한국심리학회지: 일반
Korean Journal of Psychology: General
2026, Vol. 45, No. 4, 643-686
http://dx.doi.org/10.22257/kjp.2026.9.45.4.643

OPEN ACCESS

조사 및 측정에서의 응답 스타일의 발생 원인, 영향, 탐지, 조정, 및 예방: 문헌 개관*

장 재 윤 서강대학교 심리학과

이 태 헌† 고려대학교 심리학부

사회조사와 심리검사에서는 개인의 태도, 성격, 가치, 정서, 행동 경향을 측정하기 위해 자기 보고식 설문 방법이 널리 사용된다. 그러나 자기 보고 자료는 응답자가 문항의 실제 내용에 근거하여 응답했다는 가정에 의존하기 때문에, 문항 내용과 무관하게 특정한 방식으로 응답하는 경향이 개입되면 측정의 타당도가 손상된다. 이러한 문제를 대표하는 현상 중 하나가 응답 스타일(response style)이다. 이는 단순한 무선 오차가 아니라 개인차, 문항 형식, 척도 구성, 조사 맥락, 문화적 규범 등 다양한 요인에 의해 발생하는 체계적 측정오차로 볼 수 있다. 본 연구는 자기 보고식 측정에서 나타나는 주요 응답 스타일을 중심으로 기존 연구를 개관하였다. 구체적으로 극단 응답 스타일, 중간 응답 스타일, 묵인(비묵인) 응답 스타일 등 주요 유형을 정리하고, 이들 응답 스타일의 발생 원인, 탐지 방법, 통계적 보정 절차, 척도 설계 및 조사 운영 단계에서의 예방 전략 등을 논의하였다. 더불어 비교 문화 연구의 맥락에서 응답 스타일의 영향에 관해서도 기존 연구를 중심으로 논의하였다. 마지막으로 본 연구는 연구자와 실무자가 응답 스타일을 식별하고 관리하기 위해 고려해야 할 주요 가이드라인을 제시하였다.

주요어 : 사회조사, 심리검사, 응답 편향, 응답 스타일, 묵인 응답, 극단 응답, 중간 응답, 탐지, 조정, 예방, 비교 문화

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

† 교신저자: 이태헌, 고려대학교 심리학부 교수, (02841) 서울특별시 성북구 안암로 145
Tel: 02-3290-2060, E-mail: taehun0626@korea.ac.kr

Copyright © 2026, The Korean Psychological Association. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (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.

조사 자료의 가치에 회의적인 연구자들은 주로 두 가지 우려를 제기한다. 첫째는 응답자들이 양질의 응답을 제공하는 데 필요한 내성 능력(introspective ability)을 충분히 갖추고 있는가에 대한 의문이다. 사람들은 '내가 왜 그 사람을 싫어하는지', '문제를 어떻게 풀었는지', 또는 '현 직장을 왜 선택했는지'의 질문에 대해 정확하게 답하지 못하기도 하는데, 이는 사람들이 특정 자극에 대한 자신의 반응을 매개한 인지적 과정을 온전히 내성할 수 없기 때문이다(Nisbett & Wilson, 1977). 그럼에도 응답자들은 자신이 그러한 과정을 정확히 보고할 수 있다고 믿는 경향이 있다. 이 경우 이들의 응답은 실제 내성의 결과라기보다, 자신이 이미 가지고 있던 암묵적 인과 이론이나 특정 자극이 반응의 원인으로 얼마나 그럴듯한지에 대한 판단에 근거한 설명일 가능성이 크다. 더 나아가 응답자들은 존재하지 않는 정책이나 실제로 판단하기 어려운 주제에 대해서도 자신의 태도와 의견을 제시하는 모습을 보이기도 한다(Bishop et al., 1980).

두 번째 우려는 응답자가 자신의 태도, 감정, 성향을 어느 정도 내성할 수 있더라도, 실제 응답 과정에서 문항 내용과 무관한 요인들이 개입하여 정확한 응답을 방해할 수 있다는 점이다. 이러한 요인에는 극단 응답이나 묵인 응답과 같은 응답 스타일(response style), 자신을 사회적으로 바람직하게 제시하려는 사회적 바람직성 반응(social desirability responding: SDR), 그리고 문항을 충분히 읽거나 숙고하지 않고 응답하는 불성실 응답(insufficient effort responding: IER) 등이 포함된다. 이들은 서로 다른 기제를 갖지만, 공통적으로 자기보고식 자료의 타당성과 해석 가능성을 저해할 수 있다는 점에서 중요한 문제로 다루어진다.

본 논문에서는 사회조사나 심리측정에서 만성적으로 나타나는 응답 스타일과 관련된 기존 연구를 개관하였다. 자료의 타당도를 훼손하는 응답 스타일에 대해서는 조사 방법 또는 심리측정 연구자들이 오랫동안 해결 방안을 찾고자 하였다. 응답 스타일과 관련된 우려는 현실적이고 진지하게 받아들여야 하지만, 극복할 수 없는 것은 아니다. 본 고에서는 설문조사나 심리측정에서 나타날 수 있는 다양한 응답 스타일 유형을 제시하고, 그것의 발생 원인, 탐지, 조정 및 예방에 관한 기존 연구를 개관하였다. 그리고 비교 문화 연구에서 응답 스타일의 영향 및 대처 방안에 관한 그간의 연구를 논의하였다. 마지막으로 그 결과를 바탕으로 연구자와 실무자들에게 응답 스타일에 대처하는 방법에 대한 유용한 가이드라인을 제시하고자 하였다.

응답 스타일의 유형

표 1에서 볼 수 있듯이, 응답 스타일에는 여러 유형이 있다. 먼저 묵인 응답 스타일(ARS)은 문항 내용에 대해 곰곰이 생각하기보다는 문항 내용과는 무관하게 '예', '그렇다', '동의한다'를 체계적으로 선택하는 경향성이다. 이것은 질문 문항이 난해하거나, 묻는 내용에 대해 잘 알지 못하거나, 해당 주제에 흥미가 없을 때, 조사나 측정 과정에서 피로감을 느낄 때 나타나는데, 마냥 '그렇다'고 응답하는 것이 인지적 노력 없이 설문을 쉽게 끝내는 방안이 된다. 반대로 설문 내용과 무관하게 평정 척도에서 '비동의', '매우 비동의'와 같은 응답 범주를 선호하는 경향성을 나타나는 비묵인 응답 스타일(DARS)도 있다. 또한,

비동의보다 동의를 훨씬 더 많이 보이는 경향성인 순묵인 응답 스타일(net acquiescence response style: NARS)이 있다.

중간 응답 스타일(MRS)은 문항에 대해 척도 중간에 있는 범주를 선택하는 경향성이며, 극단 응답 스타일(ERS)은 평정 척도의 극단값을 선택하는 경향성이다. 평균 응답 주위에 좁게 또는 넓게 응답 범주를 사용하는 경향성인 응답 범위(response range: RR)도 있는데, 범위가 넓다면 변량을 부풀리고, 다변량 관계들의 크기를 줄인다(Greenleaf, 1992a).

여러 종류의 응답 편향을 정확하게 탐지하는 것이 어려울 때가 있다. 예를 들어, 특정 응답 패턴이 부주의한 응답이거나 중간 또는 극단 반응의 결과일 수 있다. 그러면 정확하게 어떤 편향의 작용인지 구분하기 어렵다. 중간 응답이나 극단 응답도 질문 내용과는 무관하게 특정 유형의 응답 범주를 체계적으로 선택하는 편향들이다. 이것들은 부주의 또는 만족화(satisficing)[1]로 인해 발생한다. Krosnick 등(1996)은 이런 편향들이 응답자들이 인지적 부담을 최소화할 수 있는 지름길을 취하려고 할 때 나타난다고 하였다.

1) '제한된 합리성' 개념에 기반하여 조사 연구자들이 제안한 만족화(satisficing) 개념은 사람들이 심리적 에너지를 보존하기 위해 최적이 아닌 의사결정 전략을 동원하는 인지적 휴리스틱을 가리킨다(Krosnick, 1991). 즉, 최적의 해결안을 찾기보다는 '아주 괜찮아 보이는' 첫 번째의 만족스러운 대안을 선택함으로써 시간과 노력을 줄이는 것이다. 조사 연구에서 만족화 행동은 첫 번째 만족스러운 응답 범주를 택하는 것과 더불어 제시된 모든 문항에 동의하거나 이어지는 문항에 대해 계속 동일 선택을 반복하는 응답 스타일, 무반응이나 문항 누락과 같이 불성실하게 응답하는 IER을 모두 포함한다.

발생 원인

응답 스타일이 발생하는 원인을 크게 두 측면으로 구분할 수 있는데, 하나는 자극 측면이고 다른 하나는 응답자 측면이다. 전자는 응답 스타일이 조사나 검사 도구와 연관된 특성으로 생긴다고 본다. 후자는 응답 스타일이 응답자의 개인적 특성의 결과로 나타난다고 본다. 그러나 이 두 가지는 독립적이지 않고 서로 영향을 미칠 수 있다. 예를 들어, 자극(상황) 측면이 개인의 특정 응답 스타일을 보이려는 타고난 성향을 강화하거나 약화할 수 있다.

응답 스타일의 유발 요인: 자극

설문 자체가 응답자의 응답에 영향을 미칠 가능성이 있다. 즉, 설문의 설계와 설문 문항 자체가 응답자에게 자극(stimulus)으로 작용하기에 응답 스타일에 영향을 미칠 수 있다는 것이다. 여기서 자극은 척도 포맷(format), 자료 수집 방식, 인지적 부담, 면접자 효과, 조사 언어, 주제 관여의 측면에서 차이가 있다.

평정 척도의 포맷

리커트형 평정 척도의 포맷이라고 할 수 있는 응답 범주 수(5점 또는 7점 척도)와 명명(labeling)[2] 방식(척도 점 모두에 명명하는 경우

2) 척도에서 라벨(label)은 명확하고 간결하며 모호하지 않아야 한다. 예를 들어 5점 척도에서 '거의 그렇지 않다~별로 그렇지 않다~중간이다~약간 그렇다~상당히 그렇다'와 같은 라벨이나, 직무 만족도를 측정하는 경우 "매우 만족", "만족", "보통", "불만족" 및 "매우 불만족"과 같이 명명할 수 있다. 한편, 척도에서 앵커(anchors)는 척도의 끝점이나 기준점 역할을 하는 특수한 라벨이

또는 척도의 양 끝점에만 명명하는 경우)이 응답 스타일에 영향을 미칠 수 있다(Greenleaf, 1992b; Baumgartner & Steenkamp, 2001). 다양한 실험 조건의 온라인 조사를 통해 Weijters 등(2010)은 4~7점 척도를 사용하였을 때 응답 범주의 수가 늘어나면(긴 평정 척도일수록) ERS가 감소하는 것을 보여주었다(ARS나 NARS에는 영향이 없음). 또한, 중간점을 두거나 완전히 명명된 척도에서 NARS가 증가함을 발견하였다. 그러나 Kieruj와 Moors(2010)는 5~11점의 평정 척도를 비교하였는데, 응답 범주 수는 ERS에 아무런 영향을 미치지 않았다. 따라서 ERS의 경우에는 일관된 결과가 나타나지 않는다.[3] 또한, Weijters 등(2010)이 완전히 명명된 척도와 끝점에만 명명된 척도를 비교하였을 때, 완전히 명명된 척도에서 NARS가 증가하고, ERS가 감소하였다. 또한, 중간점은 NARS를 증가시키지만, ERS는 감소시켰다. Moors 등(2014)의 대규모 조사 실험 결과도 '전혀 동의하지 않는다 - 전적으로 동의한다'와 같이 범주에 명확한 명명이 있는 척도보다, -3~3의 숫자만 표시된 양극 수치 척도에서 ERS가 더 높아짐을 보고하였다. 따라서 가급적 모든 범주에 언어적 라벨(예: 전혀 아니다, 약간 그렇다)을 붙이는 것이 바람직하다. 단순히 숫자만 제시한 척도의 경우, 숫자 구조가 응답자의 인식에 미묘하게 영향을 미치는 데, 숫자가 어떻게 구성되느냐가 응답자의 심리적 해석을 바꾼다(Schwarz et al., 1991). 1~5 척도와 -2~2 척도는 논리적으로 동일한 5점 척도이지만, 응답자의 인지적, 정서적 인식은 다르다. 1~5 척도는 모두 양의 숫자이기에 사람들에게 대체로 긍정적 방향의 정도 차이로 읽힐 가능성이 있지만, -2~2 척도는 중앙을 기준으로 음(-)과 양(+)이 나뉘어 있어서 응답자는 자신의 답이 긍정인지 부정인지에 대해 더 민감해진다. 즉, 같은 문항이라도 숫자 범위가 바뀌면 응답 분포가 달라질 수 있는 것이다.

Weijters 등(2010)에 따르면, 응답 범주의 최적 수는 척도가 사용되는 목적에 달려있다. 그들은 만약 연구자가 평균이나 퍼센트와 같이 응답의 직접적인 요약을 보고하고자 한다면, 완전히 명명된 5점 또는 7점 척도를 사용할 것을 제안한다. 왜냐하면, 명명이 척도 점(응답 범주)의 해석에 있어 일관성을 높이기 때문이다. 이것은 명명(또는 앵커링)이 척도의 의미를 분명하게 해주기에 완전히 명명된 포맷이 신뢰도와 타당도 측면에서 가장 높은 수준을 보인다는 Krosnick(1999)의 제안과도 일치한다. 만약 연구자가 변인 간의 관계를 보거나 선형 모형을 추정하고자 할 경우, Weijters 등(2010)은 양 끝점에만 명명된 5점 또는 7점 척도가 가장 좋은 선택이라고 제안하였다. 응답자는 그러한 척도를 선형 모형에 더 잘 맞는 방식으로 사용하기 때문이다[4].

다. 예를 들어, 척도가 통증을 측정하는 경우 양쪽 극단의 기준점은 "통증 없음" 및 "상상할 수 있는 최악의 통증"이 될 수 있다.

3) Weijters 등은 응답 스타일의 대표적 지표(RIRS: representative indicators for response styles)방법을 사용하였고, Kieruj와 Moors는 ERS를 잠재-계층 확인적 요인분석(LCFA)으로 모델링하였다.

4) 다만 이러한 권고는 5점 또는 7점 척도가 모든 연구 상황에서 보편적으로 최적이라는 뜻으로 이해되어서는 안 될 것이다. 실제 자료에서 응답 범주가 적절하게 기능하는지 평가하는 데는 IRT 기반 문항 분석이 유용할 수 있다. 즉, IRT 기반 문항 분석을 통해 특정 범주가 거의 사용되지 않거나 인접 범주와 구분되지 않는지를 확인할 수 있다. 따라서 척도 포맷을 결정할 때는 5점 또는 7점 척도라는 일반적 권고와 함께, 예비 조사와

또한, ARS는 척도에서 긍정적 및 부정적으로 표현된 문항들이 비슷하게 균형을 이루면 통제될 수 있다는 제안도 있지만(Baumgartner & Steenkamp, 2001), 이것은 응답의 정확성(즉, 타당도)을 낮추는 결과를 초래할 수도 있다(Schriesheim & Hill, 1981).

한편, 응답 포맷에 따라 응답 스타일 효과가 달라질 수 있다. Böckenholt(2017)는 구조에 대한 선호와 구조 결여에 대한 반응이라는 두 하위 차원을 포함하는 성격 특성인 '구조에 대한 개인적 욕구(personal need for structure)'를 측정하기 위해 세 가지 응답 포맷, 즉 전통적 Likert 포맷, 먼저 동의/비동의를 선택한 뒤 그 강도를 평정하는 funnel 포맷, 그리고 문항을 응답 범주에 끌어다 놓는 drag-and-drop 포맷을 비교하였다. 분석 결과, 세 응답 포맷은 유사한 특성 점수를 산출했지만, 응답 스타일 효과는 포맷에 따라 다르게 나타났다. 특히 세 조건 중 두 조건에서 포맷 특유의(response-format-specific) 응답 스타일 효과가 상당한 수준으로 확인되었다. 이 연구는 응답 스타일을 단순한 잡음(nuisance)으로 보기보다, 응답자가 서로 다른 설문 형식에 어떻게 반응하는지를 보여주는 심리적 현상으로 다룰 필요가 있음을 시사한다.

