 동의하지 않는다.”는 9개, “보통이다.”는 10개, “대체로 동의한다.”는 14개, “매우 동의한다.”는 9개를 한 줄 응답의 불성실 판단 기준으로 적용하는 것이다.

한 줄 응답을 사용하여 불성실 응답을 탐지하는 방법은 같은 내용을 반대로 질문한 문항(예, 역코딩 문항)이 있거나 서로 다른 개념에 대한 측정 문항이 뒤섞여 있는 경우 특히 효과적이다(Dunn et al., 2018; Schonla & Toepoel, 2015). 이 때, 불성실 응답으로 탐지하는 기준치는 절대적인 기준을 따르기보다 문항 및 항목의 개수를 고려하여 비정상적인 패턴을 보이는 값을 찾는 것이 바람직할 것이다.

3.2.3 개인 응답 분산도

개인 응답 분산도(individual response variability)란 Dunn et al.(2018)이 한 줄 응답의 개념을 확

장하여 제시한 불성실 응답 탐지방법으로, 연속된 설문 문항에 대한 특정 응답자의 답변들의 분포를 구했을 때 응답들의 표준편차를 의미한다. 개인 응답 분산도는 한 줄 응답과 반비례 관계에 있으며, 한 줄 응답은 그 값이 클수록, 개인 응답 분산도는 그 값이 작을수록 불성실 응답일 확률이 높다고 간주한다. 설문이 복수의 개념을 측정하거나 역코딩 문항을 포함하는 경우, 정상적인 응답자는 각 개념 또는 문항 유형에 대하여 다른 형태의 응답을 보여야 하기 때문이다. 그러나 응답자가 모든 설문 문항에 대해 동일한 답변을 제공한 경우, 즉 한 줄 응답이 발생한 경우, 개인 응답 분산도는 매우 작은 값을 가지게 될 것이고, 해당 응답자의 데이터를 불성실 응답으로 간주해 제거할 수 있을 것이다.

유사한 유형의 불성실 응답을 탐지하는 것으로 보이나, 개인 응답 분산도는 한 줄 응답 탐지방법이 탐지하지 못하는 보다 미묘한 형태의 한 줄 응답 유형들도 탐지해 낼 수 있다(Dunn et al., 2018). 응답자가 특정 선택지를 사용해 비교적 짧은 길이의 한 줄 응답을 보인 후 인접 선택지를 사용해 또 다른 짧은 길이의 한 줄 응답을 보이는 행위를 반복하는 유형(예, 2번 연속 선택 후 3번 연속 선택), 그리고 특정 선택지와 인접 선택지를 번갈아가며 선택하는 행위를 반복하는 유형(예, 2번과 3번 계속 번갈아 선택) 등이 이에 속한다. 또한 Dunn et al.(2018)은 타 탐지방법에 비해 계산이 쉬워 간단하게 불성실 응답을 탐지해 낼 수 있다는 점, 설문의 전체가 아닌 일부분에도 적용이 가능할 뿐 아니라 다양한 길이의 응답에 적용할 수 있는 유연한 탐지방법이라는 점을 개인 응답 분산도의 장점으로 제시하였다.

그러나 개인 응답 분산도는 한 줄 응답 방법과 마찬가지로 불성실 응답을 분류하는 명확한 기준치가 없어 문제가 된다(DeSimone & Harms, 2018).

이에 대해 Dunn et al.(2018)은 개인 응답 분산도는 사용된 문항의 수와 같이 설문의 성격에 따라 달라질 수 있으므로 명확한 기준치를 사용하기 어렵다고 하였다. 그리고 전체 응답표본 중 탐지할 응답의 비율에 따라 기준치를 설정한 후, 불성실 응답으로 탐지된 표본 중 개인 응답 분산도가 가장 낮은 응답이 불성실 응답에 해당되는지 여부를 살펴 기준치를 다시 조정하는 것을 대신 권고하였다. 이 때 탐지할 응답의 비율은 설문의 길이나 문항의 내용 등 불성실 응답과 관련된 요소들을 고려하여 설정하는데, Dunn et al.(2018)은 10%를 사용하였다. 또한, Dunn et al.(2018)에 따르면, 개인 응답 분산도 값은 개인 응답의 중심경향성(central tendency response style)(Bachman & O'Malley, 1984; Bardo, Yeager, & Klingsporn, 1982)에 크게 영향을 받는데, 역코딩 문항을 사용하여 중심경향성 문제의 발생 여부를 확인할 수 있다. 역코딩 문항 여부에 따라 응답 경향이 달라지는 경우, 중심경향성 문제가 발생하지 않는다고 볼 수 있는 것이다. 뿐만 아니라, Dunn et al.(2018)은 개인 응답 분산도를 계산하는 경우 총 25개 이상, 150개 이하의 문항을 사용하는 것을 권장하였다.

3.3 통계적 방법

통계적 방법(statistical methods)으로 분류된 탐지방법은 사후적으로 분석 가능하다는 점에서는 앞서 설명한 비개입적 방법과 유사하지만, 보다 복잡한 통계 기법을 사용하며 불성실 응답을 탐지하기 위해 여러 단계가 필요하다는 점에서는 비개입적 방법과 차이를 보인다(DeSimone & Harms, 2018). 통계적 방법으로는 마할라노비스 거리(Curran, 2016; Meade & Craig, 2012), 일관성 관련 지표(Curran,

2016), 동의/반의 일관성(Curran, 2016; Johnson, 2005; Meade & Craig, 2012) 등이 활용되고 있다.

3.3.1 마할라노비스 거리

마할라노비스 거리(Mahalanobis distance)는 흔히 다변량(multivariate) 분석에서 이상치(outlier)를 찾아내는 방법으로 활용된다(Mahalanobis, 1936; Kline, 2015). 두 개의 변수가 존재하는 다변량 데이터의 경우, 마할라노비스 거리는 다변량 공간(multivariate space)의 중심점(\bar{X}_1, \bar{X}_2)으로부터 특정 데이터까지의 거리를 두 변수의 공변량(covariance)으로 표준화한 값을 의미한다. 변수 X_1, X_2 의 공변량 매트릭스를 C 라고 할 때, 피험자 문항(person by item) 매트릭스 X 의 X_i 값이 가지는 마할라노비스 거리를 수식으로 나타내면 다음과 같다.

$$\begin{aligned} & \text{다변량 중심으로부터의 표준화된 거리} \\ & = \text{마할라노비스 거리}(D) \\ & = \sqrt{(X_i - \bar{X})C_X^{-1}(X_i - \bar{X})^T} \end{aligned}$$

충분한 주의를 기울이지 않은 불성실한 응답은 성실한 응답들과 멀리 떨어진 이상치를 만들어내는 분명한 원인이다(Curran, 2016). 따라서 이상치 분석을 위해 활용되는 이 계산법을 불성실 응답의 탐지에도 활용할 수 있을 것이며, 이는 실제로 여러 연구를 통해 효과적인 것으로 확인되었다(Maniaci & Rogge, 2014; Meade & Craig, 2012). 그러나 이 방법 역시 표본 데이터의 분포에 따라 그 효과가 크게 달라지고, 불성실 응답 탐지의 기준치가 연구자마다 다르다는 한계점이 있다(Curran, 2016; Meade & Craig, 2012). 구체적으로, 마할라노비

스 거리를 사용한 방법은 불성실 응답이 균등분포(uniform random distribution)를 이루거나 일부 문항에 대해서 정규분포를 따를 때에는 불성실 응답을 매우 효과적으로 탐지해냈으나, 불성실 응답이 모든 문항에 대해서 정규분포를 따를 때, 즉, 불성실 응답이 대부분의 문항에 대해 지배적으로 나타나는 경우에는 탐지능력이 크게 감소하는 것으로 나타났다(Meade & Craig, 2012). 따라서, 연구자들은 샘플 데이터의 분포와 특성을 고려하여 마할라노비스 거리 방법의 사용 여부를 신중하게 결정해야 할 것이다. 활용 시 기준에 관하여 DeSimone & Harms(2018)은 마할라노비스 거리의 제곱값(D^2)이 카이 제곱(χ^2) 분포의 상위 5% 이내일 경우, 불성실 응답으로 분류하였다. 혹은 다변량 분석의 이상치 탐지기준을 준용할 수도 있다. Kline(2015)에 따르면, 카이 제곱(χ^2) 분포를 따르는 특정 케이스의 D^2 값이 매우 높아서 나머지 케이스들과 동일한 모집단에서 왔다는 영가설을 기각하는 경우가 이에 속한다. 다만 이 경우, Kline(2015)은 영가설 기각에 대한 유의수준을 .001로 보수적으로 설정할 것을 권장하였다.

3.3.2 개인 내 일관성

개인 응답의 일관성(consistency)을 기반으로 불성실 응답을 탐지하는 방법으로는 척도를 반으로 나누어 두 그룹 간의 상관을 측정하는 홀짝 일관성 방법(odd-even reliability)과 반복추출된 개인 신뢰성(re-sampled individual reliability) 방법이 있다. 두 방법 모두 크제는 유사한 개념에 대해 일관된 답변을 했는지를 살펴보는 개인 신뢰성(individual reliability; Huang et al., 2012; Huang, Bowling et al., 2015; Jackson, 1977) 방법에

속한다. 홀짝 일관성의 경우, 개인의 응답을 홀수 문항에 대한 응답과 짝수 문항에 대한 응답으로 나누어 두 개의 벡터를 만든 후 두 벡터 간의 상관을 구하는 방법으로 구할 수 있다(Curran, 2016; Meade & Craig, 2012). 동일 개념을 측정하고 있음에도 홀짝 문항 간 상관관계가 낮다면, 일관되지 않은 응답, 즉, 성실하지 않은 응답으로 간주할 수 있다. Jackson(1977)은 이러한 상관계수의 값이 0.3보다 낮을 경우 비일관적인 응답으로 분류할 것을 권장하였다. 그러나 홀짝 일관성을 구할 때에는, 표본수의 최댓값이 문항 수의 2분의 1로 제한되는 바, 상관계수의 값이 문항의 길이에 의해 영향을 받는다는(scale-dependent) 단점이 있어 이를 절대적인 기준으로 보기는 어렵다(Curran, Kortba & Denison, 2010). 한편, 반복추출 된 개인 신뢰성은 홀짝 일관성의 개념을 확장시킨 것으로 척도를 홀짝 번호로만 나누어 벡터 쌍을 1개만 추출하는 대신, 각 척도의 응답을 무작위하게 두 그룹으로 반씩 나누어 두 개의 벡터를 추출하는 작업을 반복수행한 후 그 상관값들의 평균을 구하는 방법이다. 따라서, 이는 동일한 개념을 측정하는 문항에 있어서 개인의 응답이 일관될 것이라는 개인 신뢰성의 개념을 홀짝 일관성에서 보다 충실히 실현한 형태라고 할 수 있다(Curran, 2016). 이 때, 비일관성의 탐지 기준으로는 홀짝 일관성과 같이 0.3을 적용할 수 있을 것이나, 추후 연구를 통한 검증이 필요할 것으로 보인다.