자료 수집 방식

자료 수집 방식에 따라 응답 스타일에서 차이가 나타난다. 면대면 방식 대 온라인 비대면 방식, 지필 설문 조사 대 온라인 웹 조사에 따라 ARS, ERS, MRS가 달라질 수 있다. 예를 들어, 면대면 조사에서 ARS와 ERS는 높아지고, MRS는 낮아진다(Weijters et al., 2008). 또한, 지필 조사보다 웹 조사에서 ERS가 더 높게 나타난다(Kiesler & Sproull, 1986). 온라인 조건의 참가자들이 지필 조건의 참가자들보다 ERS와 DARS를 약간 덜 나타내며, 지필이나 온라인 자료 수집과 비교하여 전화 인터뷰에서 ARS가 더 자주 나타났다(Weijters et al., 2008). 하나의 자료 수집 방식만 사용되었을 때 특정한 응답 스타일이 두드러지거나 감소할 수 있다. 따라서 두 가지 이상의 자료 수집 방식을 사용하여 자료를 통합하는 경우 응답 스타일로 인해 차이가 나타날 수 있으므로, 이 부분에 주의를 기울여야 한다.

인지적 부담

조사나 측정에서 응답자에게 인지적 부담(cognitive load)이 큰 상황이 자주 나타난다. 인지적 부담이 클수록 ARS(Knowles & Condon, 1999) 및 NARS(Cabooter, 2010)가 늘어난다. 문항에 모호하거나 복잡한 표현들이 있으면 의미를 파악하기 위해 더 많은 인지적 노력이 요구되므로, 연구자는 설문 문항이 최대한 명료하게 표현되도록 해야 한다. 또한, 응답자가 어느 정도 여유 시간이 있을 때 응답하도록 요구하는 것이 좋다. 그리고 응답자가 높은 인지적 부담 상황에서 조사에 응하고 있다고 의심되면, 응답 스타일에 대한 사후 평가를 해야 한다.

주제 관여

응답자가 설문 문항과 관련된 주제(topic)에 대해 어느 정도 개인적 관여(involvement)를 느끼는가는 설문 응답 과정에서의 주요한 심리적 요인이다. 주제 관여는 개인이 특정 이슈에 대해 가지는 중요성 인식, 관련 지식 수준,

IRT 기반 범주 기능 분석을 통해 해당 연구 맥락에서 범주 수가 적절한지 확인하는 것이 바람직하다.

정서적 투자 등을 포함하는 개념이며(Krosnick, 1991), 이러한 관여 정도는 응답자가 척도 범주를 사용하는 방식과 반응 강도에도 직접적 영향을 미친다. 가장 직접적으로 주제 관여와 ERS 간의 관계를 검토한 Gibbons 등(1999)은 개인이 설문 주제에 대해 높은 관심과 정서적 관여를 보일수록 척도의 극단 범주를 선택하는 경향이 증가함을 보고하였다. Krosnick(1991)이 제안한 '만족화(satisficing)' 모델 역시 낮은 관여 또는 낮은 관심도는 피상적이고 형식적 반응을 유발하지만, 높은 관여나 관심은 정교한 정보처리를 통해 ERS를 증가시킬 수 있다고 설명한다. ERS를 직접적으로 측정한 연구는 아니지만, 사회심리학 연구에서 태도 중요성이나 관여 정도가 응답의 극단성과 직접적으로 연결된다는 실증적 증거는 많다(예, Bishop, 1987). 소비자 심리 연구로, Baumgartner와 Steenkamp(2001)는 특정 제품이나 이슈에 대한 관여도가 높을수록 응답 스타일(ERS 및 ARS)에 차이가 나타남을 보여주었다. 따라서 주제 관여가 높을수록 응답자의 극단 범주 사용이 증가한다는 점에서 일관된 결과가 나타난다. 다만, 이 경우 ERS가 단순한 응답 편향인지 아니면 설문 문항을 처리하는 응답자의 지식 수준이나 동기를 반영하는 심리적 현상인지 구분하기 어려울 수 있다.

일반적으로 저부담 상황의 조사나 검사에서 가장 큰 도전은 응답자들의 동기 수준을 높여 평가 과정에 몰입하도록 하는 것이다. 설문지가 개인적으로 관련 있는 것으로 느끼고(주제에 높게 관여될 때) 그리고 연구 결과가 중요하고 사회에 유용하다고 생각할 때 정확하게 응답하려는 동기가 높아져 응답 스타일이 감소한다(Gibbons et al., 1999; Krosnick, 1991).

조사 언어(survey language)

간혹 비교 문화 연구를 할 때, 여러 나라 응답자를 대상으로 영어로 설문하는 경우가 있다(Rowland et al., 2010). 이때 일부 응답자는 모국어가 아닌 제2 외국어로 설문 응답을 하게 된다. 제2 외국어(예, 영어)로 된 설문에 응답하면 모국어로 하는 경우보다 ARS와 ERS는 줄어들지만, MRS와 RR은 늘어난다(Gibbons et al., 1999; Harzing, 2006). 평정 척도에 자신의 응답을 제대로 반영할 수 있기 위해 가급적 모국어로 설문에 응답하는 것이 필요하다.

응답 스타일의 유발 요인: 응답자 특성

응답 스타일은 응답자의 특성이나 성격으로 나타나기도 한다. 다음부터 인구통계적 특성(연령, 성별, 교육, 지능), 성격, 문화의 영향에 관한 연구를 개관하였다.

인구통계적 변인

먼저, 연령과 응답 스타일 간의 관계에서 ARS와 연령 간에는 정적 관계가 나타나기도 하지만(Billiet & McClendon, 2000; Weijters, Geuens, et al., 2010a), Eid와 Rauber(2000)는 아무런 관계가 없다는 결과를 제시하였다. Weijters, Geuens 등(2010a)은 DARS와 연령 간의 아무런 관계를 발견하지 못하였고, MRS의 경우에는 정적인 관계가 나타났다. ERS의 경우, 일부 연구에서는 연령과 ERS 간의 정적 관계가 나타났지만(Greenleaf, 1992a; Weijters, Geuens, et al., 2010a), 다른 연구에서는 부적인 관계를 보였고(Austin et al., 2006; Light et al., 1965), 아무런 관계가 없다는 연구 결과도 있다(Moors, 2008). 그러나 De Jong 등(2008)의 연구에서는 젊은층과 고령층이 중년층보다 더

높은 ERS를 보였다. 이러한 곡선 관계는 앞의 상반된 결과를 설명한다. 즉, 젊은층이나 중년층보다 고령층의 비율이 높으면 그리고 고령층이 더 높은 ERS를 보이면, 연령과 ERS 간의 정적인 선형 관계가 있을 것이다. 반대로 젊은층의 비율이 높아서 그들이 높은 ERS를 보인다면, 연령과 부적인 선형 관계가 나타날 것이다. 만약 젊은층과 고령층의 비율이 거의 동일하고 두 집단 모두 높은 ERS를 보인다면, 선형 모형으로는 아무런 관계를 찾지 못할 것이다.

둘째, 성별의 경우, 남성보다 여성에게서 더 높은 ARS가 나타나기도 하고(Austin et al., 2006; Weijters, Geuens, et al., 2010a), 성차가 드러나지 않는 연구도 있다(Light et al., 1965). ERS의 경우, 여성에게서 더 높은 결과를 보인 연구(De Jong et al., 2008; Weijters Geuens, et al., 2010a)와 남성에게서 더 높은 경향성을 보인 연구(Harzing, 2006), 그리고 아무런 차이를 보이지 못한 연구(Light et al., 1965; Moors, 2008)가 있다. DARS의 경우, Crandall(1973)의 연구에서 성차가 나타나지 않았다. MRS의 경우, Harzing(2006)는 여성에게서 높은 수준이 나타남을 보였고, Light 등(1965)은 성차를 발견하지 못하였다.

셋째, 교육 수준은 거의 예외 없이 응답 스타일과 부적인 관계, 즉 교육 수준이 높을수록 응답 스타일이 감소하는 경향을 보인다. Meisenberg와 Williams(2008)는 ARS와 ERS의 경우 이러한 관계가 특정 국가나 문화권에 국한되지 않고 비교적 보편적으로 나타나는 현상임을 보여주었다. 그러나 교육 수준의 효과가 모든 응답 스타일에서 동일하게 확인된 것은 아니다. 예컨대 Weijters, Geuens 등(2010a)은 ARS, DARS, ERS를 함께 검토하였는데, ARS와 ERS에서는 교육 수준과의 부적 관계가 있었지만, DARS에서는 그러한 관계가 나타나지 않았다. 또한 Moors(2008)와 De Jong 등(2008)의 연구에서는 교육 수준이 ERS에 유의한 영향을 미치지 않는 것으로 나타났다. 따라서 교육 수준은 응답 스타일을 낮추는 중요한 개인차 요인으로 볼 수 있지만, 그 효과의 크기와 일관성은 응답 스타일의 유형, 표본, 문화적 맥락, 그리고 분석 방법에 따라 달라질 수 있다.

한편, Light 등(1965)은 초등 4학년~고교 졸업반에 이르는 연령 범위(9세~18세)의 학생들을 대상으로 ERS와 IQ 간의 부적인 관계를 보여주었다. 지능이 높고 나이가 많을수록 ERS가 낮고 MRS는 더 높았다. 더 나아가 그들은 고지능의 저학년보다 고지능의 고학년에서 더 낮은 MRS가 나타나지만, 저지능의 경우에는 반대 현상이 나타남을 보여주었다. 미국 대학 입학시험으로 측정된 지능은 ARS와 부적인 관계가 있었다(Zuckerman & Norton, 1961). Trippas 등(2015)은 인지적 능력이 높은 참가자가 더 높은 추론 능력을 보이며, 응답 편향을 덜 보이는 경향이 있음을 보여주었다. 조사에서 단순히 응답자의 과거 사실이나 정보를 묻는 일상적 질문이 아니고 상당한 추론이 요구되는 조사 질문을 사용하는 경우라면 높은 인지 능력의 응답자들이 응답 편향을 덜 보일 수 있다.

전반적으로 기존 문헌에 의하면 인구통계적 변인들이 응답 스타일에 어느 정도 영향을 미치지만, 일관된 결과를 보이지 않는다. 인구통계적 변인은 응답 스타일의 직접적 결정요인이라기보다는 응답자가 어떤 인지적, 사회적 조건에 놓여 있고 어떤 측정 양식이나 맥락에 있는지를 반영하는 대리 변인(proxy)으로 가깝다. 따라서 최근에는 단순히 통제 변인으로

처리하는 접근에서 벗어나 잠재 특성과 응답 스타일을 측정 모형 차원에서 분리하고 인구통계적 변인을 응답 전략의 발현 가능성을 조절하는 변인으로 보고자 한다.

성격

응답 스타일이 비교적 안정적이고 일관된 개인차 특성을 지닌다는 점을 보여주는 연구도 적지 않다. 예컨대 Hamilton(1968)은 설문지 전반에 걸쳐 ERS가 안정적으로 나타난다는 점을 보고하였고, Naemi 등(2009)과 Weijters, Geuens 등(2010b)은 ARS와 ERS가 여러 문항군에 걸쳐 일관되게 나타남을 보여주었다. 또한 Weijters, Geuens 등(2010a)은 1년의 시차를 둔 자료 수집에서 ARS, DARS, MRS, ERS가 일정한 안정성을 보인다는 점을 확인하였으며, Billiet과 Davidov(2008)는 동일 응답자의 ARS가 4년에 걸쳐 안정적으로 유지됨을 보여주었다. 이러한 결과들은 응답 스타일이 특정 설문 상황에서 우연히 발생하는 반응 패턴이라기보다, 개인에게 비교적 지속적으로 나타나는 응답 경향성임을 시사한다.

응답 스타일이 성격 및 인지적 특성과 관련된다는 연구 결과도 이러한 해석을 뒷받침한다. 예컨대 ERS는 모호성에 대한 인내, 단순 사고와 결단력에 대한 선호, 그리고 성격 5요인 중 외향성 및 성실성과 관련이 있는 것으로 보고되었다(Austin et al., 2006; Naemi et al., 2009).[5] Ayidiya와 McClendon(1990)은 MRS가 회피성(evasiveness)과 정적인 관계를 보인다고 보고하였으며, Couch와 Keniston(1960)은 ARS가 충동성 및 외향성과 정적인 관계를 갖는다는 결과를 제시하였다. 이처럼 응답 스타일은 설문 문항의 형식이나 조사 맥락에 의해 유발될 수 있지만, 동시에 응답자의 비교적 안정적인 심리적 특성과도 관련될 수 있다.

그러나 성격의 영향에 관한 이전 연구 결과는 성격 평가에서 이미 평정 척도가 사용되었기에 성격 측정 자체가 응답 스타일에 의해 오염되었을 수 있다는 지적이 있다(Bentler et al., 1971). 이에 대응하여 Cabooter(2010)는 척도 형식이 아닌 독특한 성격 측정치를 사용하여[6] 자기-조절 초점(self-regulatory focus)과 ERS와

5) Naemi 등(2009)은 ERS가 모호성에 대한 불내성(intolerance of ambiguity), 단순 사고 성향(simplistic thinking), 결단력(decisiveness)에 의해 설명되는지를 조사하였다. 116쌍의 학부생과 해당 학생의 동료가 참여하여, 이러한 성격 요인을 측정하는 여러 설문지와 함께 세 가지 ERS 측정 도구에 응답하였다. 연구 결과, 동료가 평가한 모호성 불내성과 단순 사고 성향이 '해당 학생이 설문에 소요한 시간'과 상호작용하여 학생의 ERS를 예측하는 것으로 나타났다. 즉, 설문을 빠르게 완료하면서 동시에 모호성을 견디지 못하거나 사고가 단순한 사람일수록 ERS를 보일 가능성이 가장 높았다.

6) Cabooter의 박사학위 논문(2010) 실험에서는 다음과 같은 두 가지 유형의 조절 초점 측정 방법을 사용하였다. 첫째, 조절 초점 시나리오 기반 과제로, 참가자들은 증진 또는 예방 상황을 묘사하는 짧은 시나리오를 읽고 본인이 각각의 상황에서 어떻게 행동하거나 느낄지를 판단하는 의사결정 방식으로 응답하였다. 이는 일반적인 Likert 척도 문항이 아니라, 상황 해석 패턴 및 목표 접근/회피 경향을 기반으로 성격을 측정하는 방법이다. 둘째, 행동적 조절 초점 선택 과제로 조절 초점이 반영되는 개별적인 선택 과제를 사용하였다. 예를 들어, 증진 초점이라면 성취나 이득을 강조하는 선택지를 선호하고, 예방 초점이라면 안전이나 손실 회피의 선택지를 선호한다. 이러한 선택 패턴을 통계적으로 분석하여 개인의 증진 대 예방 지표를 추정하였다. 이 경우도 행동적 선택을 통해 성격을 추정하는 비척도형 성격 측정 방식이다.

MRS를 연구하였는데, 예방 초점은 MRS와 정적으로 관련되고, 향상 초점은 ERS와 정적으로 관련되는 것으로 나타났다. 이런 결과는 성격과 응답 스타일 간에 특정한 관계가 있음을 좀 더 분명하게 보여준다.

성격 특성이 응답 스타일을 예측한다는 사실은 응답 스타일이 설문 형식이나 조사 상황만으로 설명되는 현상이 아니며, 따라서 연구자가 이를 완전히 예방하기 어려울 수 있음을 시사한다(Kieruj & Moors, 2010). 최근에는 응답 스타일과 성격 간의 관계를 광범위한 성격 요인보다는 자기 통제나 인지적 반성(reflection)과 같은 자기 조절 관련 개인차의 관점에서 설명하려는 연구들이 나타나고 있다(Fukudome & Takeda, 2024; Roberts et al., 2019). 이러한 접근은 자기 조절 관련 특성이 응답 내용 자체보다는 문항을 숙고하는 정도와 응답 전략의 선택에 영향을 미치며, 응답 스타일이 인지적·동기적 자원 관리의 결과일 수 있음을 시사한다. 즉, 성격 및 자기조절 관련 개인차는 특정 문항 내용에 대한 응답을 직접 결정하기보다, 응답 과정에서 특정 전략이 선택될 가능성을 조절하는 요인으로 작용할 수 있다.