3.3.3 동의/반의 일관성

동의/반의 일관성(synonyms/antonyms consistency)은 개인 내 일관성과 마찬가지로 일관성을 기준으로 불성실 응답을 분류하는 방법이다. 성실한 응답자라면 의미상 유사한 항목에 대해서는 유사한 답변을,

반대되는 항목에 대해서는 상반된 답변을 할 것이라는 전제를 활용하는 것이다. 이는 설문 내 전체 문항 중 비교에 활용할 동의/반의 관계에 있는 문항 쌍들을 결정하는 방법에 따라 의미론상(semantic) 방법과 심리측정상(psychometric) 방법으로 나뉜다(Curran, 2016; Goldberg & Kilkowski, 1985; Johnson, 2005). 의미론상 방법은 연구자가 사전에 문항들의 의미를 해석하여 유사한 답변이 예상되는 문항의 쌍을 지정하여 활용하는 방법이다(예, 상사에 대한 신뢰도와 상사에 대한 애정도). 반면, 심리측정상 방법은 연구자 임의의 판단을 배제한 채 사후에 전체 응답자들의 실제 응답 중 상관관계가 높은 문항 쌍을 통계적으로 추출하여 활용하는 방법이다. 이 경우 의미상 직접적인 상관은 없지만 전체 응답자들의 응답에서 상관관계가 높게 나타난 문항 쌍이 선택될 수도 있다(예, 상사에 대한 만족도와 동료에 대한 만족도). 이런 방법으로 서로 동의 또는 반의 관계에 있는 문항 쌍들을 다수 지정한 후, 각 그룹에 대한 개인의 응답을 벡터화하여, 두 벡터 간의 상관을 구하는 방법으로 동의/반의 일관성을 계산한다. 만일 유사한 의미를 지닌 문항들 벡터 간의 상관관계가 음수로 나타나거나 또는 0.3 미만으로 낮게 나타난 경우, 해당 응답자가 유사한 의미를 지닌 문항에 대해 일관되지 않은 응답을 한 것으로 해석하여 불성실 응답으로 분류할 수 있다(Curran, 2016; DeSimone & Harms, 2018; Meade & Craig, 2012). 그러나 이 역시 홀짝 일관성과 같이 척도를 절반으로 나누는 방법을 사용하므로, 결과값이 척도의 길이나 형태에 의해 영향을 받는다는 한계점을 갖는다(Curran et al., 2010).

〈표 2〉 불성실 응답의 탐지방법 및 특징

구분	기법	세부기법 및 의미	탐지 기준치
1. 직접적 방법 (direct methods)	지시적 조작 점검 문항 (instructional manipulation check item)	<ul style="list-style-type: none"> 응답자에게 특정한 응답을 할 것을 지시하는 문항(DeSimone & Harms, 2018; Oppenheimer et al., 2009) 	<ul style="list-style-type: none"> 해당 문항에 대해 오답을 제시할 경우, 불성실 응답으로 분류(DeSimone & Harms, 2018; Oppenheimer et al., 2009)
	가짜문항 (bogus item)	<ul style="list-style-type: none"> 특정한 정답을 내재하고 있는 문항(Bagby et al., 1991; Huang, Bowling et al., 2015; Meade & Craig, 2012) 	<ul style="list-style-type: none"> 해당 문항에 대해 오답을 제시할 경우, 불성실 응답으로 분류(Huang, Bowling et al., 2015; Meade & Craig, 2012)
	자기보고식 노력 측정 문항 (self-reported effort item)	<ul style="list-style-type: none"> 응답자에게 설문에 응답하며 얼마나 주의와 노력을 기울였는지 스스로 평가하도록 하는 문항(Curran, 2016; DeSimone & Harms, 2018; Meade & Craig, 2012) 	<ul style="list-style-type: none"> 해당 문항에 대한 응답 내용을 기준으로, 불성실 응답으로 분류(Curran, 2016; DeSimone & Harms, 2018; Meade & Craig, 2012)
2. 비개입적 방법 (unobtrusive methods)	응답시간 (response time)	<ul style="list-style-type: none"> 설문을 응답하는데 소요된 문항/페이지별 또는 전체 시간(DeSimone & Harms, 2018) 	<ul style="list-style-type: none"> 응답시간이 예상보다 지나치게 짧을 경우, 불성실 응답으로 분류(DeSimone & Harms, 2018)
	한 줄 응답 (long string/straight lining)	<ul style="list-style-type: none"> 연속된 문항들 혹은 동일 페이지 내 문항들에 대해 동일한 응답을 한 최대 횟수(DeSimone & Harms, 2018; Huang et al., 2012; Meade & Craig, 2012; Schonla & Toepoel, 2015) 	<ul style="list-style-type: none"> 문항 및 항목의 개수를 고려하여 비정상적인 패턴을 보이는 값을 찾는 것이 바람직함 5점 척도의 각각의 점수별로 6~14개의 연속된 문항에 동일한 응답을 할 경우, 불성실 응답을 의심(Costa & McCrae, 2008; Huang et al., 2012)하거나 9개의 연속된 문항에 동일한 응답을 할 경우, 불성실 응답을 의심(DeSimone & Harms, 2018)
	개인 응답 분산도 (individual response variability)	<ul style="list-style-type: none"> 특정 응답자가 보인 연속된 응답들 간의 표준편차(Dunn et al., 2018) 	<ul style="list-style-type: none"> 명확한 기준치를 사용하기 어려우나, 탐지할 응답의 비율에 따라 기준치를 설정하여 불성실 응답을 탐지한 후, 불성실 응답으로 탐지된 표본 중 개인 응답 분산도가 가장 낮은 표본의 불성실 응답 해당 여부를 살펴 기준치를 다시 조정(Dunn et al., 2018) 최소 25개 항목 이상 최대 150개 항목 이하의 응답을 사용할 것을 권장(Dunn et al., 2018)
3. 통계적 방법 (statistical methods)	마할라노비스 거리 (Mahalanobis distance)	<ul style="list-style-type: none"> 다변량 중심점으로부터 특정 데이터 간의 거리를 두 변수 간의 공변량으로 표준화한 값(Mahalanobis, 1936) 	<ul style="list-style-type: none"> 명확한 기준치가 존재하지 않음(Curran, 2016; Meade & Craig, 2012) 마할라노비스 거리 제곱값(D^2)이 다음에 해당 하는 경우, 불성실 응답으로 분류 <ul style="list-style-type: none"> - 1) 카이 제곱(χ^2) 분포의 상위 5% 이내일 경우(DeSimone & Harms, 2018) - 2) 카이 제곱(χ^2) 분포 상 유의수준 .001에서 동일한 모집단에서 추출되었다는 영가설이 기각되는 경우(Kline, 2015)
	개인 내 일관성 (individual reliability)	<ul style="list-style-type: none"> 설문을 임의로 두 개의 하위 척도로 구분한 후 각 척도에 대한 개인의 응답을 벡터로 생성하였을 때, 두 벡터 간의 상관 값 하위 척도 분할 방법: 홀짝(even-odd) 구분식, 반복추출(re-sampled) 구분식 	<ul style="list-style-type: none"> 0.3 미만일 경우, 불성실 응답으로 분류(DeSimone, Harms, & DeSimone, 2015; Johnson, 2005)
	동의/반의 일관성 (synonyms/antonyms consistency)	<ul style="list-style-type: none"> 동의/반의 관계에 있는 문항 쌍 들 사이의 개인별 응답의 상관관계 문항 쌍 선정방법: 의미론식(semantic), 심리측정식(psychometric) 	<ul style="list-style-type: none"> 0.3 미만일 경우, 불성실 응답으로 분류(DeSimone & Harms, 2018; Meade & Craig, 2012)

IV. 불성실 응답의 탐지와 제거효과의 실증 분석

4.1 연구문제

다양한 불성실 응답의 탐지방법에 따른 제거효과를 실증적으로 확인하기 위하여 본 연구에서는 분석에 앞서 다음과 같이 세 가지 구체적인 연구문제를 설정하였다.

첫째, 불성실 응답의 탐지방법에 따라 탐지되는 불성실 응답의 비율과 대상이 달라지는지 살펴본다. 존재 여부가 의심되나 구체적으로 추정되지 않았던 불성실 응답의 비율을 확인하는 것은 연구자들로 하여금 불성실 응답에 대한 유의성을 인식하고 적절한 대응을 고려하게 하는 계기가 될 것이다. 탐지 대상의 경우, 앞서 이론적 배경을 바탕으로 각 방법이 탐지하는 대상이 상이할 수 있음을 설명한 바 있으며, 이러한 예측에 따라 실제로 여러 탐지방법에 의하여 적발되는 대상이 상이한지 혹은 수렴하는지 검증해볼 필요성이 있기에 이를 첫 번째 연구 질문으로 설정하였다.

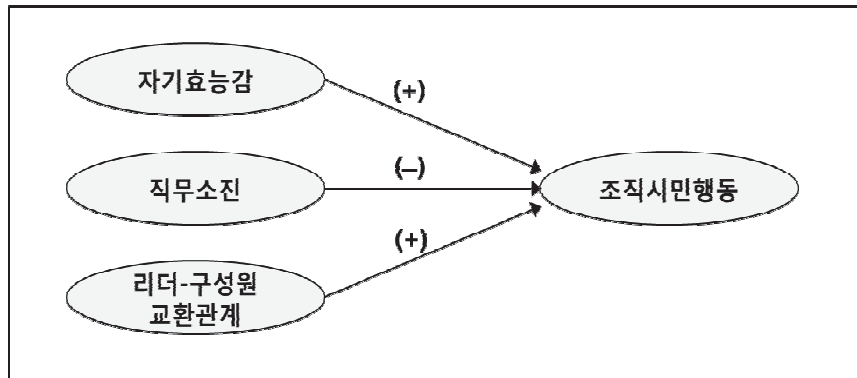
둘째, 불성실 응답이 의심되는 자료를 제거하는 것이 통계적 추론 결과에 어떤 영향을 미치는지 검증하고자 한다. 응답자가 설문 내용과 상관없는 부정확한 응답을 제공하는 경우, 연구자가 이를 적절히 탐지하고 대응하지 않는다면 연구결과가 왜곡될 위험성이 있다는 것이 다수의 선행 연구를 통해 밝혀져 왔다(Huang, Liu et al., 2015; Ward & Meade, 2018). 불성실 응답은 오차 분산, 상관관계, 요인분석 결과에 영향을 미쳐 최종적으로 통계적 가설검정에 영향을 미칠 수 있다(Meade & Craig, 2012), 다만, 그 효과의 크기가 제한적인지, 상당

한지에 대해서는 이견이 존재한다. 따라서 본 연구에서도 이러한 점을 검증하고자 한다.