응답 스타일의 영향

표 1의 '영향' 부분에서 볼 수 있듯이, 자기보고식 조사에서 나타나는 응답 스타일은 크게 두 가지 방식으로 연구 결론의 타당도에 영향을 미친다(Baumgartner & Steenkamp, 2001). 첫째, 응답 스타일은 단변량 분포, 즉 평균과 변량에 영향을 미칠 수 있다. 예를 들어 Moors(2012)는 자유방임형 리더십에서 성차를 보고한 연구 결과가 실제 리더십 지각의 차이라기보다, 남성과 여성 간 ERS 수준의 차이에 의해 나타난 것일 수 있다고 지적하였다. 즉, 특정 집단이 평정 척도의 극단 범주를 더 자주 선택하는 경향을 보일 경우, 집단 간 평균 차이가 실제 구성개념의 차이가 아니라 응답 스타일에 의해 만들어진 체계적 오차를 반영할 수 있다. 따라서 응답 스타일을 통제하지 않으면 연구자는 t 검증이나 F 검증과 같은 집단 비교 분석에서 잘못된 결론에 이를 수 있다.

둘째, 응답 스타일은 다변량 분포에도 영향을 미친다. 예를 들어, Baumgartner와 Steenkamp (2001)는 건강 의식(HCO), 품질 의식(QCO), 환경 의식(ECO), 및 자민족중심주의(ETN) 간의 상관을 조사하여 HCO - QCO 간 .40, HCO - ECO 간 .33, QCO - ECO 간 .31, HCO - ETN 간 .28, QCO - ETN 간 .19, ECO - ETN 간 .15의 결과를 얻었다. 이론적 관점에서는 HCO, QCO, ECO 간 유의한 상관이 있을 것이지만, ETN과는 상관이 없어야 한다. 이때 응답 스타일을 통제하면 상관은 상당히 감소한다(HCO - QCO, .20; HCO - ECO, .15; QCO - ECO, .13; HCO - ETN, .02; QCO - ETN, .00; ECO - ETN, .01). 따라서 응답 스타일은 변인 간의 상관 크기에 영향을 미친다. Cronbach 알파, 회귀분석, 요인분석, 구조방정식 모형과 같은 통계분석 기법들은 변인 간 상관에 기초하고 있으므로, 응답 스타일에 대한 통제 없이 변인 간 관계를 조사하는 연구는 잘못된 결과와 결론에 이를 수 있다.

일상적 맥락에서 개인의 정서, 행동, 인지 상태를 고해상도로 포착할 수 있는 경험표집법(Experience Sampling Method: ESM)와 생태학적 순간 평가(Ecological Momentary Assessment: EMA)는 순간적인 자기 보고(momentary self-

표 1. 응답 스타일의 유형 및 특징

응답 스타일	특성	영향
묵인(acquiescence) 응답 스타일 (ARS)	설문 내용과는 무관하게 문항에 동의하는 것으로, 평정 척도에서 '동의', '매우 동의'와 같은 응답 범주를 선호하는 경향	관찰된 평균을 부풀리며, 다변량 관계의 크기를 증가시킴
비묵인(disacquiescence) 응답 스타일(DARS)	설문 내용과 무관하게 문항에 부동의하는 것으로, 평정 척도에서 '비동의', '매우 비동의'와 같은 응답 범주를 선호하는 경향	관찰된 평균을 낮추며, 다변량 관계의 크기를 증가시킴
극단(extreme) 응답 스타일(ERS)	평정 척도의 가장 높고 낮은 범주에만 주로 체크하는 것으로, '매우 동의' 또는 '매우 비동의'와 같은 극단적 응답 범주를 선호하는 경향	관찰된 평균을 왜곡하고, 변량을 축소하며, 다변량 관계의 크기를 감소시킴
중간(midpoint) 응답 스타일(MRS)	설문 내용과는 무관하게 평정 척도의 중간에 있는 범주를 선호하는 경향	관찰된 평균이 중간점에 가깝게 되며, 변량을 축소하고, 다변량 관계의 크기를 증가시킴
온건 응답 스타일 (mild response style: MLRS)	평정 척도의 가장 높고 가장 낮은 응답 범주를 피해서 체크하는 경향으로, ERS와 상반됨	관찰된 평균이 중간점에 가깝게 되며, 변량을 축소하며, 다변량 관계의 크기를 증가시킴

주) Van Vaerenbergh & Thomas(2013)와 Wetzel, Böhnke, & Brown(2016)의 내용을 보완함.

report)를 반복적으로 수집하는 특성상, 응답 스타일 및 SDR과 같은 응답 편향이 데이터의 구조적 특성을 왜곡할 가능성이 지속적으로 제기되어 왔다. 즉, ESM 자료에서 개인 차원의 척도 사용 습관이 반복적으로 지적되는 데(Nezlek, 2017; Shrout & Lane, 2012), MRS, ARS, ERS에서의 개인차는 일상 정서의 개인 내 범위(range), 변량, 왜도, 극단값 빈도를 왜곡할 수 있다. 특히, 집중 종단 자료(intensive longitudinal data, ILD) 또는 반복 측정된 정서 자료를 활용하는 연구에서, 중요한 관심사 중 하나는 정서가 시간에 따라 어떻게 변화하는지의 정서 역동성(affect dynamics)이다.[7] ILD 자

7) 정서 역동성을 측정하기 위한 대표적인 두 지표는 다음과 같다. 첫째, 개인 내 변동성이다. 이는 개인의 정서 상태가 자신의 평균 상태로부터 얼마나 벗어나는지를 보여준다. 일반적으로 각 측정 시점에서의 정서 점수에 대한 표준편차로 계산된다. 변동성은 변동의 크기를 포착하지만, 시간 간 인접 응답 간의 의존성은 고려하지 않는다(Eid & Diener, 1999; Kuppens et al., 2007). 둘째, 1차 자기회귀 효과(first-order autoregressive effect 또는 inertia)이다. 이 지표는 어떤 시점의 정서 상태가 그다음 시점의 상태를 예측하는 정도를 나타낸다. 이는 정서의 지속성 또는 유연성 결핍을 의미하며, 높은 관성(inertia)은 정서 상태가 쉽게 바뀌지 않음을 나타낸다. 이러한 측정치는 단순히 정동 역동성만을 반영하는 것이 아니라, 참여자의 응답 스타일에도 체계적으로 영향을 받을 수 있다(Adams et al., 2019).

료에서는 ERS와 같은 응답 스타일이 변산성 측정에 영향을 줄 수 있다. ERS 수준이 높은 사람은 극단 범주를 과도하게 선택하는 경향이 있어, 실제로 정서 상태가 크게 변하지 않더라도 높은 변산성이 나타날 수 있다. 반면, 관성(1차 자기회귀 효과)의 경우는 ERS가 높을수록 감소할 수 있다. 이에 대해 직접적이고 정교한 증거를 제공한 Henninger 등(2025)은 1,398명의 대규모 표본을 대상으로, 확률적 보상 과제를 통해 정서 상태를 140회 반복 측정하고, 실험 전에 독립적으로 측정된 ERS와 여러 정서 변화 지표 간의 관계를 분석하였다. 연구 결과, ERS는 정서의 개인 내 변동성(intraindividual variability)과 강하게 정적으로 연관되어, ERS가 높은 개인은 실제 정서 변화가 동일하더라도 평정 값의 진폭(amplitude)이 더 크게 나타났다. 이는 변동성 지표(예, SD, RMSSD 등[8])가 응답 스타일에 의해 크게 왜곡될 수 있음을 보여주는 것이다. 반면, ERS는 정서가 시간에 따라 이어지는 방식 자체를 바꾸지는 않는다. 즉, 직전에 기분이 좋았으면 다음에도 좋은 편인지(반대로 금방 달라지는지)와 같은 관성(inertia)과 어떤 자극(보상, 스트레스)이 들어왔을 때 얼마나 민감하게 반응하는지의 반응성(reactivity)에는 크게 영향을 미치지 않았다. ERS는 단지 점수를 크게 올리거나 내리게 만드는 '표현 방식의 과장'일 뿐, 정서가 실제로 천천히 또는 빨리 변하는 사람인지, 또는 자극에 민감한지 둔감한지 같은 실제 심리적 특성까지 바꾸지는 않는다는 것이다. 이러한 발견은 ILD 또는 ESM(EMA) 기반의 정서 불안정성 연구(예, 경계성 장애, 기분 불안정)에서 높은 변동성이 정서 시스템의 불안정성을 의미하는 것이 아니라, ERS로 인해 과대 추정된 가공물(artifact)일 가능성을 시사한다.

응답 스타일의 탐지

응답 스타일을 탐지하는 여러 가지 방법이 있지만, 탐지 방법은 크게 외현(manifest) 변인 접근과 잠재(latent) 변인 모델링으로 구분된다(Wetzel et al., 2016)[9]. 전자는 빈도 지표를 사용하는 것과 같은 명시적 접근이며, 후자는 응답 스타일 측정을 위한 확인적 요인분석(CFA)이나 문항 반응 모델(item response model)과 같은 잠재 변인 모델링 접근이다.

또한 별도 문항의 사용 여부라는 측면에서도 구분된다. 이것은 관심 특성과 응답 스타

8) SD는 표준편차, RMSSD(Root Mean Square of Successive Differences)는 반복측정 자료(특히 시간 연속적 데이터)에서 단기 변동성을 재는 지표로, 연속된 시점 간 점수 차이의 제곱 평균의 제곱근이다. SD가 전체 관측 기간의 전체적 변동성을 나타낸다면, RMSSD는 시계열의 단기적인 급격한 변화, 즉 시점 간 즉각적 변동성을 반영한다. 정서의 경우, 표준편차는 정서가 평균을 중심으로 얼마나 넓게 분산되는지를 나타내지만, RMSSD는 정서가 순간적으로 얼마나 출렁이는지를 나타낸다. 그래서 RMSSD는 EMA와 ESM 연구에서 정서적 불안정성(affective instability)의 핵심 지표 중 하나이다.

9) 다만 이러한 구분은 설명의 편의를 위한 것이다. 빈도 지표와 같은 외현 변인 접근은 주로 응답 스타일의 정도를 산출하거나 선별하는 데 초점을 두지만, 잠재 변인 모형은 스타일의 탐지와 특성 점수의 조정을 동시에 수행할 수 있다. 따라서 본 절에서는 독자의 이해를 돕기 위해 탐지 방법을 먼저 소개하되, 일부 잠재 변인 접근이 이후 조정 절에서 다시 논의될 수 있음을 전제로 한다.

표 2. 응답 스타일의 탐지

방법	특징	장점	단점	관련 문헌
빈도 지표	설문의 전체 문항에 대한 동의, 비동의, 극단 응답, 중간점 응답의 수를 셈	사용하기 쉽고 추가 지표가 필요하지 않음	이질적인 문항들에서만 작동함	Greenleaf (1992a, 1992b); Baumgartner & Steenkamp (2001); Weijters et al. (2010a, 2010b)
역점수 문항에 대한 이중 동의의 수	설문에 역점수 문항들이 포함되어 있고, 역점수 문항들에서의 이중 동의의 수를 셈	사용하기 쉽고 추가 지표가 필요하지 않음	가끔 역점수 문항을 만들기가 어려움. 역점수 문항에 대한 응답은 인지적 처리 전략의 차이에 기인될 수 있음	Krosnick (1991); Weijters et al. (2009); Weijters & Baumgartner (2012)
MTMM/ CFA에서 방법 요인 명세화	동일 특성이 서로 다른 방법들로 반복 측정됨. 관찰된 변량은 참 변량과 오차 변량으로 분해될 수 있음. 내용 요인들과 별도로 방법 요인을 명세하여 응답성향을 추정함.	추가 문항 없이 기존 척도 내에서 방법효과를 모형화할 수 있음. ARS와 DARS의 순수(net) 효과를 측정함.	대체로 ARS나 문항 워딩 효과에 초점이 맞춰짐. MRS, 또는 ERS를 충분히 설명하지 못할 수 있음. 긍정 및 부정 문항이 모두 포함된 척도 문항을 사용해야 함. 모형을 식별하기 위한 방법 요인 부하량 제약, 균형 문항 구성, 식별 조건 등에 민감함 반복 측정으로 인해 일관성 편향과 기억 효과가 발생할 수 있음. 종종 식별(identification)의 문제들이 발생함	Campbell & Fiske (1959); Podsakoff et al. (2003); Billet & McClendon (2000); Eid et al. (2003)
RIRS 방법	관심 변인과 상관되지 않고, 최대한 이질적인 문항들을 많이 추가함. 그리고 가중된 RS 지표들을 계산함	계산하기 쉬움. 내용과 관련이 없는 ARS, DARS, ERS, MRS, NARS를 측정할 수 있음. 이후 분석에서 공변인으로 사용하기 쉬움	조사에 관심 변인과는 무관한 문항들이 많이 추가되어야 함	Baumgartner & Steenkamp (2001); Weijters, Schillewaert, & Geuens (2008)
RIRMACS 방법	RIRS 방법을 확장한 것으로, 관심 변인과 상관되지 않은 문항들을 추가함, 이것은 CFA에서 관찰된 변인이 됨. ARS, DARS, MRS 및 ERS는 잠재 변인이 됨.	사용하기 쉬움. RS 지표들이 이후 분석에서 공변인으로 추가될 수 있음. 구체적인 RS 지표들이 사용되면 내용과 스타일 간의 변별이 가능하고, ARS, DARS, MRS, 및 ERS의 측정이 가능함. 서로 다른 응답 스타일의 수렴 및 변별 타당도 검증이 가능함	조사에 관심 변인과는 무관한 문항들이 많이 추가되어야 함	Weijters, Schillewaert, & Geuens (2008)
LCFA	잠재계층분석과 요인분석을 결합하여 내용 요인과 응답 스타일 요인을 동시에 고려함.	추가 지표들이 필요하지 않음. 일부 모형들은 ARS와 ERS를 변별할 수 있음	연구자가 잠재 계층 분석에 친숙하지 않을 수 있음. 계층 수와 요인 수 선택과 해석이 어려움. 발견된 계층이 실제 응답 스타일인지 보장하기 어려움. DARS와 MRS를 충분히 설명하지 못하는 경우가 많음.	Moors (2003); Morren et al. (2011); Kieruj & Moors (2013)
IRT 모형	응답 스타일을 문항 반응 과정의 잠재요인/잠재계층/ 의사결정노드로 모형화	관심 특성과 응답 스타일을 동시에 추정함	모형이 복잡하고 표본 크기와 문항 수가 충분해야 함. 모형 선택과 식별 가정에 민감함	De Jong et al. (2008); Bolt & Johnson (2009); Bolt & Newton (2011); Bolt et al. (2014); Böckenholt (2012); De Boeck & Partchev (2012); Bolt & Meng (2025)

일을 평가하는 데 동일한 문항이 사용되는지, 아니면 관심 특성을 측정하는 문항과 별개로 응답 스타일을 평가하기 위한 목적의 문항이 사용되는지로 구분된다. 다음부터 주요한 탐지 방법들을 소개하였다(표 2 참조).

첫째, 응답 스타일 연구의 시초부터 사용된 것으로 빈도 지표가 있다(Greenleaf, 1992a, 1992b). 설문에 포함된 전체 문항에 대해 동의, 비동의, 극단 응답, 중간 응답의 수를 세는 것이다(Baumgartner & Steenkamp, 2001; Weijters, Geuens, & Schillewaert, 2010b). 예를 들어, ERS의 경우, 모든 설문 문항에서 응답자가 선택한 극단 범주의 수를 세는 것이다. ARS의 경우, 동의를 나타내는 응답 범주를 세면 된다. 가장 단순한 이 방법은 설문 문항들이 의미나 내용 측면에서 상당히 이질적일 때 사용할 만한 것이다. 여러 검사 도구에 포함된 이런 빈도 지표들은 검사 결과를 채점하는 과정에서 타당도 체크의 기능을 한다. 예를 들어, NEO-PI 검사에서는 ARS와 DARS를 확인하기 위해 동의 및 비동의한 문항 수를 세도록 권고한다.

둘째, 역점수(reverse scoring) 문항들에 대한 이중 동의(double agreements)를 세는 방법이다. 이 방법은 불성실 응답을 탐지하는 등 여러 목적으로 널리 사용된다(Swain et al., 2008; Weijters et al., 2013). 그런데 역점수 문항은 만들기 어렵거나 이해하기 어려울 수 있고, 신뢰도를 낮추고 별도 요인으로 분리될 위험도 있어 신중한 사용을 권고하기도 한다(Weijters & Baumgartner, 2012). 더불어 역점수 문항에 응답하는 방식에서 해석상의 이슈가 있다(Wong et al., 2003). 핵심은 동일 척도 내 여러 문항이 서로에게 정보 단서로 작동한다는 것인데, 정방향의 문항에 응답할 때 응답자는 이미 앞에서 비슷한 내용의 정방향 문항들에 응답하면서 개략적 이미지나 판단 기준을 만들게 된다. 그래서 새 문항을 볼 때 "아, 이것도 비슷한 거구나"하고 기존에 떠올린 정보를 재사용하고 굳이 새로운 기억이나 사례를 꺼내지 않는다(Weijters et al., 2009). 달리 말하면, 만족화(satisficing)로 추가 정보의 인출을 최소화하는 것이다(Krosnick, 1991). 반면, 역점수 문항에 응답할 때 문항 방향이 바뀌면서 표현이나 초점이 미묘하게 달라진다. 예를 들어, 정방향 문항인 '나는 팀에서 일하는 것을 즐긴다'에 응답한 후, 역방향 문항인 '나는 혼자 일할 때 더 편안하다'를 보면 사람들은 기존에 사용하던 기준으로 답하기보다는 다른 상황, 다른 에피소드, 다른 관점을 추가로 떠올려서 재평가하는 경향이 있다. 새로운 정보의 인출을 최대화하게 된다. 역점수 문항에 대한 응답 패턴은 문항의 방향 차이가 유발한 인지적 처리 전략의 차이, 즉, 어떤 정보를 얼마나 새로 끌어오는지와 더 관련 있을 수 있다는 것이다. 따라서 Weijters 등(2009)은 역문항에 대한 반응을 곧바로 ARS의 지표로 쓰는 건 적절하지 않을 수 있다고 보았다. Kam 등(2021)의 연구도 역문항에 대한 오응답(misresponse)이 참가자의 비합리적 행동 때문이라는 기존의 관점을 비판한다. 오히려, 특정 특성의 중간 수준에 있는 참가자에게는 겉보기에는 동등해 보이는 진술들에 대해 동의와 비동의를 모두 표현하는 것이 논리적으로 타당할 수 있다(예: "나는 키가 크지 않다. 그러나 그렇다고 키가 작은 것도 아니다."). 이러한 이유로 단일 차원의 개념을 측정하더라도 정방향 문항과 역방향 문항은 서로 다른 요인에 적재될 수 있다.[10]

10) 역문항 중에서도 정반대의(polar opposites) 표현

다특성-다방법(multitrait-multimethod: MTMM) 접근은 여러 특성을 여러 방법으로 반복 측정하여 문항반응의 변량을 특성 요인과 방법 요인으로 분리하는 방법이다(Campbell & Fiske, 1959; Podsakoff et al., 2003). 고전적 MTMM에서 방법효과는 동일한 측정방법이 여러 특성의 측정치에 걸쳐 유발하는 공통변량으로 해석된다. 응답 스타일 연구에서는 이러한 논리를 긍정형 문항과 부정형 문항의 방향에 적용할 수 있다. 구체적으로 내용 요인에는 문항의 긍정 혹은 부정 방향에 따라 서로 반대되는 부호의 요인부하량을 부여하는 반면, 모든 문항에는 동일한 방향의 방법 요인 부하량을 부여함으로써 묵인 응답 스타일(acquiescent response style: ARS)을 내용 변량과 분리한다(Billiet & McClendon, 2000). Billiet와 McClendon (2000)의 확인적 요인분석(confirmatory factor analysis: CFA) 모형에서 특성 요인은 실제로 측정하려는 내용 특성을 반영하고, 방법 요인은 문항 내용과 관계없이 긍정형과 부정형 문항 모두에 동의하려는 묵인 성향을 반영한다. 이 모형에서는 내용 요인의 부하량은 문항 방향에 따라 반대 부호를 갖도록 설정하는 반면, 모든 문항의 방법 요인 부하량은 동일한 방향과 크기, 즉 원 모형에서는 +1로 고정하였다. 이를 통해 긍정형과 부정형 문항 모두에 동의하는 반응 경향을 내용 특성과 구분할 수 있다. Billiet와 Davidov(2008)는 이 모형을 종단 패널자료에 적용하여 ARS 요인이 4년에 걸쳐 상당한 안정성을 나타낸다는 것을 확인하였다. 또한 내용 요인과 방법 요인의 상관을 0으로 제약한 모형에서도 ARS를 내용 요인과 구분하여 추정할 수 있음을 보였다. 다만 이는 모든 상황에서 내용 특성과 ARS가 실질적으로 독립적이라는 것을 의미하지는 않는다. 이와 같은 MTMM 또는 CFA 방법요인 접근은 이미 다방법 측정치나 긍정과 부정 방향으로 균형화된 문항이 포함되어 있는 경우 별도의 응답 스타

과 부정 형태로 전환된 문항(negated regular items)은 질적으로 다른 종류이므로 서로 다른 요인에 적재될 수 있다. 여기서 '정반대의 표현'은 원래 진술의 반대 의미를 가진 단어로 다시 서술한 문항이다. 예를 들어, "나는 사교적인 편이다"의 정반대의 역문항은 "나는 내성적인 편이다"이다. 이런 양극단의 표현은 새로운 진술로 받아들여지며, 참가자는 별도의 평가 과정을 거쳐야 한다. 의미 구조상 독립적인 판단을 요구하므로, 단순히 부정된 형태보다 더 많은 인지적 처리 자원이 필요하며, 정방향 문항과 동일한 차원을 측정한다고 보장할 수 없다. 반면 '부정 형태로 전환된 문항'은 기존 진술에 단순히 부정어(not, never 등)를 덧붙인 문항이다. 예를 들어, "나는 사교적인 편이다"에서 "나는 사교적인 편이 아니다"로 바꾼 것이다. 겉보기에는 정방향 문항과 동일한 의미 구조를 갖지만, 언어적 부정 처리는 인지적으로 더 복잡하다. 특히 이중 부정(double negation)이나 정서가 개입된 문장에서는 혼란을 유발할 수 있다(예: "나는 다른 사람들과 어울리는 것을 좋아하지 않는 편이 아니다"). 문장 구조의 난해함이 오응답 가능성을 높이는 것이다. Kam 등(2021)은 실험적으로 이 두 문항 유형이 서로 다른 요인에 적재된다는 것을 보여주었는데, 그들은 이러한 결과에 대해 1) 정반대의 표현은 해당 특성의 다른 차원을 측정하는 경향이 있으며, 정방향 문항과의 불일치가 발생할 수 있다, 2) 부정 문항은 같은 차원을 측정하려 하되, 언어적 부정의 인지적 복잡성 때문에 일관성 저하를 일으킬 가능성이 크다, 3) 특히 특성 수준이 중간에 있는 개인에게는 이러한 문항 쌍에 대해 동의와 비동의를 모두 보일 수 있으며, 이는 논리적으로 타당한 응답 패턴일 수 있는 것으로 해석하였다. 한편, 일반적으로 부정 문항이 문항 난이도가 더 높지만, 극단적인 표현(예, 'bad'보다는 'terrible')을 사용하면 부정 문항이 별도 요인으로 구분될 가능성은 낮아진다(Kam, 2023).

일 측정문항을 추가하지 않고도 ARS나 문항 방향에 따른 방법효과를 추정할 수 있다는 장점이 있다. 그러나 Billiet와 McClendon(2000)의 모형은 서로 다른 내용 특성을 측정하는 긍정형과 부정형 문항이 적절하게 균형을 이루는 설계에 크게 의존한다. 따라서 부정형 문항이 거의 없거나 특성 - 방법 조합과 지표의 수가 충분하지 않으면 내용 요인과 방법 요인을 안정적으로 분리하기 어렵고 모형 식별 문제가 발생할 수 있다(Eid et al., 2003; Podsakoff et al., 2003). 또한 이 접근은 ERS(extreme response style)나 MRS(midpoint response style)처럼 특정 반응범주의 선택 패턴에 기초한 응답 스타일을 직접 탐지하기 어렵다. 동일한 특성을 여러 방법으로 반복 측정하는 설계에서는 일관성 동기, 기억효과 또는 반복측정에 따른 반응 의존성이 나타날 가능성도 있다.

이러한 접근의 한계를 보완하기 위해 이후 연구에서는 응답 스타일 변량(stylistic variance)과 실제 내용 변량(substantive variance)을 보다 명확하게 분리할 수 있도록, 관심 특성과 개념적으로 관련이 없는 별도의 문항 집합을 활용하는 방법이 발전하였다(Baumgartner & Steenkamp, 2001; Weijters et al., 2008). 이는 문항 내용과 관계없이 나타나는 일반적인 동의 및 비동의 경향을 측정하려 했던 초기 연구의 전통을 확장한 것으로 볼 수 있다(Couch & Keniston, 1960). 대표적 응답 스타일 지표(representative indicators for response styles: RIRS) 방법은 관심 특성과 관련이 없는 여러 척도에서 대표성 있게 표집한 문항들을 이용하여 응답 스타일을 측정한다. 서로 다른 내용 영역의 문항을 사용함으로써 특정 문항 내용의 영향을 상쇄하고, 반복적으로 나타나는 반응범주 선택 경향을 응답 스타일로 해석하는 것이다. 이를 통해 ARS, DARS, ERS, MRS 및 NARS 등의 지표를 계산할 수 있다. 일반적으로 응답 스타일을 통제변수로 사용하는 연구에서는 응답 스타일 지표별로 약 5개 문항을 사용할 수 있으며, 응답 스타일 자체를 주요 연구대상으로 삼는 경우에는 지표별로 약 10～14개의 문항을 사용하는 것이 권장된다(Weijters et al., 2008).

RIRS를 확장한 RIRMACS(representative indicators for response styles means and covariance structure) 접근은 여러 응답 스타일을 단순한 관찰 지표가 아니라 잠재 요인으로 모형화한다. 이를 통해 응답 스타일 지표의 측정오차를 고려하면서 응답 스타일 간 수렴타당도와 변별타당도를 평가하고, 응답 스타일이 관심 내용 요인과 다른 변수 간의 관계에 미치는 영향을 추정하고 조정할 수 있다. RIRS와 RIRMACS는 여러 유형의 응답 스타일을 함께 검증할 수 있고, 추정된 응답 스타일을 후속 분석에서 공변인이나 잠재 예측변수로 활용할 수 있다는 장점이 있다. 그러나 관심 특성과 직접 관련이 없는 문항을 다수 추가해야 하므로 검사 시간과 응답자 부담이 증가하며, 기존의 2차 자료에는 필요한 대표 문항이 포함되어 있지 않을 수 있다는 현실적인 제약이 있다(Weijters et al., 2008).

조사에 별도의 응답 스타일 척도를 충분히 포함하기 어렵거나 문항 수준의 2차 자료를 분석하면서 Likert 응답을 연속변수로 간주하지 않고 범주별 반응확률을 모형화하려는 경우, 잠재계층 요인분석(latent class factor analysis: LCFA)을 활용할 수 있다. LCFA는 여러 잠재요인을 각각 몇 개의 순서화된 이산 수준으로 표현하고, 문항이 특정 요인에만 적재되도록 하는 등 CFA와 유사한 확인적 측정구조를 설

정하는 잠재계층모형의 한 형태이다(Moors, 2003; Morren et al., 2011). 이 모형에서는 내용 요인과 응답 스타일 요인을 함께 추정할 수 있으며, Morren et al.(2011)의 제한된 LCFA에서는 문항반응과 내용 요인의 관계를 순서형·단조적으로 제한하는 한편, ERS 요인과 문항범주의 관계는 명목적으로 설정하여 범주별 효과를 자유롭게 추정한다. 이에 따라 표준적인 CFA 방법요인 모형으로 표현하기 어려운 ERS의 U자형 비단조 효과를 모형화할 수 있다. 또한 긍정 및 부정 방향의 균형 문항이 확보된 경우에는 ARS를 동의 방향으로 작용하는 별도의 순서형 스타일 요인으로 추가하여 ERS와 함께 추정할 수 있다. Kieruj 와 Moors(2013)는 이러한 동시 모형을 제시하였으나, 그들의 자료에서는 ERS에 대한 증거는 강한 반면 ARS에 대한 증거는 약했다. LCFA는 별도의 응답 스타일 지표 없이도 서로 다른 내용영역의 기존 문항반응 패턴을 이용할 수 있고, 이문화 자료에서는 내용 요인과 응답 스타일 요인에 대한 집단 차이를 동시에 추정할 수 있다는 장점이 있다(Morren et al., 2011). 이 접근법은 원칙적으로 MRS나 DARS를 비롯한 다른 범주선호 패턴 및 복수의 스타일 요인으로도 확장할 수 있지만, 여러 스타일을 동시에 구분하려면 각 스타일에 대응하는 범주별 효과에 이론적 제약을 부과하고 적절한 문항구조를 확보해야 한다. 또한 잠재요인의 수준 수, 요인적재의 영 제약과 동일성 제약, 모형 식별 방식을 결정해야 하며 잠재 수준이나 스타일 요인의 수가 증가할수록 계산량과 해석의 복잡성이 커진다.

IRT 기반 응답 스타일 모델링은 응답 스타일을 문항 반응을 산출하는 잠재 과정의 일부로 다룬다. 최근의 논의를 종합하면, 이러한 접근은 크게 다차원 IRT 모형, 혼합 IRT 모형, IRTree 모형으로 구분할 수 있다(Bolt & Meng, 2025). 첫째, 다차원 IRT 모형은 관심 특성과 응답 스타일을 서로 다른 잠재 차원으로 설정하고, 하나의 문항 반응이 두 차원의 영향을 동시에 받는다고 가정한다. 이러한 흐름의 출발점 중 하나로 De Jong 등(2008)은 ERS를 관찰된 극단응답 패턴의 배후에 있는 연속적이고 안정적인 개인 특성으로 간주하는 이분 IRT 모형을 개발하여, 각 문항이 ERS를 어느 정도 반영하는지를 다르게 설정할 수 있도록 하였다. 다만 이 접근에서는 원래의 다범주 문항을 극단/비극단의 이분 지표로 변환하므로, 비극단 범주들 사이의 차이, 관심 특성과 응답 스타일의 동시 작용을 충분히 활용하지 못한다는 한계가 있었다. 이러한 한계를 보완하는 방향에서 Bolt와 Newton(2011)은 각 문항 반응이 관심 특성 요인과 ERS 요인에 동시에 관련되는 다차원 IRT(Multidimensional IRT) 접근으로 확장하였다. 이들은 하나의 문항 반응이 실제 특성 요인(substantive trait)과 응답 스타일(ERS) 요인의 영향을 동시에 받을 수 있도록 모델링하였다. 여기서 응답 스타일 요인은 CFA에서의 '방법 요인'과 유사하게 별도의 잠재 차원으로 정의되며, 다차원 IRT 모형 내에서 함께 추정된다[11]. Bolt 등(2014)은 이러한

11) 구체적으로, 원래 문항 반응(예, 1~5 Likert)을 실제 특성 요인을 추정하는 데 그대로 사용하고, 동일한 문항에 대해 ERS 요인에 적합한 범주별 슬로프 구조를 지정함으로써, 한 문항이 내용 요인과 응답 스타일 요인 모두에 적재되도록 구성된다. 예를 들어, ERS 요인에 대해서는 양극단 범주(1, 5)에 양의 슬로프, 중간 범주(2~4)에 음의 슬로프를 부여함으로써, ERS 성향이 높은 사람일수록 극단 범주를 선택할 가능성이 높아지도록 모델링한다. 또한 이 접근은 여러 개의 척도(예,

접근을 확장하였는데, 그들은 응답 스타일을 측정할 때 기존의 특성 문항 대신 '앵커링 빈넷'이라는 보조 문항을 사용했다.[12] 이 접근의 장점은 빈넷을 활용하면 거의 모든 유형의 응답 스타일을 모델링할 수 있고, 응답 스타일 요인을 특성 요인과 완전히 분리할 수 있다는 점이다. 결과적으로 이 방법은 기존의 ERS 측정 방식이 갖는 한계(예, ERS가 높은 응답자와 실제 점수가 높은 응답자의 구분이 어렵다는 문제)를 해결하며, 실제 특성 요인과 응답 스타일을 하나의 통합된 다차원 모형에서 동시에 추정할 수 있는 강점이 있다.

둘째, 혼합 IRT 모형 또는 혼합 라쉬 모형은 응답자가 동일한 반응 과정을 따른다고 가정하기보다, 서로 다른 응답 전략이나 응답 스타일을 보이는 잠재 계층이 존재한다고 가정한다. 이 접근은 ERS 집단과 비ERS 집단처럼 질적으로 다른 응답자 하위집단을 식별할 수 있으며, 각 계층 내에서 별도의 문항 및 특성 모수를 추정할 수 있다(Rost et al., 1997; Austin et al., 2006). 그러나 응답 스타일을 연속적인 강도 차이로 다루기보다 범주형 이질성으로 다룬다는 점에서, 동일 계층 안의 응답자들이 보이는 응답 스타일 강도의 차이를 충분히 반영하지 못할 수 있다.