셋째, 2개 이상의 탐지방법을 조합하는 다중 허들 방식(multi-hurdle)을 어떻게 적용할 수 있는지를 살펴본다. 다수의 연구들은 각 기법에 따라 탐지되는 불성실 응답의 성질이 상이함을 이유로 들어, 복수의 탐지방법을 조합하여 사용하는 다중 허들 방식을 사용할 것을 권고하고 있다(Curran, 2016; DeSimone et al., 2015; Dunn et al., 2018). 그러나 해당 방식의 사용 방법을 상세히 제시한 연구는 부족하여, 그 활용과정에서 연구자들이 적용에 혼동을 겪을 수 있다. 주의해야 할 사항은 복수의 탐지방법에 의해 동시에 불성실한 것으로 탐지되는 응답자를 걸러내는 것을 목표로 삼지 않아야 한다는 것이다. 한 가지 탐지방법, 즉 한 가지 허들을 사용해 극단적인 불성실 응답을 걸러낸 후, 남은 응답자를 대상으로 다시 새로운 허들을 적용하는 순차적(sequential)인 방식으로 적용하는 것이 핵심이며, 이를 실증연구 예시를 통해 보여주하고자 한다. 결론부에서는 보다 권장할 만한 방법인 사전적 방법과 사후적 방법을 조합한 다중 허들 방식에 관하여 보다 구체적으로 논하고자 한다.

4.2 불성실 응답의 영향분석을 위한 연구모형

불성실 응답 탐지방법의 이해와 활용을 돕고자 사회과학 분야에서 주로 설정하는 <그림 1>과 같은 가상의 연구모형을 구성하여 실증분석을 실시하고자 한다. 본 가상 연구모형은 대규모 설문조사를 통해 수집된 다수의 변수들 중에서 인사조직 분야에서 빈번히 논의되며 인과관계가 명확한 4개의 변수들을 선택하였다. 이는 불성실 응답의 영향 외에 변수의 개념, 인과관계, 측정도구의 타당성 등의 영향을 최



〈그림 1〉 불성실 응답의 영향분석을 위한 가상 연구모형

소화하기 위함이다. 가상모형의 독립변수로는 자기 효능감(self-efficacy), 직무소진(burnout), 리더-구성원 교환관계(LMX: leader-member exchange)를, 종속변수로는 조직시민행동(OCB: organizational citizenship behavior)을 선택하였다. 연구문제 2를 통해, 일반적인 실증연구의 전개순서를 따라가며 요인분석, 기술통계, 회귀분석 결과에 불성실 응답이 각각 어떠한 영향을 미치는지 검증하고자 한다.

4.3 자료수집과 분석방법

본 실증연구에서는 여러 하위 기업으로 구성된 한국의 대기업 집단 A에서 경영진단 목적으로 수집된 대규모 설문조사 자료 중 일부를 활용하였다. A 집단에서는 회사의 조직역량 강화와 시스템 경영을 위한 방향과 방법을 도출하고, 조직에 대한 구성원들의 전반적인 태도를 확인하고자 설문조사를 실시하였다. 설문조사는 크게 임원, 팀장, 팀원급으로 구분하여 진행되었고 기업집단 내 전수조사로 실시되었다. 전체 조사대상이 된 4,769명의 팀원 중 3,030명이 분석 가능한 유효답변을 제출하였고, 설문조사를 통해 측정된 다양한 개인 및 조직특성 변수들 중

위 가상 연구모형에 해당하는 4가지 변수들을 본 연구의 분석 대상으로 활용하였다. 본 자료는 수집 시 불성실 응답에 대한 직접적 탐지방법(예, 가짜문항 등)이 사전에 적용되지 않았으며, 비개입적 방법 중 문항에 대한 응답시간이 수집되지 않았으므로 이하의 분석대상에서 제외하였다.

해당 변수들은 기존 문헌을 통해 타당도와 신뢰성을 갖춘 측정도구를 활용하여 총 51개 문항의 7점 리커트 척도로 측정되었으며, 그 정의와 측정도구는 다음과 같다. 우선, 자기효능감은 개인의 인지적 자원을 활용하여 주어진 상황적 요구들을 충족시키기 위해 동기부여하고 행동할 수 있다고 믿는 개인의 신념을 의미한다(Wood & Bandura, 1989). 측정도구로는 Chen, Gully, & Eden(2001)의 일반적 자기효능감 척도(NGSE: new general self-efficacy scale)를 사용하였다. 예를 들어, 일반적 자기효능감 척도는 “나는 내가 세운 목표를 대부분 달성할 수 있다.”와 같이 본인의 목표 달성이나 어려운 상황에서도 극복할 수 있는지에 대해 측정하는 8개 문항으로 구성되어 있다. 다음으로, 직무소진은 사람을 대하는 일을 수행하는 사람들에게서 자주 발생하는 정서적 탈진과 낮은 성취감, 비인격화의 세 가지 특징

을 갖는 증후군을 일컫는다(Maslach & Jackson, 1981). 본 연구에서는 22개 문항으로 구성된 Maslach & Jackson(1981)의 직무소진 측정도구를 활용하였다. 이는 “나는 아침에 일어나서 일을 해야 한다는 생각이 들면 지친다.”와 같이 정서적 탈진을 측정하는 문항 9개, “나는 타인과의 문제에 효과적으로 대처한다.”와 같이 성취감을 측정하는 문항 8개, 그리고 “나는 다른 이들에게 어떤 일이 생기든지 전혀 신경 쓰지 않는다.”와 같이 비인격화를 측정하는 문항 5개로 구성되어 있다. 세 번째 독립변수인, 리더-구성원 교환관계(LMX)는 리더와 개별 구성원들 간 차별화된 사회적 교환관계를 의미하는데, 해당 관계의 질이 높을수록 상호호혜적인 결과를 가져오는 것으로 밝혀져 왔다(Graen & Uhl-Bien, 1995). 본 연구에서는 Scandura & Graen(1984)의 측정도구를 사용하였는데, 이는 “우리 팀장은 항상 나의 임무수행에 대해 만족한다.”와 같은 7개 문항으로 이루어져 있다. 마지막으로, 종속변수인 조직시민행동(OCB)은 조직의 구성원이 주변 구성원들 또는 속한 조직을 위하여, 공식적으로 주어진 과업의 범위를 넘어 수행하는 일련의 행위들을 의미한다(Brief & Motowidlo, 1986). 본 연구에서는 널리 활용되는 Williams & Anderson(1991)의 14개 문항을 사용하여 조직시민행동을 측정하였다. 이는 “나는 업무량이 많은 동료들을 돕는다.”와 같은 개인에 대한 조직시민행동(OCBI: organizational citizenship behavior-individual)을 측정하는 문항 7개, 그리고 “나는 질서를 유지하기 위한 비공식적인 규칙들을 준수한다.”와 같은 조직에 대한 조직시민행동(OCBO: organizational citizenship behavior-organizational)을 측정하는 문항 7개로 구성되어 있다.

본 연구에서는 설문조사가 진행되고 난 이후에 사

후적 탐지를 통해 불성실 응답을 확인하는 방법을 사용하였다. 구체적으로는 위의 <표 2>에 제시된 9가지 탐지방법 중 주어진 자료의 특성상 활용이 불가능한 3가지 직접적 방법과 응답시간을 제외한 총 5가지 탐지방법을 적용하였다. 비개입적 방법 중에는 (1)한 줄 응답과 (2)개인 응답 분산도를, 통계적 방법 중에서는 (3)마할라노비스 거리와 (4)홀짝 일관성, 그리고 (5)심리측정상 동의 일관성을 이용하여 불성실 응답을 측정하였다. 한 줄 응답의 경우, 문항의 전체 길이에 따라 불성실 응답으로 해석하는 기준의 설정이 달라지므로, 단순히 선행연구의 기준을 따르는 것은 권장되지 않는다. 따라서 한 줄 응답의 개인별 최댓값을 나타낸 빈도분포 그래프에서 나타나는 변곡점(scree)을 근거로 기준을 설정하였으며, 본 연구의 경우에는 총 51개의 문항 중 23개 이상을 연속으로 동일 응답한 경우를 불성실 응답으로 간주하였다. 개인 응답 분산도의 경우 DeSimone et al.(2015)와 Johnson(2005)의 기준에 따라 평균에 비해 2 표준편차 이상 큰 값을 보인 경우 불성실 응답으로 간주하였으며, 마할라노비스 거리 방법의 경우, Desimone & Harms(2018)의 기준보다 보수적인 .99의 신뢰수준으로 불성실 응답으로 분류하였고, 홀짝 일관성은 .30 미만에 해당하는 경우를 불성실 응답으로 분류하였다. 심리측정상 동의 일관성의 경우 문항 간 상관관계가 .60 이상인 경우 유사한 문항으로 추출하도록 했으며, 이렇게 추출된 문항의 응답으로 구성된 벡터 간 상관관계가 .30 미만에 해당하는 경우를 불성실 응답으로 분류하였다.

실증분석에는 R 3.5.2(R Core Team, 2018)와 AMOS 23.0, SPSS 26.0 프로그램을 활용하였고, 분석방법 및 절차의 구체적인 내용은 다음과 같다. 우선, R의 불성실 응답('careless' 버전 1.1.1) 패키지를 활용하여 개별 설문 응답자에 대한 불성실

응답의 수치들을 계산하였다(Yentes & Wilhelm, 2018). 연구자들은 해당 패키지에 탑재된 명령어를 사용하여 개인별 한 줄 응답의 최대 및 평균값, 개인 응답 분산도, 마할라노비스 거리, 홀짝 일관성, 심리 측정상 동의 및 반의 일관성을 측정할 수 있다. 각 개인들의 탐지방법별 결과 값을 도출한 후에는 사전에 설정된 기준에 따라 불성실 응답과 성실 응답을 분류하였다. 이후에는 확인적 요인분석과, 기술통계 분석, 회귀분석을 순차적으로 실시하였으며, 단일 또는 복수의 불성실 응답 탐지방법에 의하여 탐지되는 불성실 응답을 제외하기 전과 후를 비교하여, 통계적 추론 결과가 어떻게 변화하는지 분석하였다.

4.4 연구결과

4.4.1 연구문제 1: 불성실 응답의 탐지방법에 따라 탐지되는 불성실 응답의 비율과 대상이 달라지는가?

먼저, 본 실증연구의 첫 번째 연구문제인 ‘불성실 응답의 탐지방법에 따라 탐지되는 불성실 응답의 비율과 대상이 달라지는가?’를 살펴보았다.