셋째, IRTree 모형은 다분 평정 반응을 하나의 단일 판단으로 보지 않고, 여러 하위 의사결정의 연쇄로 분해한다. 예를 들어 응답자는 먼저 긍정 방향과 부정 방향 중 하나를 선택하고, 이후 극단 범주를 선택할지 또는 중간 범주를 선택할지를 결정한다고 볼 수 있다. Böckenholt(2012)는 ERS나 ARS가 단순 편향이 아니라 행동적 의사결정 전략임을 수학적으로 설명하고자 하였다. 그가 제시한 방법은 리커트 응답을 하나의 단일 판단으로 보지 않고, '중간점을 넘는가?', '긍정 방향을 선택하는가?', '극단을 선택하는가?'와 같은 여러 하위 의사결정의 연쇄로 분해한다. 그리고 각 하위 결정은 특성 또는 응답 스타일의 영향을 받을 수 있는데, 이를 가짜-결과(pseudo-outcomes)라는 이진 변수들로 표현해 모형에서 각각을 구분한다. 이 방식은 단일 응답 안에 섞여 있던 '내용 판단'과 '응답 습관(스타일)'을 분리해낼 수 있어, ARS, DARS뿐 아니라 ERS, MRS까지 더욱 정밀하게 탐지할 수 있다는 장점이 있다. 이러한 tree 구조를 사용하면 내용 판단과 응답 스타일 판단을 서로 다른 노드로 분리할 수 있으므로, ARS, DARS, ERS, MRS를 응답 과정 수준에서 동시에 다룰 수 있다(Böckenholt,

---

Enjoyment of Science, Value of Science)를 동시에 분석할 때 ERS 요인의 안정성을 활용해 ERS를 더 정확하게 추정할 수 있으며, 결과적으로 실제 특성 추정에서의 편향을 줄이고 정확도를 높이는 효과를 보였다.

12) 앵커링 빈넷(anchoring vignettes)은 모든 응답자에게 동일하게 제시되는 '짧은 상황 설명(시나리오)'이나 '가상의 인물 · 사례'로, 각 응답자가 이 시나리오를 평정하도록 하는 보정 도구이다. 즉, 응답자가 실제 문항에 답하기 전에 혹은 답한 후에, "이런 상황/사람은 이 척도에서 어느 정도라고 생각하나요?"같은 형식으로 정해진 준거 사례를 평가하게 한다. 앵커링 빈넷의 목적은 사람들이 척도를 사용하는 기준이 서로 상이한 문제를 조정하거나 탐지하기 위한 것이다. 예를 들어, 어떤 사람은 깐깐한 응답자처럼 매우 높은 기준을 가지고 있고, 어떤 사람은 모든 것을 후하게 평가할 수 있다. 빈넷에 대한 응답 패턴을 보면 ARS, ERS, MRS와 더불어 척도 사용상의 개인차 등을 특성 변인과 분리해서 파악할 수 있다. 빈넷은 모든 응답자에게 동일하게 제시되는 짧은 시나리오이기 때문에, 사람마다 척도를 쓰는 방식이 어떻게 다른지(예, ARS, ERS 등)를 특성 변인과 섞이지 않은 채 순수하게 측정할 수 있다.

2012; De Boeck & Partchev, 2012; Böckenholt & Meiser, 2017). 다만 IRTree는 어떤 tree 구조가 이론적으로 적절한지에 대한 사전 가정이 필요하고, 원래의 다분 반응을 여러 이분 과정으로 재표현하는 과정에서 해석이 복잡해질 수 있다.

지금까지 살펴보았듯이, 응답 스타일을 평가, 탐지하기 위해 다양한 방법들이 제안되었으며, 이들 방법은 응답 스타일을 범주형 변수로 볼지 연속형 변수로 볼지에 대한 기본 가정, 외현 변수와 잠재 변수를 사용하는 접근, 응답 스타일을 모델링하기 위해 별도의 항목을 사용할지 여부, 그리고 특성 변량과 응답 스타일 변량을 분리하는 방법 등에 따라 차이가 있다. 그러나 이러한 다양한 방법이 응답 스타일을 탐지하는 능력을 포괄적으로 비교한 연구나 방법 간의 수렴 타당도를 조사한 연구는 아직 많이 이루어지지 않았다. 앞으로 이 이슈에 대한 더 많은 연구가 필요하다.

응답 스타일의 조정

조사나 검사 자료에서의 변량은 관심 특성의 개인차를 반영하는 참 변량과 무선 오차 변량으로 구성된다. 그러나 응답 편향이 자료에 존재하면 그것은 체계적 변량인 것이며, 적절히 모델링되지 않으면 참 변량으로 가려지게 된다. 앞서 언급했듯이, 응답 편향은 변인 간의 허위(spurious) 관계를 만들어내거나, 관계를 과대 추정함으로써 통계분석의 결과에 영향을 미친다. 응답 스타일의 영향에 대해 특성 점수를 조정하는 다양한 사후 방법이 제안되었다. 이들 중 일부는 이미 응답 스타일의 탐지 부분에서 소개하였으며, 응답 스타일의 탐지와 조정은 함께 이루어진다. 따라서 본 절에서 말하는 조정은 탐지와 완전히 분리된 별도 절차만을 의미하지 않는다.

선형 회귀 기반 조정법은 Webster(1958)가 처음 제안한 절차로, 먼저 응답자가 특정 응답 스타일을 얼마나 보이는지를 나타내는 빈도 지표를 산출하는 것에서 출발한다. 이후 관찰된 특성 점수를 응답 스타일 지표에 회귀시키고, 이 회귀에서 얻은 기대값과 실제값의 차이(회귀 잔차)를 특성 점수의 '조정된' 형태로 사용한다. 이 잔차는 응답 스타일에 의해 설명되는 변량이 제거된 값으로 간주된다. 이 방법은 하나의 응답 스타일 지표뿐 아니라 여러 지표를 동시에 포함하는 다중 회귀 형태로 확장될 수 있으며, 이때 특성 문항과 중복되지 않는 이질적 문항을 사용해 응답 스타일 지표를 구성한다(Baumgartner & Steenkamp, 2001; Weijters et al., 2008). 이 방법의 장점은 구조방정식 모형이나 요인점수 같은 모형 기반 추정치를 요구하지 않고, 단순 합산 점수(raw sum scores)만으로도 구현할 수 있어 적용이 매우 간단하다는 데 있다. 반면, 이 방법은 특성과 응답 스타일 간의 관계가 선형적이라고 가정하는데, 실제 자료에서는 이 가정이 성립되지 않는 경우도 있어 이는 중요한 한계로 지적된다.

응답자가 보이는 ERS나 MRS의 정도를 하나의 지표(index)로 구성하여 분석 과정에서 통제하는 방식도 있다. 대표적 예로 Johnson 등(2005)이 제안한 ERS 총점 지표가 있다. 각 문항에서 극단값을 선택하면 1로, 그렇지 않으면 0으로 코딩한 뒤 이를 합산하면 개인의 ERS 성향을 나타내는 간단한 지표가 된다. 이 지표는 회귀분석 등에서 통제 변인으로 포함

되어 ERS에 의한 편향을 부분적으로 조정할 수 있다. 다만 이 방법의 타당도는 지표를 구성하는 문항들이 내용에서 충분히 이질적일 때 가장 높다. 문항 내용이 서로 지나치게 유사하면, 어떤 사람의 극단적 응답이 실제 태도 때문인지, 아니면 응답 스타일 때문인지 구분하기 어려워지기 때문이다. MRS에도 비슷한 방법을 적용할 수 있는데, 서로 관련 없는 내용의 문항들을 제시해 응답자가 주제와 상관없이 늘 중간 범주를 선택하는지 확인하는 방식도 가능하다. 그러나 이러한 문항을 설문에 추가하면 문항 수가 늘어나고, 응답자에게 부자연스럽거나 부담스럽게 느껴질 수 있다는 한계가 있다.

앞서 탐지 절에서 살펴본 IRT 기반 접근은 응답 스타일의 탐지뿐 아니라 특성 점수의 조정에도 활용된다. 혼합 라쉬 모형(mixed Rasch models)은 응답 행동에서 체계적으로 다른 패턴을 보이는 하위집단을 식별하는 데 활용될 수 있다. 이 모형에서는 참가자들이 서로 다른 잠재 계층(latent classes)에 속한다고 가정하며, 각 계층 내에서 별도의 특성 추정치를 산출한다. Rost 등(1997)은 이러한 절차가 응답 스타일로 인한 체계적 영향을 분리하고 조정하는 효과가 있다고 보았다. 혼합 라쉬 접근의 중요한 특징은 최종적으로 산출된 특성 추정치들이 동일한 척도상에서 표현되기 때문에 하위집단 간 비교가 가능하다는 점이다. 또한 이 방법은 특성 문항만을 기반으로 모형을 구성하므로, 응답 스타일을 측정하기 위한 추가 문항이 필요하지 않다는 장점이 있다. 그러나 이 방법의 한계는 ERS 집단과 비ERS 집단처럼 '질적' 차이를 보이는 집단들을 구분할 수는 있지만, 응답 스타일 강도의 '양적' 차이를 추정하는 데에는 적합하지 않다는 점이다.

ARS를 설명하기 위해 특성 요인에 방법 요인을 추가하는 전통적 모형에서는 긍정 및 부정 문항 모두가 동일한 방향으로 방법 요인을 반영한다고 가정한다(Maydeu-Olivares & Coffman, 2006). 그러나 이것은 ARS에만 적용이 가능하고, 문항 방향성이 균형 잡힌 척도가 필요하다는 제약이 있다. 이 한계를 보완하기 위해 다차원 문항 반응 모형(Multidimensional IRT)은 특성 요인과 더불어 ERS와 같은 응답 스타일을 별도의 차원으로 모델링한다. 이 경우 개인의 특성 수준은 '특성 차원'과 '응답 스타일 차원'의 정보를 동시에 고려하여 추정되며, 따라서 특성 점수는 ERS로부터 부분적으로 조정된 것으로 간주된다(Bolt & Johnson, 2009). 여러 특성 영역에 걸친 문항들을 ERS 차원을 추정하는 데 함께 사용하면, 특성에 기인된 변량과 ERS에 기인된 변량을 더 잘 분리할 수 있다는 점도 제시되었다(Bolt & Newton, 2011). 이러한 접근은 주로 ERS를 다루는 데 활용되었지만, ARS나 MRS 등 다른 응답 스타일에도 쉽게 확장될 뿐 아니라, 복수의 응답 스타일을 동시에 모델링하는 데에도 적용 가능하다(Wetzel & Carstensen, 2017). 이 방법의 주요 장점은 동일한 문항 집합으로 서로 다른 차원(특성과 응답 스타일)을 동시에 모델링할 수 있다는 점이다. 그러나 이 점은 단점이기도 하다. 특히 특성 차원이 하나나 두 개에 불과할 때, 특성 요인과 응답 스타일 요인 간 종속성이 높아져 '높은 특성 수준을 가진 사람'과 '높은 ERS를 보이는 사람'을 구분하기 어려워질 수 있다. 또 다른 한계는, 이 접근으로 추정된 응답 스타일이 척도 전반에 일반화되지 않고 특정 설문 척도(scale-specific)에만 적용된다는 점이다. 탐지 방법에서 이미 보았듯이, 이러한 문제를 해결하기 위한 대안

으로 Bolt 등(2014)은 특성 요인을 추정할 때는 평정 척도 자료를 사용하고, 응답 스타일 요인을 추정할 때는 앵커링 빈넷(anchoring vignettes) 자료를 사용하는 모형을 제안하였다. 이 방식은 척도 특유성 문제를 완화할 잠재력이 있다.

마지막으로, 탐지 방법에서 언급되었듯이, Böckenholt(2012)는 응답 행동을 특성과 관련된 하위 과정과 응답 스타일과 관련된 하위 과정으로 분리하여 모델링할 수 있다는 아이디어에 근거해, 문항반응이론(IRT)을 응용한 접근을 제안하였다. 이 접근에서는 응답자가 다분 평정 척도(예: 1~5점, 1~7점)에 표시한 하나의 최초 응답을, 두 종류의 이분 문항으로 재코딩한다. 하나는 '응답자가 해당 특성 수준에 상응하는 방향으로 문항에 동의(비동의)했는가'와 같이 관심 특성과 직접 관련된 하위 과정을 나타내는 유사 문항(pseudo-item)이고, 다른 하나는 '극단 범주를 선택했는가(ERS)' 또는 '중간 범주를 선택했는가(MRS)'와 같이 응답 스타일(ERS, MRS 등)과 관련된 하위 과정을 반영하는 유사 문항이다. 이렇게 하면, 원래의 다분 평정 반응 하나가 '특성 관련 이분 문항'과 '응답 스타일 관련 이분 문항' 두 갈래로 나뉘어, 각각을 별도의 잠재 요인으로 IRT 모형에 투입할 수 있게 된다. Khorramdel과 von Davier(2014)는 이러한 IRT 기반 다과정 모형을 성격 5요인 설문 자료에 적용하였다. 구체적으로, 응답 과정 중 특성 요소와 연관된 유사 문항들에 대한 반응만을 사용해 잠재 특성 요인을 추정한 결과, 이 요인이 응답 스타일(ERS, MRS 등)의 영향으로부터 상당 부분 분리되어 추정될 수 있음을 보였다. 즉, 동일한 응답 자료로부터 특성 차원과 응답 스타일 차원을 분리 추정함으로써, 응답 스타일의 교란을 통제한 특성 측정이 가능함을 경험적으로 보여준 셈이다. 그러나 이 접근에는 뚜렷한 한계도 있다. 다분 평정 척도에서 얻어지는 원래 정보는 범주 간 '정도의 차이(예: 약간 동의 - 대체로 동의 - 전적으로 동의)'를 반영하는 서열 및 강도 정보를 포함하지만, 이를 이분 문항(예: 동의 대 비동의, 극단 대 비극단, 중간 대 비중간)으로 재코딩하는 과정에서 이 미세한 정보가 본질적으로 동의/비동의 수준의 구분으로 축소된다. 그 결과, 특성에 대한 세밀한 변별력이 줄어들 수 있고, 척도의 정보량 및 신뢰도에도 손실이 발생할 수 있다는 점이 이 방법의 주요 단점으로 지적된다(Bolt & Meng, 2025).

응답 스타일을 조정한 또는 응답 스타일의 효과를 통제한 특성 추정치를 얻기 위한 다양한 방법이 제안되었지만, 실제로 각 방법이 응답 스타일 변량을 얼마나 효과적으로 제거하는지에 대한 체계적 비교 연구는 거의 이루어지지 않았다. Wetzel 등(2016)은 이러한 공백을 메우기 위해 여러 방법을 비교한 결과, 다차원 IRT 접근이 혼합 라쉬 모형보다 응답 스타일로부터 조정된 특성 추정치를 산출하는 데 더 적합하다고 보고했다. 흥미로운 점은, 이 연구에서 ERS가 존재하는 자료를 단일 차원 IRT 모형으로 분석했을 때(즉 응답 스타일을 전혀 모델링하지 않았을 때조차), 잠재 특성 수준의 회복이 극단적으로 나빠지지는 않았다는 사실이다. 이는 ERS가 존재하더라도 단일 차원 모형이 완전히 실패하지는 않을 수 있음을 시사한다. 최근의 IRT 기반 연구(Bolt & Meng, 2025)는 동일한 자료라도 모형이 응답 스타일을 연속 잠재차원으로 보는지, 잠재계층으로 보는지, 또는 tree 구조의 하위 결정 과정으로 보는지에 따라 특성 점수와 표준오

차가 달라질 수 있음을 보였다. 따라서 응답 스타일 조정은 모든 연구에서 기계적으로 적용해야 하는 절차라기보다, 자료 구조와 연구 목적에 맞게 선택해야 하는 분석 전략으로 간주하는 것이 적절할 것이다. 실제 연구에서는 원자료 분석과 응답 스타일 조정 분석을 함께 제시하면서, 주요 결론이 조정 방법에 얼마나 민감한지 보고하는 방식이 현실적이라 할 것이다.

응답 스타일 효과를 가장 효과적으로 조정하여 신뢰할 수 있는 특성 추정치를 얻는 최적의 방법에 대해서는 여전히 명확한 결론이 없으며, 이 분야에서는 추가적 비교 연구와 검증이 필요하다. 무엇보다 응답 스타일 효과를 효과적으로 조정하여 특성 추정치를 얻는 방법에 대해서는 더 많은 연구가 필요하다.

응답 스타일의 예방

사전에 응답 스타일을 예방하는 방법도 여러 가지 제안되었다(Stantcheva, 2023). 첫째, ARS를 줄이기 위해 질문이나 문항을 모호하게 구성하지 않는다. 모호하거나 복잡한 문항에 직면하면 응답자는 의미를 해석하려 애쓰기보다, 부담을 피하기 위해 단순히 '동의한다'라고 응답하는 경향이 있다. 따라서 간결하고 명확한 문항 작성이 기본적인 예방책이다.

둘째, '동의 - 비동의' 형식의 질문 자체를 피한다(Krosnick, 1999). '동의 - 비동의', '참 - 거짓', '예 - 아니오' 형식의 응답을 요구하는 질문은 ARS를 직접적으로 유발할 수 있다. 대신 문항 내용에 맞춘 구체적이며 좀 더 의미 기반의 응답 범주를 제시하는 것이 바람직하다. 예를 들어, "나는 건강이 좋다"라는 문항에 대해 '완전히 비동의-어느 정도 비동의-중간-어느 정도 동의-완전히 동의'의 응답 범주를 제시하기보다는 '당신은 자신의 건강을 어떻게 평가하는가?'라는 질문으로 바꾸고 '매우 나쁨-나쁨-중간-좋음-매우 좋음'의 선택지를 주는 것이다.

셋째, 양극 구조의 문항에서는 모든 가능한 관점을 반영한 응답 범주를 제공한다. 즉, 예 - 아니오(또는 진 - 위) 형식 대신, 응답자가 실제 의견 스펙트럼을 표현할 수 있는 선택지를 제시한다. 예를 들어, "학자금 대출은 모든 사람에게 탕감되어야 한다고 생각하나요?"에 대해 '예-아니오'를 제시하기보다 '학자금 대출은 모든 사람에게 탕감되어야 하는가, 아니면 그렇지 않아야 하는가?'와 함께 '탕감되어야 한다-탕감되지 않아야 한다'처럼 질문 내용을 직접 반영한 선택지를 제공하는 것이 더 낫다.

넷째, 긍정 표현 문항과 부정 표현 문항을 함께 사용하면, ARS를 어느 정도 줄일 수 있다. 예를 들어, '나는 일할 때 집중력이 높다'의 긍정 표현과 '나는 일할 때 집중력을 유지하기 어렵다'의 부정 표현의 경우 방향은 다르지만, 같은 특성(집중력)을 묻고 있다. 문항을 이렇게 절반은 긍정 방향, 절반은 부정 방향으로 섞어 놓으면 정상적인 응답은 긍정 문항과 부정 문항에서 반대 방향의 응답을 할 것이지만, ARS가 강한 사람은 두 문항에 모두 동의할 가능성이 있다. 이런 패턴을 보면 ARS 영향력을 파악하거나 부분적으로 상쇄할 수 있다. 긍정 표현 문항과 부정 표현 문항을 신중하게 함께 사용하면, ARS를 탐지하거나 통계 모형에서 분리하는 데 도움이 될 수 있다. 예를 들어, 동일한 구성개념을 긍정 방향과 부정 방향에서 모두 묻는 문항이 포함되어 있

으면, 문항 내용과 관계없이 모든 문항에 동의하는 응답 패턴을 확인할 수 있고, CFA 방법 요인 모형에서는 내용 요인과 묵인 응답 요인을 구분하는 데 필요한 구조적 정보를 제공할 수 있다. 그러나 이것은 기계적으로 역문항을 추가하는 것이 항상 바람직하다는 뜻은 아니다. 긍정 표현 문항과 부정 표현 문항을 함께 사용하는 방법은 긍정 및 부정 문항이 모두 묵인(acquiescence)에 비슷한 정도로 민감할 때 효과적이다. 왜냐하면, 어떤 문항은 그 자체로 더 동의하기 쉬운 문항(예: 사회적 바람직성이 높은 문항)일 수 있고, 다른 문항은 원래 '동의하기 어려운' 문항일 수 있다. 만약 긍정 문항과 부정 문항 중 한쪽만 ARS의 영향을 더 많이 받는 문항들이라면, 긍·부정 균형을 맞췄더라도 ARS를 정확히 감지하거나 상쇄하기 어렵다. 역문항은 부정문 처리부담, 긍정문과 부정문의 의미 비대칭성, 문항 극단성, 번역 등가성 문제로 인해 신뢰도를 낮추거나 정방향 문항과 별도 요인으로 적재될 수 있다. 따라서 긍정/부정 문항의 병행은 ARS를 줄이는 단순 처방이 아니라, 응답 스타일을 탐지하고 모형에서 식별하기 위한 설계 전략으로 이해해야 한다. 이 전략을 사용할 때는 예비조사, 인지 면접, 요인구조 점검, 문항 기능 검토를 통해 긍정 문항과 부정 문항이 실제로 같은 구성개념의 반대 방향을 측정하는지 확인하는 것이 필요하다.

다섯째, 집단 수준의 ARS를 추정하려면 응답자들에게 두 문항 버전을 무작위로 배정함으로써 집단(국가, 연령 집단 등) 단위의 ARS 경향을 추정할 수 있다. 예를 들어, '정부는 세금을 더 낮춰야 한다'와 반대 표현인 '정부는 세금을 더 낮출 필요가 없다'의 두 문항을 각각 다른 사람들에게 무작위로 제시하는 것이다. 만약 한 집단에 ARS가 강하게 존재한다면 긍정적 버전을 받은 사람과 부정적 버전을 받은 사람 모두 '동의'하는 식의 패턴이 나타날 것이다. 즉, 문항의 실제 내용과 상관없이 그냥 동의 버튼을 누르는 경향이 발견되면 그 집단의 ARS 수준을 추정할 수 있다.

여섯째, 온라인상의 ESM이나 EMA 연구에서, ERS를 줄이기 위한 한 가지 방법은 응답자가 0~100%처럼 연속적인 선(line) 위에서 자신의 상태를 표시하는 방식인 연속 슬라이더(continuous slider)나 시각 아날로그 척도(visual analog scale)를 사용하는 것이다.

이러한 설계 기반의 예방 방법은 조사 문항을 처음 설계하는 단계에서부터 응답 스타일이 나타날 가능성을 줄이는 것이다. 이 접근은 응답자가 의미를 깊이 처리하지 않고 기계적으로 응답하는 만족화(satisficing)를 최소화하는 데 초점을 둔다. 이를 위해 문항과 응답 범주는 맥락적으로 명확해야 하고, 서로 분명히 구분되는 선택지를 제공해야 한다. 예를 들어, 3점 척도처럼 응답 범주가 너무 적으면 응답을 극단값 또는 중간값으로 몰아가 ERS나 MRS처럼 보이게 된다. 또한 MRS를 줄이기 위해 중간 범주를 삭제하는 것은 적절한 해결책이 아니다. 실제로 응답자가 중간적 태도를 가질 가능성이 있기 때문이다. 그보다는 5점 척도에서 7점이나 9점 척도로 선택지 수를 늘려 응답자가 어쩔 수 없이 중간값을 선택하는 상황을 줄이는 것이 더 바람직하다.

비교 문화 연구에서의 응답 스타일

국가 간 태도나 가치를 비교하는 연구는 대개 리커트 척도 문항에 대한 응답을 국가 단

위 평균으로 비교하는 방식을 사용한다. 그러나 사람들의 응답은 문항의 실제 내용뿐 아니라 응답 스타일의 영향을 받기 때문에, 국가 비교 연구는 관심 현상의 국가 간 실제 차이를 포착하기보다, 설문 응답 방식의 차이를 반영하는 것일 수 있다. 따라서 비교 문화 맥락에서 응답 스타일은 동일한 심리적 구조를 가진 응답자 집단이 응답 방식의 차이로 서로 다른 결과를 산출할 수 있다는 점에서 중요한 위협 요소이다. 예를 들어, Stening과 Everett (1984)는 9개 국가의 주재원 및 현지 관리자를 대상으로 한 조사에서 국가에 따라 ERS와 MRS에서 두드러진 차이가 있으며, 이것이 편포된 자료에 이르게 한다는 것을 보여주었다.

응답 스타일의 발생 기제는 크게 세 가지 층위에서 이해할 수 있다(Kemmelmeier, 2016). 첫째는 인지적 기제이다. 설문 응답은 문항을 읽고, 그 의미를 해석하며, 관련 정보를 기억에서 인출하고, 판단을 형성한 뒤, 이를 적절한 응답 범주에 대응시키는 과정을 요구한다. 이 과정에서 응답자가 충분한 인지적 자원이나 동기를 투입하지 못할 경우, 최적의 응답을 산출하기보다 충분히 그럴듯한 응답을 선택하는 만족화(satisficing) 전략을 사용할 수 있다. 이 경우 응답자는 문항의 세부 의미를 깊이 처리하기보다 동의하기 쉬운 방향을 선택하거나, 중간 범주를 반복적으로 사용하거나, 익숙한 응답 범주를 기계적으로 선택할 수 있다.

둘째는 의사소통 기제이다. 설문 응답은 단순한 개인 내 판단의 결과가 아니라, 연구자나 질문자에게 자신의 의견을 전달하는 의사소통 행위이기도 하다. 따라서 문화적으로 선호되는 의견 표출 방식은 응답 스타일에 영향을 미칠 수 있다. 예를 들어 어떤 문화에서는 자기 입장을 분명하고 강하게 표현하는 것이 솔직함이나 확신의 표시로 받아들여질 수 있지만, 다른 문화에서는 극단적 표현을 피하고 완곡하게 말하는 것이 신중함이나 예의로 여겨질 수 있다. 이러한 차이는 ERS나 MRS의 차이로 나타날 수 있다. 마찬가지로 '항상', '전혀', '절대'와 같은 절대적 표현을 사용하는 데 대한 문화적 편안함이나 거부감도 응답 범주 선택에 영향을 줄 수 있다. 이런 점에서 응답 스타일은 단순한 척도 사용 습관이 아니라, 문화적으로 학습된 표현 방식의 일부로 이해될 수 있다. 또한 문항의 의미가 모호하거나, '자주', '가끔', '거의 없다'와 같은 언어적 빈도 표현이 문화권마다 다르게 해석될 경우, 동일한 경험이나 태도도 서로 다른 응답 범주로 표현될 수 있다.

셋째는 규범과 가치의 기제이다. 응답 스타일은 개인이 속한 사회의 관계 규범, 권위 구조, 자기표현 규범, 그리고 바람직한 사람됨에 대한 문화적 기대와도 관련될 수 있다. 예를 들어 조화, 체면, 겸손을 중시하는 문화에서는 자신의 의견을 강하게 드러내기보다 중간적이거나 완화된 응답을 선택하는 경향이 나타날 수 있다. 반대로 자기표현, 확신, 경쟁성, 명확한 입장 표명을 중시하는 문화에서는 극단 범주를 사용하는 것이 더 자연스럽게 받아들여질 수 있다. 또한 권위에 대한 순응이나 사회적 기대에 민감한 문화적 맥락에서는 묵인 응답(ARS)이 강화될 수 있으며, 사회적으로 승인되는 방향의 응답을 선택하려는 경향도 함께 나타날 수 있다. 규범이 강하고 일탈에 대한 허용도가 낮은 문화적 경직성(tightness)이 높은 사회에서도 응답자는 자신의 실제 판단을 그대로 표현하기보다, 사회적으로 적절하다고 여겨지는 응답 범주를 더 신중하게 선택할 가

능성이 있다.

이처럼 응답 스타일은 단일한 원인으로 설명되기 어렵다. 인지적 층위에서는 문항 처리 부담과 만족화 전략이, 의사소통 층위에서는 의견과 감정을 표현하는 문화적 방식이, 규범과 가치 층위에서는 조화, 체면, 권위, 자기표현, 규범의 강도와 같은 문화적 가치가 응답 과정에 개입한다. 응답 스타일은 한편으로 측정의 타당도를 위협하는 방법 효과이지만, 다른 한편으로는 특정 문화권에서 의견을 표현하고 사회적 관계를 조율하는 방식이 설문 응답에 반영된 결과로도 이해될 수 있다.

한편, 응답 스타일이나 사회적 바람직성 응답(SDR)과 같은 응답 편향은 주로 모호한 태도, 가치, 자기평가 문항에서 자주 발생한다. 왜냐하면 이러한 문항들은 응답자가 문항 의미나 응답 범주를 주관적으로 해석할 여지가 크기에 언어적 표현 방식이나 문화적 규범의 영향을 쉽게 받을 수 있기 때문이다. 그러나 He와 van de Vijver(2016)는 22개국 OECD 성인 기초역량 조사 자료를 분석하여, 응답 스타일이 태도 문항에만 국한되지 않음을 보여주었다. 이들은 응답자가 제공한 사실 보고형 문항, 인지검사 자료, 면접자 관찰 자료를 활용하여 ERS, MRS, ARS를 분석하였다. 그 결과, 행동 빈도나 경험 여부처럼 비교적 명확한 응답 기준을 가진 사실 문항(factual items)에서도 응답 스타일이 나타났으며, 이러한 응답 스타일은 기존의 태도 및 자기평가 문항에서 발견된 것과 유사한 국가 간 패턴을 보였다.

문화 및 국가 수준에서 응답 스타일의 차이

먼저 응답 스타일에서의 문화 또는 국가 간 차이를 보여주는 연구를 살펴보자. Clarke(2000)는 응답자가 속한 국가 또는 문화 집단에 따라 ERS 수준이 달라진다는 점을 보고하였다. 즉, ERS는 국가 간 차이뿐 아니라 동일 국가 내 민족적, 언어적, 사회문화적 하위집단 간에도 다르게 나타날 수 있으며, 이러한 차이는 응답 척도의 포맷과도 상호작용할 수 있다고 보았다. Meisenberg와 Williams(2008)의 연구에서 평균 IQ 수준이 낮은 국가들이 더 높은 ERS를 보이며, 부패한 국가들에서 ERS와 ARS 모두 높게 나타나며, 민주주의와 정치적 자유는 ARS와 ERS와 유의한 관계가 없는 것으로 나타났다. Rammstedt 등(2017)은 4만 명에 이르는 20개 유럽 국가의 유럽 사회조사(ESS) 자료를 이용하여 ARS의 개인 수준 및 국가 수준 결정요인을 동시에 추정하는 다층모형을 제시하였다. 분석 결과, 전체 ARS 변량의 15%가 국가 간 차이에서 기인했으며, 국가 수준에서는 부패 수준이 가장 강력한 예측 요인으로 나타났고, 집합주의 역시 ARS를 유의하게 증가시키는 방향으로 작용했다. 반면 국가의 경제적 부(wealth)는 추가 설명력을 보이지 않았다. 개인 수준에서는 낮은 교육 수준과 높은 보수성(conservatism)이 ARS를 증가시키는 요인으로 나타났으며, 연령과 성별의 효과는 통계적으로는 유의하나 실질적으로는 미미하였다. 전체적으로 국가 수준 예측 변인은 ARS의 74%, 개인 수준 변인은 약 10%의 변량을 설명하여, ARS가 개인적 성향의 산물이기보다는 사회문화적 환경의 영향을 더 크게 받는 체계적 응답 스타일로 나타났다. 즉, ARS는 정치 수준, 경제 구조, 사회적 신뢰, 문화 가치가 중첩된 다수준 현상이라고 할 수 있다. 또한 그들은 ARS가 높은 국가일수록 성격 척도나 가치 척도의 요인 구조와 신뢰도에서 더 큰 왜곡이 발생한다는 점을 지적하며, ARS를 무시한 상

태에서 국가 간 평균 차이를 비교하는 것은 상당한 위험을 내포한다고 지적하였다.

문화 차원 연구

북미와 동아시아의 고등학생들을 대상으로 한 비교 문화 연구를 수행한 Chen 등(1995)은 개인주의가 ERS와는 정적인 관계, MRS와는 부적인 관계가 있음을 보였다. Johnson 등(2005)은 5개 대륙 19개국[13]의 종업원들을 대상으로 수집한 자료를 통해 Hofstede의 네 가지 문화 지향성이 ERS와 ARS에 미치는 영향을 위계적 선형모형으로 분석하였다. 분석 결과, 권력거리와 남성성은 각각 독립적으로 ERS와 정적인 관련성을 보이고, 개인주의, 불확실성 회피, 권력거리, 남성성은 각각 ARS와 부적인 관련성을 보이는 것으로 나타났다. De Jong 등(2008)도 ERS와 개인주의, 권력거리, 남성성 간의 정적인 관계를 발견하였다. Lau 등(2005, 2013)은 '만족'과 '행복'과 같은 개념의 인식에서 문화 간 차이에 주목하였고, 주관적 안녕감(SWB) 조사에서 아시아인들은 해당 질문의 리커트 척도 응답에서 코카서스인들보다 ERS가 덜 나타나고 강렬한 정서도 자제하면서 적당히 표현하는 경향을 보여 MRS가 더 많이 나타남을 보였다.