〈표 3〉은 개별 불성실 응답 탐지방법을 적용하였을 때의 탐지 결과를 보여주고 있다. 요약하면, 적용된 탐지방법에 따라 최소 0.5%에서 최대 14%까지의 응답자들이 불성실 응답으로 탐지되어, 탐지방법에 따라 탐지비율의 편차가 존재하는 것으로 나타났다. 구체적으로, 한 줄 응답방식에 따르면 총 3,030명 중 16명이 불성실 응답자로 탐지되어 가장 낮은 탐지 비율을 나타내었다(0.53%). 반면, 마할라노비스 거리 방법을 적용한 경우 탐지 비율이 가장 높게 나타났는데, 이때의 불성실 응답자 수는 422명(13.93%)으로 나타났다.

〈표 4〉는 5가지 방법을 동시에 적용하였을 때의 탐지 결과를 보여주고 있다. 불성실 응답으로 탐지된 횟수에 따른 응답자 분류를 살펴보면 5개 탐지방법 모두에서 불성실 응답으로 탐지되지 않은 응답자의 수는 2,316명(76.44%)으로 나타났다. 이는 온전히 성실한 응답자의 수로 볼 수도 있으나, 각 방법이 지닌 한계를 고려할 때 단순히 1회 탐지된 응답자들은 불성실 응답으로 단정해서는 안 되므로 잠정적인 비율로만 고려될 필요가 있다. 불성실 응답으로 1회 탐지된 응답자의 수는 586명(19.34%)이었으며, 1회 이하로 탐지된 인원의 합산은 2,902명

〈표 3〉 탐지방법별 기술통계량 및 불성실 응답의 탐지비율

구분	기법	평균	표준 편차	최솟값	최댓값	구분 기준치	탐지된 불성실 응답자 수	탐지된 불성실 응답 비율 ¹⁾
비개입적 방법	한 줄 응답 ²⁾	7.67	3.99	2.00	47.00	23개 이상	16	0.53%
	개인 응답 분산도	1.52	.45	.20	3.03	2.41 이상	91	3.00%
통계적 방법	마할라노비스 거리	50.98	29.09	8.82	342.43	.99 신뢰도	422	13.93%
	홀짝 일관성	.92	.16	-1.00	1.00	.3 미만	35	1.16%
	심리측정상 동의 일관성	.66	.24	-.35	1.00	.3 미만	281	9.27%

1) 전체 응답자 3,030명 중 해당 비율

2) 전체 문항 51개 중 연속으로 동일 응답한 문항 수

〈표 4〉 불성실 응답으로 탐지된 횟수별 응답자 분류

불성실 응답으로 탐지된 횟수 ¹⁾	탐지된 응답자의 수(명)	탐지된 응답자의 비율
0	2,316	76.44%
1	586	19.34%
2	125	4.13%
3	3	0.10%
4	0	0.00%
5	0	0.00%
합계	3,030	100.00%

1) 동일한 응답자가 5가지 방법 중 몇 가지 방법에 의하여 동시에 불성실 응답으로 분류되었는지 나타냄

(95.78%)이었다.

주목해야 할 것은 2회 이상(최소 2회~최대 5회) 동시에 탐지된 응답자의 총합이 128명(4.23%)으로 작은 비중을 차지한다는 점이다. 이는 특정 불성실 응답자가 서로 다른 방법에 의해서 동시에 탐지되지 않다는 사실을 시사하며, 동시에 각 탐지방법이 탐지하고자 하는 불성실 응답의 대상이 수렴하지 않음을 나타낸다. 이는 앞서 살펴본 이론적 배경과 일치하는 결과이다.

4.4.2 연구문제 2: 불성실 응답이 의심되는 자료를 제거하는 것은 통계적 추론 결과에 어떤 영향을 미치는가?

본 연구의 두 번째 연구문제인 ‘불성실 응답이 의심되는 자료를 제거하는 것은 통계적 추론 결과에 어떤 영향을 미치는가?’를 살펴본다. 기존 자료와 각각의 불성실 응답 탐지방법을 이용해 불성실 응답이 의심되는 응답을 제거한 자료를 활용하여 요인분석과 상관분석, 회귀분석을 실시하고 분석결과를 비교하였다.

우선, 불성실 응답의 제거가 변수의 하위요소 구성 타당도에 미치는 영향을 살펴보기에 앞서 기존 자료

의 확인적 요인분석 결과를 〈표 5〉, 〈표 6〉에 제시하였다. 기존 자료의 확인적 요인분석 모델 적합도는 CMIN/DF가 27.268, GFI가 .857, RMSEA가 .093, NFI가 .895, IFI가 .898, TLI가 .882, CFI가 .898로 확인되었다. 기존 자료의 확인적 요인분석 결과에 따르면, 기존 자료의 자기효능감, 직무소진, 리더-구성원 교환관계와 조직시민행동의 모든 측정지표들과 구성개념 간의 표준화계수들도 .60 이상($p < .001$)으로 분석하기에 적합한 수치를 보였다(Bagozzi, Yi, & Phillips, 1991). 구조방정식 모형 분석은 표본크기에 비해 추정모수의 수가 커질수록 전체 모형 적합도가 떨어져서 추정값이 적절하지 않을 수 있다(Kline, 2015). 직무소진과 조직시민행동은 항목목록을 활용하여 요인분석을 실시하였는데, 항목목록은 개별문항을 바탕으로 한 분석에 비해 추정모수의 수가 적다는 점에서 훨씬 간명하며 표집 오차로 인한 문제들을 완화할 수 있으며, 오차 간의 상관이나 이중 요인 부하가 발생할 가능성이 적은 이점들이 있다(Little, Cunningham, Shahar, & Widaman, 2002).

본 연구에서 선정된 적합도 지수는 χ^2 , CMIN/DF, GFI, RMSEA, NFI, IFI, TLI, CFI로 불성실 응답 탐지방법별 확인적 요인분석 모델 적합도는

〈표 5〉 기존 자료의 확인적 요인분석 결과

분류	Cronbach's α	문항	기존 자료			
			표준화계수 요인부하량	Estimate	S.E	C.R
자기 효능감	.945	1	.726	1.000	-	-
		2	.849	1.096***	.023	47.059
		3	.844	1.042***	.022	46.763
		4	.831	1.113***	.024	45.976
		5	.859	1.083***	.023	47.672
		6	.878	1.152***	.024	48.771
		7	.804	1.009***	.023	44.432
		8	.821	1.081***	.024	45.409
리더- 구성원 교환 관계	.910	1	.582	1.000	-	-
		2	.797	1.714***	.052	33.272
		3	.550	1.063***	.041	25.703
		4	.839	1.824***	.053	34.307
		5	.859	1.956***	.056	34.769
		6	.856	1.816***	.052	34.699
		7	.873	1.835***	.052	35.087
직무 소진	.908	정서적 탈진	.629	1.000	-	-
		성취감	.717	0.882***	.031	28.492
		비인격화	.638	1.067***	.040	26.602
조직시민 행동	.863	OCBI	.727	1.000	-	-
		OCBO	.665	0.912***	.032	28.479

1) *** $p < .001$

〈표 6〉에 제시하였다. 이 때, χ^2 검정은 통계적 유의성을 제공하기는 하지만 사례 수와 관찰변인들이 다변량 정규분포를 따라야 한다는 가정에 영향을 받는다. 특히 사례 수가 500을 넘어가게 되면 무조건 귀무가설을 기각하게 되는 데, 본 연구의 사례 수는 3,030개 이므로 χ^2 검정은 해석하지 않았다. 이외의 모든 적합도 지수는 본 모형이 연구의 자료를 잘 설명하도록 제시되었다. 이외의 적합도 지수가 모두 수용 가능한 것으로 나타나 모형의 적합도가 기준을 충족시켰고 연구 모형이 표집된 데이터 분석을 위한

적합 모형임을 알 수 있었다(Kline, 2015).

분석결과를 요약하면, 불성실 탐지방법의 적용은 확인적 요인분석 결과에 긍정적인 영향을 미치는 것으로 나타났다. 다섯 가지 탐지방법을 활용하였을 때, CMIN/DF 수치가 모두 개선되었지만, 마할라노비스 방법과 홀짝 일관성 방법에서 그 감소 폭이 비교적 컸다. 마할라노비스 거리 방법을 적용하여 데이터를 제거한 이후의 확인적 모형 분석결과, CMIN/DF는 23.837, GFI는 .854, RMSEA는 .094, NFI는 .900, IFI는 .904, TLI는 .889, CFI는

〈표 6〉 기존 자료와 불성실 응답 제거자료의 확인적 요인분석 모형 적합도 비교

구분	적합도 판단기준	기존 자료	한 줄 응답	개인 응답 분산도	마할라노비스 거리	홀짜 일관성	심리측정상 동의 일관성
절대 적합 기준	χ^2	$p > .05$.000	.000	.000	.000	.000
	CMIN/DF	낮을수록 좋음	27.268	27.119	26.098	23.837	22.167
	GFI	$\geq .90$.857	.857	.857	.854	.878
	RMSEA	$\leq .1$.093	.093	.092	.094	.084
증분 적합 지수	NFI	$\geq .90$.895	.895	.891	.900	.923
	IFI	$\geq .90$.898	.898	.895	.904	.926
	TLI	$\geq .90$.882	.882	.878	.889	.914
	CFI	$\geq .90$.898	.898	.895	.904	.926

.904로 보고되었다. 홀짜 일관성 탐지방법을 적용하여 데이터를 제거한 이후의 확인적 모형 분석결과, CMIN/DF는 22.167, GFI는 .878, RMSEA는 .084, NFI는 .923, IFI는 .926, TLI는 .914, CFI는 .926로 보고되었다.

또한, 불성실 응답의 제거가 변수 간 상관관계에 미치는 영향을 살펴보기 위해 기존 자료와 각각의 불성실 응답 탐지방법을 사용하여 불성실 응답으로 의심되는 응답을 제거한 자료의 상관분석을 실시하였다(〈표 7〉 참조). 예상대로 기존 자료에서 자기효능감, 리더-구성원 교환관계는 조직시민행동과 정적 상관관계를, 직무소진은 부적 상관관계를 나타냈다. 불성실 응답 탐지방법을 적용한 자료에서는 관계의 방향이 크게 달라지지는 않았으나, 그 강도가 개인 응답 분산도, 심리측정상 동의 일관성의 경우 기존 자료보다 낮게 나타났다. 그러나 마할라노비스 거리를 사용하였을 때에는 상관관계의 강도가 기존 자료나 다른 탐지방법을 적용하였을 때보다 높게 나타나는 것을 확인할 수 있었다.