다른 연구들(Harzing, 2006; Smith, 2004)도 전반적으로 다수의 연구가 개인주의 문화(예, 미국, 북유럽)에서 ERS가 더 높게, ARS가 집합주의 문화에서 더 높게 나타난다고 보고하였다. 이러한 결과는 자기표현(self-expression)을 중시하는 개인주의 문화에서는 강한 확신을 드러내는 ERS가 사회적으로 장려되지만, 집합주의 문화에서는 조화와 대인 간 순응성을 강조하는 규범이 ARS 및 MRS 응답을 촉진한다는 해석과 부합한다.

그러나 이러한 해석과 일치하지 않는 결과들도 있다. 미국 및 필리핀 학생을 대상으로 성격 검사 응답을 조사한 Grimm과 Church (1999)의 연구에서 ARS 또는 ERS에 미치는 개인주의의 일관된 효과가 나타나지 않았고, MRS에 미치는 문화의 영향도 발견하지 못하였다. 한편, 26개국을 대상으로 한 Harzing (2006)의 연구는 응답 스타일의 문화적 기원을 더 정밀하게 파악하기 위해 기존의 Hofstede 문화 차원과 함께 GLOBE 연구(House et al., 2004)의 문화 차원도 활용하였다. GLOBE 연구에서는 각 문화 차원을 가치 지향적인 'should'(사회가 바람직하다고 여기는 규범이나 당위성)와 관행을 나타내는 'as is'(사회에서 실제 관찰되는 행동이나 관행)의 두 범주로 구분한다. 문화 변인이 ERS, ARS, MRS 등에 미치는 효과가 어떻게 달라지는지 분석한 결과, 문화 차원과 응답 스타일 간의 관계는 일관되게 유지되지 않았으며, 특정 문화 요인이 응답 스타일과 나타내는 관계의 방향성과 유의성은 Hofstede 지수를 사용할 때와 GLOBE 지수를 사용할 때 서로 다르게 나타나는 경우가 있었다. 이는 문화 차원의 개념화(가치 대 관행)나 측정 방식에 따라 응답 스타일에 대한 문화적 영향의 해석이 달라질 수 있음을 시사하며, 응답 스타일 연구에서 문화 변인 선택 자체가 실증적 결론에 체계적 차이를 낳을 수 있음을 보여준다.

또한, 대규모 자료를 분석한 Marín 등(1992)의 연구에서 히스패닉 응답자(스페인과 라틴계 문화권)는 비히스패닉 백인 응답자들보다

13) 호주, 벨기에, 브라질, 체코, 독일, 헝가리, 인도, 일본, 말레이시아, 포르투갈, 터키, 영국, 멕시코, 필리핀, 폴란드, 싱가포르, 홍콩, 프랑스, 이탈리아

ERS를 더 선호하는 분명한 결과를 보였을 뿐만 아니라, ARS도 높게 나타났다. 유사하게 Van Herk 등(2004)의 연구에서 지중해 국가들(그리스, 이탈리아, 스페인)이 서유럽 국가들(영국, 독일, 프랑스)보다 더 높은 ARS와 ERS 수준을 보였으며, 이것을 Hofstede의 문화 차원인 개인주의가 감소하면 ARS와 ERS는 증가하는 것으로 해석하였다.[14] 이러한 연구 결과와 관련하여, Benítez 등(2016)은 양적 및 질적 분석을 모두 사용한 국가 간 차이 연구를 통해 ERS가 단순한 척도 사용 습관이 아니라 문화적 정보의 처리 전략과 정서 규범의 영향을 받을 수 있음을 보여주었다. 먼저 그들은 여러 국제 연구에서 수집된 삶의 질(QoL) 지표로부터 ERS 지수를 산출하여 스페인과 네덜란드 응답자를 비교하였다. 분석 결과, 동의형 응답에서는 스페인 응답자가 네덜란드 응답자보다 ERS를 더 많이 보였지만, '전혀 없다(never)'를 포함한 빈도 척도에서는 오히려 네덜란드 응답자가 더 높은 ERS를 보이는 반대 패턴이 나타났다. 다음으로 이러한 차이를 해석하기 위해 두 나라에서 각각 25명을 대상으로 인지적 인터뷰를 한 결과, 네덜란드 응답자들은 각 문항을 비교적 독립적으로 평가하는 경향이 있었던 반면, 스페인 응답자들은 문항의 세부 문맥보다 정서적 직관에 더 의존하는 경향을 보였다. 또한 스페인 응답자들은 '전혀 없다'라는 표현에서 더 강한 함의를 느꼈으며, 직무만족과 같이 개인의 안정성이나 불안정성을 환기하는 주제에서는 더 강한 정서 반응을 보였다. 이는 극단적 표현을 쉽게 사용하지 않으려는 'never say never'식의 문화적 신념과 특정 주제에 대해 불만족, 불안, 확신과 같은 감정을 얼마나 강하게 또는 직접적으로 표현하는 것이 적절한지에 대한 정서 규범이 ERS 형성에 영향을 줄 수 있음을 시사한다. 따라서 이 연구는 ERS의 문화 차가 단순히 집합주의 - 개인주의와 같은 거시적 문화 차원만으로 설명되기 어렵고, 문항 의미 해석, 응답 전략, 정서 표현 규범이 함께 작동한 결과로 이해되어야 함을 보여준다.

최근 Bonjeer와 Vonkova(2024)는 PISA 2006에 참여한 33개국의 15세 청소년 자료로 응답 스타일과 Hofstede와 GLOBE 연구(House et al., 2004)[15]의 문화 차원 간 단순 및 편상관을 분석하였다. 분석 결과, Hofstede와 GLOBE의 두 문화 프레임에서의 문화 차원과 응답 스타일 간 관계는 성인 대상의 기존 연구 결과와 유사하지만, 일부 지표에서는 차이가 있었다. 예를 들어, 1인당 국내총생산(GDP per capita)을 통제한 후 Hofstede의 권력거리는 ARS 및 DARS와 유의한 관계를 보이지 않았고, Hofstede의 남성성은 ERS와 유의한 관계를 보였다. 이에 연구자들은 문화적 가치가 학생들의 설문 응답 행동에도 영향을 미친다는 점과 국제 비교 연구에서 응답 스타일을 고려해야 할 필요성을 강조하였다.

14) 그러나 이 연구는 Hofstede의 모든 문화 차원(불확실성 회피, 남성성, 권력거리 등)을 포함한 것은 아니었다. 그리고 동일군의 국가들이 다른 문화 차원들에서는 서로 다를 수 있기에 순전히 '개인주의' 효과라고 보기 어려운 측면이 있다.

15) GLOBE의 문화 차원은 권력거리, 불확실성 회피, 제도적 집합주의(institutional collectivism), 내집단 집합주의(in-group collectivism), 성 평등주의, 주장성, 미래 지향성, 성과 지향성, 인간 지향성의 9가지이다.

응답 스타일에 대한 국가 및 문화 수준 요인의 영향

대체로 인구통계적 요인과 성격 요인은 응답 스타일 변량의 일부만을 설명하는 반면, 비교문화 연구에서는 국가 및 문화 수준 특성이 응답 스타일의 국가 간 차이를 상당 부분 설명하는 것으로 보고되어 왔다. 예를 들어 벨기에 표본을 사용한 Weijters 등(2010a)은 성별, 연령, 교육 수준 등 인구통계적 변인이 응답 스타일의 개인차를 일부 설명하지만, 그 설명량은 응답 스타일 유형에 따라 1.4%에서 8.3% 정도에 그친다고 보고하였다. Meisenberg와 Williams(2008) 역시 개인 수준에서는 연령, 교육, 소득과 같은 인구통계적 요인이 ARS와 ERS와 관련되지만, 그 설명력은 제한적이라고 보았다. 반면 국가 수준 분석에서는 국가 평균 IQ, 부패 수준, 경제발전 수준과 같은 국가 특성이 ARS와 ERS의 국가 간 차이를 상당 부분 설명하는 것으로 나타났다. 이와 유사하게 De Jong 등(2008)은 Hofstede의 문화 차원만으로도 ERS의 국가 간 변량 중 상당한 부분을 설명할 수 있다고 보고하였다.

그러나 이러한 결과를 문화의 고유한 효과로 곧바로 해석하는 데에는 주의가 필요하다. 국가 수준에서 측정된 문화 지표는 성격, 경제발전, 교육 수준, 제도적 특성, 언어 및 조사 관행과 밀접하게 얽혀 있을 수 있기 때문이다. 실제로 Hofstede와 McCrae(2004)는 33개국의 국가 평균 성격 점수와 Hofstede의 문화 차원 점수를 비교하여, 여러 문화 차원이 성격 5요인의 국가 평균과 유의하고 상당한 상관을 보인다는 점을 보고하였다. 이는 응답 스타일을 예측할 때 문화 지표의 높은 설명력이 순수하게 문화 고유의 효과를 반영하는지, 아니면 국가 수준 성격, 사회경제적 조건, 교육 수준 등과 공유되는 변량을 포함하는지 신중하게 검토해야 함을 시사한다.

따라서 응답 스타일의 국가 간 차이를 설명할 때는 문화 차원만을 독립적인 원인으로 간주하기보다, 개인 수준의 인구사회학적 특성과 국가 수준의 문화 및 사회경제적 특성을 함께 고려할 필요가 있다. Van Vaerenbergh와 Thomas(2013)가 지적하듯이, 연구 표본 간 인구통계적 구성이나 사회적 맥락이 다를 경우 이러한 변인을 통제하지 않으면 응답 스타일 비교에서 체계적 오류와 해석 왜곡이 발생할 수 있다. 결국 응답 스타일은 개인의 응답 습관이나 성격 특성만으로 설명되기 어렵고, 개인이 속한 국가적·문화적·사회경제적 맥락 속에서 형성되는 다층적 현상으로 이해하는 것이 적절하다.

응답 스타일 조정 관련 논쟁

최근 Ulitzsch 등(2024)은 Sorokowski 등(2023)이 45개국을 대상으로 수행한 사랑 경험(love experience)에 대한 국제 비교 연구를 재분석하면서, 비교문화 연구에서 응답 스타일이 국가 간 비교 결과를 크게 왜곡할 수 있음을 보여주었다. 이들은 서로 다른 국가의 응답자들이 리커트 척도를 동일한 방식으로 사용하지 않으며, 동일한 잠재 태도 수준을 가지고 있더라도 국가별 ERS와 MRS 경향에 따라 서로 다른 응답 범주를 선택할 수 있다고 보았다. 예컨대 9점 척도로 측정된 사랑 경험 관련 문항에서 실제 태도 분포가 동일하다고 가정하더라도, 각국의 응답 척도 사용 방식 차이로 인해 응답 분포와 국가 평균은 전혀 다른 양상으로 나타날 수 있었다. 이러한 차이는 국가 간 평균 비교뿐 아니라, 국가 수준 변인 간의

관계를 해석하는 데에도 영향을 미칠 수 있다. 그들은 Sorokowski 등(2023)이 보고한 국가의 발전 수준과 사랑 경험 간의 곡선형 관계를 재분석하였다. 응답 스타일을 조정하지 않은 모형에서는 기존 연구와 유사하게 유의한 곡선형 관계가 나타났지만, 응답 스타일을 통계적으로 조정한 후에는 인간개발지수(HDI)와 사랑 경험 간의 관계는 더 이상 유의하지 않았다. 이는 비교문화 연구에서 관찰되는 국가 간 차이나 국가 수준 변인 간 관계가 실제 문화적, 심리적 차이를 반영하기보다, 국가별 응답 척도 사용 방식의 차이에서 비롯될 수 있음을 보여준다. 따라서 이들은 ERS, MRS와 같은 응답 스타일을 고려하지 않으면 허구적 문화 차이나 허구적 국가 간 차이가 도출될 수 있다고 주장하였다.

이에 대해 Sorokowski와 Kowal(2024)은 응답 스타일이 비교문화 자료의 해석을 어렵게 만들 수 있다는 점에는 동의하면서도, Ulitzsch 등이 기존 연구 결과를 단순히 잘못된 결론으로 간주하는 것은 성급하다고 보았다. 이들은 비교문화 연구에서 우선 검토되어야 할 문제는 측정 불변성(measurement invariance)이며, 자신들이 사용한 사랑의 삼각형 척도(TLS-45)가 여러 국가에서 측정 불변성을 충족했다고 주장하였다. 또한 응답 스타일 조정은 통계적으로 복잡하고 고도의 모형화를 요구하므로, 모든 국제 비교 연구에서 반드시 적용해야 하는 절대적 기준으로 삼기는 어렵다고 보았다. 이들의 관점에서 응답 스타일은 중요한 방법론적 문제이지만, 응답 스타일 조정 없이도 측정 불변성 검증, 척도 타당화, 자료 품질 관리 등을 통해 비교문화 자료의 신뢰성을 일정 부분 확보할 수 있었다.

Ulitzsch 등(2024)과 Sorokowski와 Kowal(2024)의 논쟁은 응답 스타일 조정 여부가 대규모 비교문화 연구의 결론을 근본적으로 바꿀 수 있음을 보여줌과 동시에 비교 문화 연구에서 응답 스타일을 어떻게 다루어야 하는지에는 여전히 이견이 존재함을 드러낸다. 한쪽에서는 응답 스타일을 통제하지 않으면 허구적 문화 차이나 국가 수준 관계가 도출될 수 있다고 강조하는 반면, 다른 쪽에서는 응답 스타일 조정이 항상 필수적인 것은 아니며 기존의 측정 불변성 검증과 타당화 절차도 중요하다고 본다. 따라서 향후 국제 비교 연구에서는 응답 스타일을 무조건 제거해야 할 잡음으로만 보거나 반대로 별다른 조정 없이 무시하기보다는, 측정 불변성 검증과 응답 스타일 조정 절차를 함께 고려하고 분석 결과의 강건성을 재차 확인할 필요가 있다.

측정 불변성 논쟁

국가 간 비교 연구에서 측정 불변성(invariance) 검증은 서로 다른 문화 집단이 동일한 심리적 구성개념을 같은 방식으로 이해하고 응답하는지를 평가하는 핵심 절차이다. 조사를 통해 얻은 평균, 변량, 공변량이 응답 스타일에 의해 편향된다면(Baumgartner & Steenkamp, 2001), 전통적인 측정의 등가성 검증은 응답 스타일에 대해 조정되어야 할 것이다. 실제로 Welkenhuysen-Gybels 등(2003)과 Kankaraš와 Moors(2011)는 ARS와 ERS에 대해 각각 조정이 이루어졌을 때 측정 등가성 검증의 결과는 상당히 달라질 수 있음을 보여주었다.

그러나 Welzel 등(2023)과 Fischer 등(2023)의 최근 논쟁은, 불변성 위반이 반드시 의미적 비등가성(semantic inequivalence)을 의미하는지, 그리고 응답 스타일과 사회적 바람직성 응답

(SDR)이 이러한 구조적 차이를 어떻게 생성하거나 왜곡하는지를 둘러싸고 근본적 질문을 다시 제기한다.

먼저, Welzel 등(2023)의 관점은 불변성 위반은 '수학적 산물'이며 편향 문제는 과장되었다는 것이다. 그들은 불변성 위반의 주요 원인이 문화적 응답 방식의 차이나 SDR이 아니라, 집단 평균값의 극단성에서 비롯된 닫힌 척도의 수학적 제약이라고 주장한다. 평균이 바닥 또는 천장에 가까울수록 표준편차와 공분산이 기계적으로 축소되며, 이는 요인부하량 감소와 모형 불일치로 이어진다. 이러한 과정은 응답자의 의미적 이해나 응답 전략의 차이와 무관하게 발생하며, 따라서 비불변성이 '편향'의 증거로 해석되지 않아야 한다고 본다. 특히 도덕이나 가족 규범과 같이 특정 사회에서 강한 문화적 합의(consensus)가 존재하는 문항일수록, 응답의 동질성은 응답 편향이 아니라 집단 규범 자체의 강도로 이해해야 한다고 주장한다. 이러한 논리에 따라 Welzel 등은 응답 편향의 영향력은 상대적으로 축소하고 관련 사회적, 정치적 변인들과의 이론적 상관 구조가 유지되는지를 나타내는 nomological validity가 비교 연구의 타당성 판단에서 핵심 기준이 되어야 한다고 제안하였다.