마지막으로, 불성실 응답의 제거가 변수 간의 영향관계와 회귀모형의 설명력에 미치는 영향을 살펴

보기 위해, 본 연구에서 제시한 연구모형을 바탕으로 회귀분석을 실시하였다(〈표 8〉 참조). 먼저, 기존 자료에서 자기효능감은 조직시민행동에 통계적으로 유의한 정적 영향을 미치는 것으로 나타났다($\beta = .38, p < .001$). 직무소진은 조직시민행동에 유의한 부적 영향을 미쳤으며($\beta = -.26, p < .001$), 리더-구성원 교환관계는 조직시민행동에 유의한 정적 영향을 주었다($\beta = .09, p < .001$). 불성실 응답을 제거한 자료에서도 자기효능감과 리더-구성원 교환관계는 조직시민행동에 유의한 정적 영향을 미치고 직무소진은 조직시민행동에 유의한 부적영향을 미치는 것으로 나타났다. 탐지방법에 따른 표준화 회귀계수 변동의 폭은 크지 않았다. 회귀모형 전체가 지닌 설명력은 기존 자료의 경우 .36으로 나타났다. 불성실 응답 탐지를 통해 불성실 응답을 제거한 자료의 경우, 설명력이 기존 자료에서보다 다소 낮게 나타났다. 그러나 마할라노비스 거리를 이용해 불성실 응답을 제거한 자료의 경우, 설명력이 .41로 나타나 기존 자료에 비해 설명력이 증가하는 것을 확인할 수 있었다.

〈표 7〉 기존 자료와 불성실 응답 제거자료의 기술통계량 및 상관관계 비교¹⁾

구분	변수	평균	표준편차	자기효능감	직무소진	리더-구성원 교환관계	
기존 자료	자기효능감	5.40	.86				
	직무소진	3.30	.74	-.47			
	리더-구성원 교환관계	4.50	1.11	.26	-.39		
	조직시민행동	5.44	.64	.53	-.48	.30	
비 개 입 적 방 법	한 줄 응답	자기효능감	5.40	.85			
		직무소진	3.30	.74	-.47		
		리더-구성원 교환관계	4.50	1.11	.26	-.39	
		조직시민행동	5.44	.63	.53	-.48	.30
	개인 응답 분산도	자기효능감	5.36	.83			
		직무소진	3.34	.70	-.44		
		리더-구성원 교환관계	4.48	1.08	.25	-.37	
		조직시민행동	5.41	.62	.50	-.44	.29
통 계 적 방 법	마할라 노비스 거리	자기효능감	5.39	.81			
		직무소진	3.26	.70	-.51		
		리더-구성원 교환관계	4.59	1.03	.30	-.40	
		조직시민행동	5.44	.63	.57	-.52	.34
	홀짝 일관성	자기효능감	5.41	.85			
		직무소진	3.29	.74	-.47		
		리더-구성원 교환관계	4.51	1.12	.26	-.39	
		조직시민행동	5.45	.62	.51	-.47	.29
	심리 측정상 동의 일관성	자기효능감	5.40	.83			
		직무소진	3.24	.73	-.44		
		리더-구성원 교환관계	4.52	1.15	.25	-.41	
		조직시민행동	5.50	.60	.47	-.46	.29

1) 모든 상관계수는 $p < .01$ 수준에서 유의함

4.4.3 연구문제 3: 2개 이상의 탐지방법을 조합하는 다중 허들방식은 어떻게 적용할 수 있는가?

〈표 9〉는 본 연구자료에 2단계 다중 허들방식을 적용한 예시적 결과이다. 본 결과는 다중 허들방식이 2개 이상의 탐지방법의 교(交)집합이 아닌 합(合)

집합에 의해서 탐지되는 응답자들을 제외하는 원리로 작동됨을 안내하고자 작성된 것이다. 연구문제 1을 통하여 각 탐지방법이 탐지하는 대상이 서로 상이함을 확인한 바 있다. 따라서 2개 이상의 탐지방법에 의하여 동시에 탐지된 대상자를 색출하는 것은 바람직한 접근이라고 보기 어렵다. 대신, 사전에 설정한 순서에 따라 특정 기준을 통과한 대상자들을

〈표 8〉 기존 자료와 불성실 응답 제거자료의 회귀분석 결과 비교¹⁾

구분	탐지방법	설명력(R제곱)			변수	표준화 회귀계수 ²⁾			
		기존	제거 후	변화량		기존	제거 후		
기존 자료		.36			자기효능감	.38			
					직무소진	-.26			
					리더-구성원 교환관계	.09			
비개입적 방법	한 줄 응답		.35	-.00	자기효능감		.38		
							직무소진		-.26
							리더-구성원 교환관계		.09
	개인 응답 분산도		.32	-.03	자기효능감		.37		
							직무소진		-.23
							리더-구성원 교환관계		.11
통계적 방법	마할라노비스 거리		.41	.05	자기효능감		.39		
							직무소진		-.28
							리더-구성원 교환관계		.11
	홀짝 일관성		.34	-.02	자기효능감		.37		
							직무소진		-.26
							리더-구성원 교환관계		.10
	심리측정상 동의 일관성		.31	-.05	자기효능감		.33		
							직무소진		-.27
							리더-구성원 교환관계		.10

1) 종속변수: 조직시민행동

2) 모든 계수는 $p < .001$ 수준에서 유의함

〈표 9〉 다중 허들방식의 적용 결과

(단위: 명)

조합순	1차 허들	1차생존	1차탈락	2차허들	2차생존	2차탈락	1+2차탈락	전체탈락비율
(1)	한 줄 응답	3,014	16	개인응답분산	2,923	91	107	3.53%
(2)	한 줄 응답	3,014	16	마할라노비스	2,594	420	436	14.39%
(3)	한 줄 응답	3,014	16	홀짝 일관성	2,980	34	50	1.65%
(4)	한 줄 응답	3,014	16	동의 일관성	2,741	273	289	9.54%
(5)	개인응답분산	2,939	91	마할라노비스	2,554	385	476	15.71%
(6)	개인응답분산	2,939	91	홀짝 일관성	2,904	35	126	4.16%
(7)	개인응답분산	2,939	91	동의 일관성	2,658	281	372	12.28%
(8)	마할라노비스	2,608	422	홀짝 일관성	2,575	33	455	15.02%
(9)	마할라노비스	2,608	422	동의 일관성	2,380	228	650	21.45%
(10)	홀짝 일관성	2,995	35	동의 일관성	2,745	250	285	9.41%

1) 최초 샘플 수 3,030명

다음 기준에 의하여 탈락시키는 순차적인(sequential) 접근을 취하는 것이 권장된다(Curran, 2016).

다중 허들 방식을 적용한 경우, 한 가지 방식을 적용하였을 때보다 불성실 응답으로 분류되는 인원이 소폭 증가했다(최소 1.65%에서 최대 21.45%). 단, 합집합 방식으로 적용되었기 때문에, 위의 <표 3>에서 나타난 두 가지 개별 방식을 통해 탐지된 인원의 단순 합산보다는 적은 인원이 탐지되었다. 예를 들어, (4)번 조합의 한 줄 응답과 동의 일관성 방식을 다중 허들로 적용한 경우, 단순히 두 방식에 의하여 각각 탐지된 인원들을 합산한 숫자(각 16명, 281명으로 합계 297명)에 비하여 적은 인원이(289명)이 탐지되게 된다.

V. 논의 및 결론

본 연구는 설문조사에서 나타나는 불성실 응답의 개념과 유형을 체계적으로 정리하고, 이에 대응할 수 있는 다양한 탐지방법들을 소개한 후, 대규모 설문 데이터를 활용하여 각 탐지방법의 적용 결과와 연구모형에 미치는 영향들을 검토하였다. 다음에서는 연구결과가 주는 함의와 이를 향후 연구에 활용함에 있어 주의를 기울여야 할 사항들을 논하고자 한다. 이러한 논의가 향후의 설문조사 기반 연구에서 불성실 응답에 효과적으로 대처하고, 각 연구의 신뢰도와 정확성을 높일 수 있는 계기가 되기를 기대한다.

먼저, 연구문제 1을 통해서 불성실 응답 탐지방법의 비율과 대상에 대하여 살펴본 바, 본 연구 대상 자료에서는 최소 0.5%~최대 14%의 불성실 응답을 확인하였다. 이러한 결과를 접하는 독자들은 해

당 비율이 너무 높은 것이 아닌지, 해당 방식에 의해 불성실 응답으로 탐지된 모든 응답자들을 탈락시키는 것은 자료 수에 지나친 손실이 아닌지 우려할 수 있다. 우선, 본 실증 연구결과와 비율은 기존 문헌에서 보고된 3~12%의 수치와 크게 상이하지 않다(Johnson, 2005; Maniaci & Rogge, 2014; Meade & Craig, 2012). 또한, 단일 방법에 의하여 탐지되었다고 하여 무조건 탈락의 대상으로 고려해서는 안 된다(Curran, 2016). 특히, 본 연구의 자료는 기업 관리차원에서 하향식(top-down)으로 이루어져 참여자들의 성실 응답에 대한 동기가 부족할 수 있으며, 수많은 변수를 동시에 측정하였다는 점에서 피로도가 높아 불성실 응답이 일반적인 연구에 비하여 과대하게 추정되었을 가능성이 있다. 이러한 연구맥락에도 불구하고, 성실 응답자일 확률이 높은 응답자들, 즉, 1개 이하의 탐지방법에 의하여 탐지된 인원은 95.78%의 비중을 차지했다. 이러한 결과는 자율적 참여에 기반하고 소수의 관심 변수에 집중한 연구에서 보수적인 다중허들 방식을 적용하였을 때 연구자가 감수해야할 표본 탈락의 비중이 높지 않을 수 있다는 점을 시사한다.

다음으로, 연구문제 2를 통해 불성실 응답의 제거가 통계적 추론결과에 미치는 영향을 확인한 결과, 모형의 설명력이나 요인분석 모형의 적합도에서 변화가 있었으나 이전 연구에서 추정된 정도로는 크지 않은 것으로 나타났다. 그 중 마할라노비스 거리 방법이 요인구조의 모형 적합도를 높여주고, 회귀분석의 설명력을 소폭 높이는 것으로 나타났으나 그 변화폭은 최대 5%수준에 그쳤다. 이는 불성실 응답을 실증적으로 연구한 기존 문헌에서 불성실 응답의 영향력에 다소 제한이 있음을 주장하는 결과와 맥락을 같이 한다(Gallen & Berry, 1996; Meade & Craig, 2012; Breitsohl & Steidelmüller, 2018).

다만, 얻을 수 있는 통계적 효과의 크기가 상당하지 않다고 하여, 불성실 응답의 존재를 인식하고 있으며 적절한 대처방안이 존재함에도 불구하고 아무런 조치도 취하지 않는 것은 2중 오류를 극대화할 가능성이 높으므로 바람직한 선택이 아니다. 연구자들은 설문 결과의 코딩과정에서 지나치게 같은 응답이 반복되거나 남들과 다른 응답 패턴이 보일 경우, 해당 응답이 불성실 응답임을 직관적으로 알 수 있다. 기존에는 연구자들이 불성실 응답을 인식하고도 명확한 탈락기준이 없어서 넘어가야 했다면, 본 연구에서 소개한 방법을 적용하여 사전에 설정한 기준에 따라 이러한 응답을 걸러내는 것은 변수들의 측정 타당도와 연구의 신뢰성을 높이고, 설명하고자 하는 관계를 명확히 하는데 큰 도움이 될 것이다.

마지막으로, 연구문제 3을 통하여 다중 허들방식의 적용 사례를 보여주었다. 본 연구에서는 예시적으로 5가지 통계적 탐지방법을 활용하여 조합 가능한 10가지의 2단계 다중 허들 방식을 적용한 결과를 모두 보여주었으며, 실제 연구자들은 자신의 연구의 맥락에 따라 한 가지 조합만을 선택하여 활용할 수 있을 것이다. 다만, 본 연구에 나타난 허들의 조합들은 모두 비개입적, 통계적 방식 간의 조합이기에 그 효과성과 의미가 상당히 제한됨을 반드시 유의해야 한다. 향후, 연구자들이 다중 허들방식을 실제 연구에 적용할 때는 연구 설계 단계에서 미리 고려해야 하는 사전적 방식(예, 가짜문항)과 본 연구에서 활용된 사후적 방식(예, 한 줄 문항)을 조합하여 다중 허들방식을 구성한다면, 보다 효과적으로 불성실 응답을 탐지해낼 수 있을 것이다.

직접적 또는 사전적 방식은 본 실증연구에서 다른 통계적인 접근 방식들에 비하여, 실제 불성실 응답이 아님에도 불성실 응답으로 탐지할 가능성이 낮다. 통계적 방식에서는 특정 응답자가 특이값을 나

타내더라도, 불성실한 응답 태도 때문이 아니라 문항구성의 특징이나 실제 해당 응답자의 개인특성의 특이성으로 인해 그러한 결과가 나왔을 가능성이 존재한다. 그러나 특정 답변을 고르라는 질문에 엉뚱한 답변을 하거나(예, 지시적 조작 점검 문항), 응답자 스스로 본 조사를 성실하지 않게 응답하였다고 답하는 경우(예, 자기보고식 노력측정 문항), 불성실 응답에 해당할 가능성이 매우 크다. 또한, 이러한 방식은 복잡한 사후적 통계분석에 비하여 연구자가 이해하고 수행하기에 직관적이며, 설문 참여자 입장에서 몇 개의 문항에만 추가적으로 응답하면 되는 것이기에 쉽게 활용할 수 있다. 따라서 향후 연구자들이 이를 연구 설계 단계에서 적극 고려하여 활용할 수 있도록, 다음의 <표 10>에 활용 예시를 정리하였다. 총 3가지 방법과 예시문항이 제시되어 있으나, 이 중 활용할 방법과 문항의 수는 문헌조사를 거쳐 연구의 맥락과 연구자의 재량에 따라 정할 수 있다.

특히, 모든 연구 상황에서 공통적으로 유의해야 할 점은 불성실 응답 탐지기준의 설정에 관한 사항이다. 기존의 연구들 또한 구체적인 기준을 언급함에 있어서 상당히 신중한 태도를 취하고 있다. 이는 각 탐지방법이 설문문항의 구성, 내용 및 길이에 의하여 많은 영향을 받기 때문이다. 예를 들어, 한 줄 응답 방법은 연구 설문 문항이 길어질수록 자연스럽게 그 평균값이 높아지기 때문에 연구자 입장에서 불성실 응답의 특정한 기준치를 설정하기가 어렵다. 유연한 기준을 적용할 경우 귀중한 시간과 자원을 들여 모은 자료 중 상당수가 실제 불성실 응답이 아님에도 불성실 응답으로 판별되어 탈락하는 1중 오류가 발생할 수 있다. 또한, 두 개 이상의 방법을 활용하여 다중 허들 방식을 적용할 경우에는 과도한 자료의 탈락을 방지하기 위하여 더욱 극단에 가까운, 보수적인 탐지기준의 설정이 요구된다. 따라서

〈표 10〉 직접적(=사전적) 불성실 응답 탐지방법 방법의 예시

출처	문항 예시 (일부 발췌)	활용예시
1. 지시적 조작 점검 문항(instructional manipulation check item)		
DeSimone & Harms (2018)	‘이 문항에 대해서는 (3번 답안)을 선택하세요.’	<ul style="list-style-type: none"> • 지시적 조작 점검 문항과 가짜문항을 합쳐서 총 3개 내외의 문항을 설문에 골고루 배치 • 해당 문항에 대해 지시된 답변을 고르지 않은 경우, 오답으로 간주 • 복수의 문항 중 오답이 몇 개 이상일 때 불성실 응답으로 간주할지는 연구자가 사전검토 및 설정
Breitsohl & Steidelmüller (2018)	‘이 문항에 대해서는 (매우 동의하지 않음)을 선택하세요.’	
2. 가짜문항(bogus item)		
Meade & Craig (2012)	‘나는 현재 컴퓨터를 사용 중이다.’	<ul style="list-style-type: none"> • 연구의 지시사항(instruction)에 응답의 성실성을 체크하기 위한 일부 문항들이 포함되어 있음을 미리 표기하기를 권장 • 연구와 문화적 맥락에 따라, 의도하지 않은 해석이 발생할 확률이 가장 적은 문항들을 취사선택하여 사용 • 예로, 7점 척도 중, 긍정의 답을 요하는 경우 6-7점(동의 또는 매우 동의의 경우)에만 성실응답으로 간주. 나머지 1-5점을 선택한 경우 모두 오답으로 간주 • 복수의 문항 중 오답이 몇 개 이상일 때 불성실 응답으로 간주할지는 연구자가 사전검토 및 설정
	‘나는 전 세계 모든 나라를 방문해보았다.’	
	‘나는 하루에 잠을 1시간도 자지 않는다.’	
	‘나는 한국어를 한 마디도 알아듣지 못한다.’	
DeSimone & Harms (2018)	‘나는 정확히 354명의 절친한 친구가 있다’	
Maniaci & Rogge (2014)	‘나는 칭찬받으면 기분이 좋다.’	
	‘나는 증오받기보다 사랑받는 것이 좋다.’	
	‘나는 놀림을 받는 것이 싫다.’	
	‘나는 텔레마케팅 전화 받는 것을 즐긴다.’	
	‘나는 복권에 당첨되어도 기쁘지 않을 것 같다.’	
3. 자기보고식 노력 측정 문항(self-reported effort item)		
Meade & Craig (2012)	‘답변 중에 TV, 휴대폰 등 다양한 방해요인이 있을 수 있습니다. 당신은 이 연구에 얼마만큼의 주의를 기울였나요?’	<ul style="list-style-type: none"> • 설문의 맨 마지막에 제시 • 답변내용에 상관없이 연구에 대한 피드백이나 보상 등은 제공됨을 명시하여 솔직한 답변 유도 • 잘못된 데이터의 포함은 연구에 참여한 사람들과 연구자들의 시간과 자원이 낭비될 수 있음을 함께 언급함
	‘솔직하게 답변하여, 우리가 당신의 데이터를 분석에 활용해도 괜찮습니까?’	

‘정상적인 응답이라면 절대 나타날 수 없는’ 값을 기준으로 삼는 것이 권장된다. Curran(2016)은 자료의 손실을 최소화하면서도 불성실 응답을 걸러내는 방안으로, 가장 눈에 띄게 극단적인 불성실 응답의 형태를 보인 응답자들을 탈락시키는 방식을 권장하였다. 예를 들어, 전체 문항의 60%를 연속으로

한 줄 응답한 경우, 문항 당 응답시간이 0.5초보다도 짧은 경우 등으로 보수적인 기준을 설정하여 적용할 수 있을 것이다.

본 연구의 목적은 설문조사의 신뢰도와 타당도를 위협할 수 있는 불성실 응답에 대한 연구자들의 인식을 제고하고 탐지뿐만 아니라 해결방안을 활용하

는 것이다. 이어서 본 연구의 결론을 토대로 향후의 연구방향을 제시하면 다음과 같다.

첫째, 불성실 응답에 대한 국내 연구자들의 보다 많은 관심과 이해가 요구된다. 이는 연구방법론에 있어서 인식이 보편화된 동일방법편의(common method bias)(박원우·김미숙·정상명·허규만, 2007)에 대한 논의와 같이, 연구자들 사이에 관심이 확대되고 공감대가 형성될 경우 사전 설계단계에서부터의 고령을 통해 보다 효과적으로 극복할 수 있는 문제이기 때문이다. 또한 본 연구에 활용된 선행연구들은 서구 문화와 맥락을 바탕으로 하지만, 이는 보편적 현상이기 때문에 국내에서도 유의한 영향력과 시사점을 갖는지, 문화적 맥락에 따라 다른 결과가 나타나는지 등에 관하여 후속연구를 통해 반복적인 검증과 토의가 이뤄질 필요성이 있다.

둘째, 본 연구에서는 이미 발생한 불성실 응답에 대한 탐지방법을 제시하였으나, 불성실 응답 발생을 사전에 예방할 수 있는 방법에 대한 답은 제시하지 않고 있다. 이는 매우 광범위한 주제로 단일 연구에서 함께 다루기에는 다소 한계가 있었지만, 향후 실용적 측면에서 높은 가치가 있는 주제이다. 예를 들어, 응답자들이 느끼는 피로에 의하여 불성실 응답이 발생한다고 여겨지는 경우가 많은데(Liu et al., 2013), 설문지의 길이를 줄이는 것이 가장 효과적인 예방책이지만 이는 상황에 따라서는 적용이 어려울 수 있다. 대신, 설문지의 지시사항에 경고문구들을 삽입하거나, 시각적 요소를 통해 주의를 환기시키거나, 금전적 인센티브를 제공하는 등의 대안적 장치들의 효과를 검증한 후 활용할 수 있을 것이다(Deutskens et al., 2004).

셋째, 기업 현장에서 불성실 응답의 영향에 대한 논의는 거의 진행되지 않았기에, 불성실 응답이 기업의 데이터에 어떤 영향을 미치는지 논의가 필요하

다. 최근 기업들은 인적자원정보시스템 분석의 효율성을 극대화하기 위해 분석에 포함시키는 데이터의 종류를 다양화(예, 인구통계자료, 직무와 경력, 자격증, 성과자료, 근태 자료, 사내 게시판의 답글 및 구성원 설문조사)하고 있으며 데이터의 종류 뿐 아니라 질적인 면에서의 중요성도 강조하는 추세이다(Guenole, Ferrar, & Feinzig, 2017). 그러나 기업에서 아무리 최신의 방법을 활용하여 데이터를 수집한다고 할지라도, 불성실 응답으로 인한 데이터 품질의 저하가 발생할 수 있다. 따라서 HR 분석 실무자는 설문의 측정과 분석을 좀 더 세밀하고 정확하게 진행해야 하며 또한 분석 프로세스의 타당성을 확보하여 의사결정자가 의사결정을 올바르게 할 수 있는 시스템을 마련해야 한다.

본 연구는 문헌조사와 실증연구에 기초하여 유의미한 발견과 연구방향들을 제시하였으나, 동시에 한계점도 일부 지니고 있다. 우선, 실증분석에 활용된 데이터가 단일 대기업 집단의 샘플로서, 해당 기업 집단의 문화나 특성에 의하여 편중된 영향을 받았을 가능성이 있다. 하지만 샘플 내에 다양한 산업군의 회사와 직급, 직무를 포괄했다는 점에서 오히려 일반화 가능성이 적지 않다고 볼 수 있다. 또한, 연구의 주된 목표는 불성실 응답의 영향을 관찰하는 것으로, 해당 기업집단의 특성이 응답의 내용적 측면에는 영향을 주었을 수 있으나, 불성실 응답의 패턴이나 통계적 비율은 연구 맥락에 상관없이 고르게 나타나는 경향이 있으므로 연구의 결론에 크게 영향을 주지 않았으리라 추정할 수 있다.

둘째, 문헌연구 결과를 바탕으로 효과적인 사전적 탐지방법의 적용을 적극 제안하였으나, 그 효과를 직접 검증하지 못하였다. 사전적 탐지방법은 불성실 응답에 대한 사전적 이해와 계획을 바탕으로 연구 설계 시 고려되어야 하는 것으로 연구자가 불성실 응

답의 의미와 영향력, 탐지방법에 대한 이론적 배경을 이해하고 있을 때에 사용 가능한 방법이다. 따라서 향후 연구자들은 이러한 사전적 탐지방법을 설문조사 실시 전에 고려한다면 자신의 연구모형 내의 변수 간 관계뿐만 아니라 동시에 불성실 응답의 효과도 연구할 수 있는 기회를 얻게 될 것이다.

셋째, 불성실 탐지방법에 있어서 본 연구에서 정리 및 제시된 방법들이 모든 방법을 포괄하고 있는 것은 아니다. 근래 논의가 이루어지고 있는 다항 거트만 방식(polytomous Guttman error)(Niessen et al., 2016)을 비롯하여 기타 다양한 탐지방법들도 존재한다. 단, 본 연구는 불성실 응답의 개념을 소개하려는 목적으로, 이해와 활용이 비교적 용이하며 기존 문헌에서 주로 활용된 방법에 집중하였다. 향후, 기타 분석방법 또는 연구맥락(예, 오프라인 조사) 등에 있어 본 연구에서 미처 다루지 못한 보다 넓은 범위의 연구가 이어질 필요성이 있다.

설문조사는 단연 응용심리학 분야에서 가장 폭넓게 사용되는 연구방법이다(Breitsohl & Steidelmüller, 2018). 그럼에도 불구하고, 경영학 전반, 특히 인사조직 연구 및 마케팅 관련 학술지에서 불성실 응답에 관한 충분한 연구를 찾아볼 수 없다는 점은 그 범용성과 활용도를 고려할 때 보완이 필요한 부분이다. 일부 심리측정 분야 연구자들을 중심으로 유의미한 발전이 진행되고 있음에도 정작 이를 활용할 응용 과학자들에게 전파되어 적절히 활용되지 않고 있다는 점이 학술적, 그리고 실무적 아쉬움으로 남아 있다. 모든 설문조사에 필연적으로 존재하는 불성실 응답은 연구자들이 많은 관심과 노력을 쏟아 밝히고자 하는 연구문제에 대한 명확한 이해를 방해할 것이다. 본 연구는 불성실 응답에 대한 학술적 논의를 진척시킬 뿐 아니라 이의 탐지와 해결에 대한 실용적인 방법을 안내하고자 하였다. 따라서 불성실 응

답의 개념과 탐지방법에 대한 정보를 가급적 이해하기 쉽고, 향후 활용 가능한 방법으로 제시하였다. 본 연구가 불성실 응답에 대한 보다 많은 국내 연구자들의 관심과 실제 적용을 돕는 디딤돌로 활용되어, 경영학 전반에 발생하는 다양한 현상에 대한 명확한 이해를 돕는 데 도움이 되기를 기대한다.

참고문헌

- 박원우 · 김미숙 · 정상명 · 허규만(2007), “동일방법편의(Common Method Bias)의 원인과 해결방안,” **인사조직연구**, 15(1), pp.89-133.
- 백영민 · 김은미 · 이준웅(2012), “자기응답방법에서 나타나는 인터넷이용시간 과도응답과 그 원인,” **한국언론학보**, 56(2), pp.121-142.
- 이윤석 · 이지영 · 이경택(2008), “온라인조사의 응답오차에 대한 연구: 설문 응답 시간과 응답 성실성의 관계,” **조사연구**, 9(2), pp.51-83.
- Akbulut, Y.(2015), “Predictors of inconsistent responding in web surveys,” *Internet Research*, 25(1), pp.131-147.
- Bachman, J. G., and P. M. O'Malley(1984), “Yeasaying, nay-saying, and going to extremes: Black-white differences in response style,” *Public Opinion Quarterly*, 48(2), pp.491-509.
- Baer, R. A., J. Ballenger, D. T. R. Berry, and M. W. Wetter(1997), “Detection of random responding on the MMPI-A,” *Journal of Personality Assessment*, 68(1), pp.139-151.
- Bagby, R. M., J. R. Gillis, and R. Rogers(1991), “Effectiveness of the Millon Clinical Multi Axial Inventory validity index in the detection of random responding,” *Psychological*

- Assessment*, 3(2), pp.285-287.
- Bagozzi, R., Y. Yi, and L. Phillips(1991), "Assessing construct validity in organizational research," *Administrative Science Quarterly*, 36(3), pp.421-458.
- Bardo, J. W., S. J. Yeager, and M. J. Klingsporn (1982), "Preliminary assessment of format-specific central tendency and leniency error in summated rating scales," *Perceptual and Motor Skills*, 54(1), pp.227-234.
- Barge, S., and H. Gehlbach(2012), "Using the theory of satisficing to evaluate the quality of survey data," *Research in Higher Education*, 53(2), pp.182-200.
- Beach, D. A.(1989), "Identifying the random responder," *Journal of Psychology: Interdisciplinary and Applied*, 123(1), pp.101-103.
- Behrend, T. S., D. J. Sharek, A. W. Meade, and E. N. Wiebe(2011), "The viability of crowd sourcing for survey research," *Behavior Research Methods*, 43(3), pp.800-813.
- Berry, D. T., M. W. Wetter, R. A. Baer, L. Larsen, C. Clark, and K. Monroe(1992), "MMPI-2 random responding indices: Validation using a self-report methodology," *Psychological Assessment*, 4(3), pp.340-345.
- Breitsohl, H., and C. Steidelmüller(2018), "The impact of insufficient effort responding detection methods on substantive responses: Results from an experiment testing parameter invariance," *Applied Psychology*, 67(2), pp.284-308.
- Brief, A. P., and S. J. Motowidlo(1986), "Prosocial organizational behaviors," *Academy of Management Review*, 11(4), pp.710-725.
- Bowling, N. A., and J. L. Huang(2018), "Your attention please! Toward a better understanding of research participant carelessness," *Applied Psychology*, 67(2), pp.227-230.
- Bowling, N. A., J. L. Huang, C. B. Bragg, S. Khazon, M. Liu, and C. E. Blackmore(2016), "Who cares and who is careless? Insufficient effort responding as a reflection of respondent personality," *Journal of Personality and Social Psychology*, 111(2), pp.218-229.
- Calsyn, R. J., and J. P. Winter(1999), "Understanding and controlling response bias in needs assessment studies," *Evaluation Review*, 23(4), pp.399-417.
- Cannell, C. F., P. V. Miller, and L. Oksenberg(1981), "Research on interviewing techniques," *Sociological Methodology*, 12, pp.389-437.
- Chen, G., S. M. Gully, and D. Eden(2001). "Validation of a new general self-efficacy scale," *Organizational Research Methods*, 4(1), pp.62-83.
- Costa, P. T., and R. R. McCrae(2008), "The revised NEO personality inventory (NEO-PI-R)," In G. J. Boyle, G. Matthews, and D. H. Saklofske (Eds.), *The SAGE Handbook of Personality Theory and Assessment*, London, England, SAGE, pp.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), pp.596-612.
- Curran, P. G.(2016), "Methods for the detection of carelessly invalid responses in survey data," *Journal of Experimental Social Psychology*, 66(5), pp.4-19.
- Curran, P. G., L. Kotrba, and D. Denison(2010, April), "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., A. J. DeSimone, P. D. Harms, and D. Wood(2018), "The differential impacts of two forms of insufficient effort responding," *Applied Psychology*, 67(2), pp.309-338.
- DeSimone, J. A., and P. D. Harms(2018), "Dirty data: The effects of screening respondents who provide low-quality data in survey research," *Journal of Business and Psychology*, 33(5), pp.559-577.
- DeSimone, J. A., P. D. Harms, and A. J. DeSimone (2015), "Best practice recommendations for data screening," *Journal of Organizational Behavior*, 36(2), pp.171-181.
- Deutskens, E., K. De Ruyter, M. Wetzels, and P. Oosterveld(2004), "Response rate and response quality of internet-based surveys: An experimental study," *Marketing Letters*, 15 (1), pp.21-36.
- Dunn, A. M., E. D. Heggestad, L. R. Shanock, and N. Theilgard(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), pp.105-121.
- Gallen, R. T., and D. T. Berry(1996), "Detection of random responding in MMPI-2 protocols," *Assessment*, 3(2), pp.171-178.
- Goldberg, L. R., and J. M. Kilkowski(1985), "The prediction of semantic consistency in self-descriptions: Characteristics of persons and of terms that affect the consistency of responses to synonym and antonym pairs," *Journal of Personality and Social Psychology*, 48 (1), pp.82-98.
- Green, S. B., and T. Stutzman(1986), "An evaluation of methods to select respondents to structured job analysis questionnaires," *Personnel Psychology*, 39(3), pp.543-564.
- Graen, G. B., and M. Uhl-Bien(1995), "Relationship-based approach to leadership: Development of leader-member exchange (LMX) theory of leadership over 25 years: Applying a multi-level multi-domain perspective," *The Leadership Quarterly*, 6(2), pp.219-247.
- Guenole, N., J. Ferrar, and S. Feinzig(2017), *The Power of People: How Successful Organizations Use Workforce Analytics to Improve Business Performance*, New York, NY, FT Press.
- Hough, L. M., N. K. Eaton, M. D. Dunnette, J. D. Kamp, and R. A. McCloy(1990), "Criterion-related validities of personality constructs and the effect of response distortion on those validities," *Journal of Applied Psychology*, 75(5), pp.581-595.
- Huang, J. L., N. A. Bowling, M. Liu, and Y. Li(2015), "Detecting insufficient effort responding with an infrequency scale: Evaluating validity and participant reactions," *Journal of Business and Psychology*, 30(2), pp.299-311.
- Huang, J. L., P. G. Curran, J. Keeney, E. M. Poposki, and R. P. DeShon(2012), "Detecting and deterring insufficient effort responding to surveys," *Journal of Business and Psychology*, 27(1), pp.99-114.
- Huang, J. L., M. Liu, and N. A. Bowling(2015), "Insufficient effort responding: Examining an insidious confound in survey data," *Journal of Applied Psychology*, 100(3), pp.828-845.
- Jackson, D. N.(1977), *Jackson Vocational Interest Survey Manual*, London, Canada, Research

- Psychologists.
- Johnson, J. A.(2005), "Ascertaining the validity of individual protocols from Web-based personality inventories," *Journal of Research in Personality*, 39(1), pp.103-129.
- Kam, C. C. S., and J. P. Meyer(2015), "How careless responding and acquiescence response bias can influence construct dimensionality: The case of job satisfaction," *Organizational Research Methods*, 18(3), pp.512-541.
- Kim, D. S., C. J. McCabe, B. L. Yamasaki, K. A. Louie, and K. M. King(2018), "Detecting random responders with infrequency scales using an error-balancing threshold," *Behavior Research Methods*, 50(5), pp.1960-1970.
- Kline, R. B.(2015), *Principles and Practice of Structural Equation Modeling*, New York, NY, Guilford.
- Krosnick, J. A.(1991), "Response strategies for coping with the cognitive demands of attitude measures in surveys," *Applied Cognitive Psychology*, 5(3), pp.213-236.
- LaRose, R., and H. Y. S. Tsai(2014), "Completion rates and non-response error in online surveys: Comparing sweepstakes and pre-paid cash incentives in studies of online behavior," *Computers in Human Behavior*, 34, pp.110-119.
- Little, T. D., W. A. Cunningham, G. Shahar, and K. F. Widaman(2002), "To parcel or not to parcel: Exploring the question, weighing the merits," *Structural Equation Modeling*, 9(2), pp.151-173.
- Liu, M., N. A. Bowling, J. L. Huang, and T. A. Kent(2013), "Insufficient effort responding to surveys as a threat to validity: The perceptions and practices of SIOP members," *The Industrial-Organizational Psychologist*, 51(1), pp.32-38.
- Mahalanobis, P. C.(1936), "On the generalized distance in statistics," *Proceedings of National Institute of Science of India*, 12, pp.49-55.
- Maniaci, M. R., and R. D. Rogge(2014), "Caring about carelessness: Participant inattention and its effects on research," *Journal of Research in Personality*, 48, pp.61-83.
- Maslach, C., and S. E. Jackson(1981), "The measurement of experienced burnout," *Journal of Occupational Behavior*, 2(2), pp.99-113.
- McGrath, R. E., M. Mitchell, B. H. Kim, and L. Hough(2010), "Evidence for response bias as a source of error variance in applied assessment," *Psychological Bulletin*, 136(3), pp.450-470.
- Meade, A. W., and S. B. Craig(2012), "Identifying careless responses in survey data," *Psychological Methods*, 17(3), pp.437-455.
- Niessen, A. S. M., R. R. Meijer, and J. N. Tendeiro (2016), "Detecting careless respondents in web-based questionnaires: Which method to use?," *Journal of Research in Personality*, 63, pp.1-11.
- Oppenheimer, D. M., T. Meyvis, and N. Davidenko (2009), "Instructional manipulation checks: Detecting satisficing to increase statistical power," *Journal of Experimental Social Psychology*, 45(4), pp.867-872.
- R Core Team(2018), *R: A Language and Environment for Statistical Computing (Version 3.5.2)* [Computer Software], R Foundation for Statistical Computing, Vienna, Austria, Available at <http://www.R-project.org/>.
- Roivainen, E., J. Veijola, and J. Miettunen(2016), "Careless responses in survey data and the

- validity of a screening instrument," *Nordic Psychology*, 68(2), pp.114-123.
- Scandura, T. A., and G. B. Graen(1984), "Moderating effects of initial leader - member exchange status on the effects of a leadership intervention," *Journal of Applied Psychology*, 69(3), pp.428-436.
- Schmitt, N., and D. M. Stults(1985), "Factors defined by negatively keyed items: The result of careless respondents?," *Applied Psychological Measurement*, 9(4), pp.367-373.
- Schneider, S., M. May, and A. A. Stone(2018), "Careless responding in internet-based quality of life assessments," *Quality of Life Research*, 27(4), pp.1077-1088.
- Schonla, M., and V. Toepoel(2015), "Straight lining in Web survey panels over time," *Survey Research Methods*, 9(2), pp.125-137.
- Ward, M. K., and A. W. Meade(2018), "Applying social psychology to prevent careless responding during online surveys," *Applied Psychology*, 67(2), pp.231-263.
- Ward, M. K., A. W. Meade, C. M. Allred, G. Pappalardo, and J. W. Stoughton(2017), "Careless response and attrition as sources of bias in online survey assessments of personality traits and performance," *Computers in Human Behavior*, 76, pp.417-430.
- Ward, M. K., and S. B. Pond, III.(2015), "Using virtual presence and survey instructions to minimize careless responding on Internet-based surveys," *Computers in Human Behavior*, 48, pp.554-568.
- Warwick, C., J. Rimmer, A. Blandford, J. Gow, and G. Buchanan(2009), "Cognitive economy and satisficing in information seeking: A longitudinal study of undergraduate information behavior," *Journal of the American Society for Information Science and Technology*, 60(12), pp.2402-2415.
- Williams, L. J., and S. E. Anderson(1991), "Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors," *Journal of Management*, 17(3), pp.601-617.
- Wood, R., and A. Bandura(1989), "Social cognitive theory of organizational management," *Academy of Management Review*, 14(3), pp.361-384.
- Woods, C. M.(2006), "Careless responding to reverse-worded items: Implications for confirmatory factor analysis," *Journal of Psychopathology and Behavioral Assessment*, 28(3), pp.189-194.
- Yentes, R. D., and F. Wilhelm(2018), *careless: Procedures for Computing Indices of Careless Responding*, R packages version 1.1.1, Available at <https://github.com/ryentes/careless>.

Careless Responding in Surveys: Detection Methods and the Impact of Screening on Research

Won-woo Park* · Sung Hyuck Mah** · Su hyun Bae*** · Seon Young Ji**** ·
Yoo Woo Lee***** · Ja Young Kim*****

Abstract

This study investigates the phenomenon of 'careless responding' that prevails in surveys. Specifically, we focus on methods of detection and the effects of screening careless responding. First, this study discusses the diverse definitions and types of careless responding. Methods of detection are introduced in the following triangular categorization: direct, unobtrusive, and statistical. The theoretical review portion provides a comparative summary of detection methods and their advantages and disadvantages to help future researchers apply a suitable one to their own study designs. Furthermore, this study conducts an empirical study to identify the impact of screening careless responding on statistical results, analyzing survey data from 3,030 employees working in a major conglomerate in Korea. The major findings are as follows: (1) According to the applied method, the target and the proportion of screened responses varied, ranging from 0.5% to 14% at maximum. Only a small proportion was detected coincidentally by two or more methods. (2) The screening of careless responding only had a slight impact on statistical figures of factor analysis, correlations, and regression. Such less than moderate impacts were in line with earlier findings and less threatening to the validation of research models. (3) Lastly, we demonstrated a multi-hurdling method that adopted two methods sequentially. To conclude, the study discusses possible applications of detection methods and

* Seoul National University, First Author
** Seoul National University, Corresponding Author
*** Seoul National University, Co-Author
**** Seoul National University, Co-Author
***** Seoul National University, Co-Author
***** Seoul National University, Co-Author

avenues for future research.

Key words: Research Methodology, Surveys, Careless Responding

-
- 저자 박원우는 서울대 경영대학에서 학사 및 석사학위를, 그리고 미 Pittsburgh대학에서 경영학(인사조직) 박사학위(1989년)를 취득하였다. Pittsburgh대학에서 조교수로 근무한 후 귀국하여 중앙대와 경희대 교수를 거쳐 1998년부터 서울대에 재직 중이다. 학계에선 한국경영학회 부회장, 한국인사조직학회 부회장, 한국윤리경영학회 회장 등으로 봉사하였으며, 주요 연구분야는 groupthink, empowerment, trust, efficacy, goal orientation, culture change, 및 happiness인데, 그간 130여 편의 국내외 학술논문과 16 편의 단행본 도서를 출간하였고, 서울대학교 경영대학의 우수강의상을 수차례, 2018년엔 서울대학교 교육상을 수상하였다.
 - 저자 마성혁은 현재 서울대학교 경영대학 인사조직전공 박사과정에 재학 중이다. 연세대학교에서 경영학 학사학위를, 서울대학교에서 경영학 석사학위를 취득하였다. 주요 연구분야는 리더십, 직업과 소명의식, 일과 삶의 균형, 연구방법론 등이다.
 - 저자 배수현은 현재 서울대학교 교육학 박사과정을 수료하였다. 뉴욕주립대학교 알바니에서 심리학과를 졸업하였으며, 서울대학교에서 교육학 석사학위를 취득하였다. 주요 연구분야는 직업교육, HRD 등이 있다.
 - 저자 지선영은 현재 서울대학교 경영대학 석사과정에 재학 중이며 성균관대학교에서 경영학 학사학위를 취득하였다. 주요 관심분야는 의미있는 일(meaningful work), 조직문화, 연구방법론 등이다.
 - 저자 이유우는 현재 서울대학교 농·산업교육과 박사과정을 수료하였으며 중앙대학교 대학원 경영학과 인사조직전공에서 석사학위를 취득하였다. 주요 연구분야는 연령관리, 경력 전환, 연구방법론 등이다.
 - 저자 김자영은 현재 서울대학교 경영대학 전략·국제경영전공 박사과정에 재학 중이다. 연세대학교 실내건축학과를 졸업하였으며, 서울대학교 경영학 석사학위를 취득하였다. 주요 연구분야는 국제합작투자의 설계, 벤처기업의 성장 및 국제화 전략, 팀 및 기업의 창의성 등이다.