이에 대해 Fischer 등(2023)은 응답 스타일과 SDR은 실제로 불변성 위반을 유발하는 실제 요인이라고 반박한다. 그들은 Welzel의 관점이 과도하게 단순화되어 있으며, 국제 비교 자료에서 관찰되는 구조적 차이를 설명하기에는 불충분하다고 지적하였다. 무엇보다 다양한 문화권에서 ARS, ERS, MRS 등 응답 스타일이 다르게 나타난다는 상당한 증거는 문항 사용 방식의 문화적 차이가 실제로 요인 구조의 변형을 초래할 수 있음을 시사한다(SDR 역시 권위주의적, 전통주의적 맥락에서 특정 응답을 선호하도록 압력을 가하여 집단 평균뿐 아니라 항목 간 공분산 구조까지 왜곡할 수 있다). Fischer 등은 Welzel 등이 강조하는 nomological validity는 측정 편향(measurement bias)을 무효화하는 것이 아니라 편향이 존재하더라도 이론적 상관은 유지될 수 있음을 보여주는 사례일 뿐이라고 반박하였다. 즉 이론적 관계가 타당하다고 해서 측정 과정이 문제없다는 결론을 내릴 수는 없으며, 편향을 통제하지 않은 상태에서의 강한 상관은 오히려 잘못된 확신을 강화할 위험이 있다는 것이다.

이 논쟁의 핵심은 국가 간 관찰된 평균 차이와 요인 구조 차이가 어느 정도까지 문화적 현실(cultural reality)을 반영하며, 어느 정도까지 응답 과정의 결과인지의 문제로 수렴된다. Welzel 등의 접근은 평균 차이를 존재론적 실제로 보고, 편향보다는 문화적 합의에 초점을 두고 해석한다. 반대로 Fischer 등의 접근은 응답 스타일 및 SDR의 누적된 영향으로 평균 차이가 체계적으로 왜곡될 가능성을 강조한다. 이 논쟁이 부각하는 점은 응답 편향이 측정 불변성 논쟁의 부차적 요소가 아니며, 응답 스타일과 SDR은 측정 구조와 국가 간 비교의 타당성을 직접적으로 형성하는 핵심 요인이라는 것이다. Welzel 등이 지적한 수학적 제약이 존재하더라도, 실제 자료에서는 응답 스타일과 SDR이 그 효과를 증폭하거나 다른 방향으로 왜곡시킬 수 있다. 반대로 Fischer 등의 접근이 옳다고 하더라도, 모든 비불변성이 응답 편향 때문이라고 단정할 수도 없다. 이는 국제 비교 연구에서 (1) 수학적 제약, (2) 응답 과정의 문화적 차이, (3) nomological 구조를 종합적으로 고려하는 다층적 타당성 평가가 필요함을 시사한다.

문화 간 비교가능성을 높이기 위한 방법론적 절차

응답 스타일은 비교 문화 연구에서 국가 간 평균 차이와 변인 간 관계를 왜곡할 수 있으므로, 연구자는 연구 설계와 분석 단계에서 이를 체계적으로 고려할 필요가 있다. 다만 응답 스타일을 단순히 제거해야 할 잡음으로만 보는 접근은 충분하지 않다. 응답 스타일은 측정오차의 원천이면서 동시에 문화적으로 학습된 의사표현 방식이나 응답 규범을 반영할 수 있기 때문이다. 따라서 비교 문화 연구에서는 응답 스타일을 무조건 제거하기보다, 그것을 어떻게 측정하고, 모형화하며, 결과 해석에 반영할 것인지에 대한 방법론적 판단이 필요하다.

첫째, 응답 스타일을 직접 측정하여 분석 모형에 포함하는 방법을 고려할 수 있다. 예를 들어 Lee 등(2022)은 응답자의 ARS를 간단히 선별하기 위한 단축형 screener를 개발하였다. 이 도구는 장문의 ARS 지표를 사용하기 어려운 조사 현장에서 응답자의 전반적 동의 경향을 간결하게 파악할 수 있게 해준다는 점에서 유용하다. 다만 이 척도는 주로 미국 성인을 대상으로 개발된 것이므로, 비교 문화 연구에 사용할 경우에는 각 문화권에서 동일한 방식으로 작동하는지 별도의 타당화가 필요하다. 이와 함께 He와 van de Vijver(2019)의 일반화된 응답 스타일(generalized response style: GRS) 척도도 응답 스타일을 직접 측정하려는 접근의 예로 볼 수 있다. GRS는 ARS, ERS, MRS, SDR과 같은 개별 응답 스타일을 하나의 넓은 의사소통 스타일로 이해하며, 응답자가 자신의 의견이나 감정을 증폭하여 표현하는지 또는 절제하여 표현하는지를 포착하려는 개념이다. 이러한 직접 측정 접근은 응답 스타일을 본 설문 점수에서 사후적으로 추정하는 방법과 달리, 응답 스타일 자체를 공변량이나 잠재 요인으로 모형에 포함할 수 있게 해준다는 장점이 있다.

둘째, 개인 내 표준화(within-person standardization 또는 ipsatization)와 같은 자료 변환 절차는 신중하게 사용해야 한다. 개인 내 표준화는 각 응답자의 전체 응답 평균을 기준으로 개별 문항 점수를 재조정하는 방법이다. 예를 들어 어떤 응답자가 대부분의 문항에 높게 응답하는 경향을 보이면, 이 절차는 그 사람의 응답을 개인 평균으로부터의 상대적 위치로 변환하여 전반적 동의 경향이나 높은 점수 사용 경향을 줄이고자 한다. 그러나 이 방식은 응답자의 절대적 점수를 상대적 프로파일로 바꾸기 때문에, 실제로 존재하는 집단 간 평균 차이나 문화 차이까지 함께 제거하거나 왜곡할 수 있다. Venaik 등(2021)은 Gelfand 등(2011)의 문화적 경직성 - 유연성(tightness - looseness) 자료를 재분석하여, 개인 내 표준화가 ARS를 제거한다는 명분 아래 관심 변인의 실제 범위를 과도하게 축소하거나 확대하고, 국가 수준 변인과의 관계를 왜곡할 수 있음을 보여주었다. 따라서 ipsatization은 편리한 보정 절차이지만, 비교 문화 연구에서 기계적으로 적용하기보다는 원자료, 보정 자료, 대안적 지표를 함께 비교하여 결과의 강건성을 검토하는 방식으로 사용해야 한다.

셋째, ARS를 조정하기 위해서는 반대 방향 문항(opposite-items) 접근법을 활용할 수 있다. 이 접근법은 의미상 서로 반대되는 문항 쌍을 함께 사용하여, 실제 태도와 무관하게 모든 문항에 동의하거나 부동의하는 경향을 추정하려는 방법이다. 예를 들어 동일한 구성개념을

긍정 방향과 부정 방향에서 각각 묻는 문항을 포함하면, 응답자가 내용과 관계없이 동의 방향으로 응답하는지 확인할 수 있다. Venaik 등(2021)은 이러한 접근이 단순한 개인 내 표준화보다 ARS를 더 직접적으로 식별하고 조정하는 데 적합할 수 있다고 보았다. 다만 반대 방향 문항을 사용할 때도 번역, 문항 난이도, 부정문 처리 부담, 문화권별 문항 해석 차이를 함께 고려해야 한다. 반대 방향 문항이 항상 동일한 의미의 반대 극을 측정하는 것은 아니기 때문이다.

넷째, 특정 문화 차원을 측정할 때는 단일 지표에만 의존하지 않고 대안적 지표를 함께 검토할 필요가 있다. 예를 들어 문화적 경직성 - 유연성의 경우, Gelfand 등(2011)의 주관적 평정 기반 지수 외에도 Uz(2015)의 대안적 지수를 활용할 수 있다. Uz의 지수는 개인의 평균 평정을 단순히 국가 수준으로 집계하기보다, 한 사회 내부에서 가치, 규범, 행동이 얼마나 동질적 또는 이질적으로 분포하는지를 나타내는 문화 수준의 표준편차에 근거하여 경직성 - 유연성을 추정한다. 이처럼 대안적 지표를 함께 비교하면, 특정 문화 차원의 결과가 실제 문화 차이를 반영하는지, 아니면 특정 척도 형식이나 응답 스타일에 의존한 것인지 더 신중하게 판단할 수 있다.

다섯째, 여러 비교가능성 개선 절차를 함께 평가하는 접근이 필요하다. He 등(2017)은 16개국 대학생 자료를 사용하여 성격과 가치 척도의 비교문화적 비교가능성을 높이기 위한 여러 절차를 비교하였다. 이들은 사전적 절차로 앵커링 비네트와 응답 스타일 및 SDR의 직접 측정을, 사후적 절차로 ipsatization과 문항 묶음(item parcelling)을, 자료 모형화 절차로 리커트 자료를 연속형으로 처리하는 방식과 순서형 범주로 처리하는 방식을 검토하였다. 분석 결과, 앵커링 비네트는 측정 불변성을 충족하지 못해 편향이 없는 절차라고 보기 어려웠고, ipsatization은 내적 합치도와 비교가능성을 낮추는 경향을 보였다. 또한 어떤 절차도 scalar 불변성을 확보하지는 못했다. 반면 앵커링 비네트와 문항 묶음은 비교가능성을 일부 개선했으며, 리커트 척도 자료를 연속형이 아니라 순서형 범주로 처리할 때 비교 가능성이 더 높게 나타났다. 응답 스타일 및 SDR 보정은 전반적인 비교 가능성에는 큰 영향을 미치지 않았지만, 국가 수준 성실성의 외적 타당도를 높이는 데에는 도움이 되는 것으로 나타났다.

이러한 결과는 비교문화 연구에서 응답 스타일 조정 절차가 만능 해결책이 아님을 보여준다. 어떤 절차도 비교 가능성 문제를 완전히 해결하지 못하며, 절차에 따라 신뢰도, 측정 불변성, 외적 타당도, 국가 평균 해석에 미치는 영향이 달라질 수 있다. 따라서 연구자는 특정 방법을 기계적으로 적용하기보다, 연구 목적과 자료 구조에 맞게 여러 절차를 비교하고, 원점수와 보정 점수의 결과가 어떻게 달라지는지 투명하게 보고해야 한다. 특히 응답 스타일 조정은 때로 실제 문화 차이를 제거하거나 왜곡할 수 있으므로, 조정 결과를 곧바로 더 '정확한' 결과로 간주하기보다는 민감도 분석의 일부로 해석하는 것이 바람직하다.

나가는 말

본 고에서는 자기 보고식 사회조사와 심리 측정에서 나타나는 응답 스타일을 중심으로,

주요 유형과 발생 원인, 측정 결과에 미치는 영향, 탐지와 조정 방법, 예방 전략, 그리고 비교문화 연구에서의 문제를 살펴보았다. 묵인 응답 스타일(ARS), 비묵인 응답 스타일(DARS), 극단 응답 스타일(ERS), 중간 응답 스타일(MRS)은 문항의 실제 내용과 관계없이 특정한 응답 범주를 반복적으로 선택하는 경향이다. 이러한 응답 스타일은 단순한 무선 오차가 아니다. 관찰된 점수의 평균과 변량을 바꾸고, 변인 간 상관과 회귀계수, 요인구조 및 집단 간 차이를 왜곡할 수 있는 체계적 측정오차이다(Baumgartner & Steenkamp, 2001; Wetzel, Böhnke, & Brown, 2016).

응답 스타일의 영향은 반복측정 자료에서 더욱 주의 깊게 살펴볼 필요가 있다. 경험표집법(ESM)이나 생태학적 순간 평가(EMA)에서는 같은 응답자에게 여러 차례 자기 보고를 요구한다. 이 경우 개인이 평정 척도를 사용하는 습관이 반복적으로 자료에 반영될 수 있다. 예를 들어 ERS가 높은 응답자는 실제 정서 변화가 크지 않더라도 극단 범주를 자주 사용하기 때문에 정서 점수의 변동성이 크게 나타날 수 있다. 따라서 반복측정 자료에서 높은 변동성이 관찰되었다고 해서 이를 곧바로 실제 심리 상태의 불안정성으로 해석해서는 안 된다(Henninger, Vanhasbroeck, & Tuerlinckx, 2025).

그러나 응답 스타일을 언제나 제거해야 할 잡음으로만 보는 것도 적절하지 않다. 본문에서 살펴보았듯이 응답 스타일은 응답 범주의 수와 라벨, 자료 수집 방식, 문항의 난이도와 모호성, 주제에 대한 관심, 조사 언어와 같은 측정 상황의 영향을 받는다. 동시에 교육 수준, 인지 능력, 성격 및 자기조절 특성과 같은 응답자의 개인차와도 관련된다. 일부 응답 스타일은 여러 문항과 시점에 걸쳐 비교적 안정적으로 나타나지만, 실제로 어느 정도 나타나는지는 문항과 조사 상황에 따라 달라질 수 있다. 따라서 응답 스타일은 응답자에게 고정된 성향이나 설문 형식만으로 설명하기보다, 응답자의 특성과 측정 상황이 함께 작용하여 나타나는 응답 행동으로 이해하는 것이 적절하다.

비교문화 연구에서는 이러한 점이 더욱 중요하다. 응답 스타일은 국가 간 평균 차이나 변인 간 관계를 왜곡하는 측정오차가 될 수 있다. 그러나 동시에 특정 문화에서 의견과 감정을 표현하는 방식, 권위에 반응하는 방식, 조화와 체면을 중시하는 정도와 같은 문화적 의사소통 규범을 반영할 수도 있다. 어떤 문화에서는 자신의 의견을 강하게 표현하는 것이 자연스럽지만, 다른 문화에서는 중간적이고 완화된 표현이 더 적절하게 받아들여질 수 있다. 따라서 국가 간 응답 스타일의 차이를 단순한 개인 습관이나 측정오차로만 해석해서는 안 된다. 개인 수준의 특성뿐 아니라 조사 언어, 사회경제적 조건, 문화적 가치와 의사소통 규범을 함께 고려해야 한다.

응답 스타일을 탐지하기 위해서는 여러 방법이 제안되어 왔다. 가장 간단한 방법은 동의, 비동의, 극단 또는 중간 응답의 빈도를 계산하는 것이다. 이 밖에도 역문항과 균형 문항, MTMM 및 CFA 방법요인 모형, RIRS와 RIRMACS, 잠재계층 요인분석, 다차원 IRT, 혼합 IRT 및 IRTree 모형을 사용할 수 있다. 그러나 이들 방법은 각각 서로 다른 가정과 장단점을 갖는다. 빈도 지표는 계산하기 쉽지만 문항 내용과 응답 스타일을 분명하게 구분하기 어려울 수 있다. 역문항과 균형 문항은 ARS를 탐지하는 데 도움이 되지만, 부정문 처

리의 어려움이나 문항 의미의 차이로 인해 별도의 방법효과가 나타날 수 있다. 잠재변인이나 IRT 기반 모형은 관심 특성과 응답 스타일을 함께 추정할 수 있지만, 충분한 문항과 표본이 필요하고 모형의 식별과 제약 방식에 따라 결과가 달라질 수 있다.

같은 자료를 사용하더라도 응답 스타일을 어떻게 정의하고 모형화하는가에 따라 결과가 달라질 수 있다. 예를 들어 응답 스타일을 연속적인 개인차로 볼 수도 있고, 서로 다른 응답 유형을 가진 잠재계층으로 볼 수도 있으며, 일련의 하위 의사결정 과정으로 볼 수도 있다. 이러한 선택에 따라 추정되는 특성 점수와 표준오차가 달라질 수 있다(Bolt & Meng, 2025). 따라서 하나의 방법을 모든 연구에 적용할 수 있는 최선의 방법으로 보기보다는, 연구 목적, 척도 구조, 문항 수, 표본 크기와 모형 가정을 고려하여 적절한 방법을 선택해야 한다.

응답 스타일에 대한 사후적 조정도 신중하게 사용해야 한다. 회귀 잔차화나 개인 내 표준화는 적용하기 쉽지만, 응답 스타일과 함께 실제로 존재하는 개인차나 집단 차이까지 제거할 수 있다. 잠재변인이나 IRT에 기초한 조정도 선택한 모형과 가정에 따라 결과가 달라질 수 있다. 따라서 조정된 점수를 곧바로 편향이 제거된 참점수로 간주해서는 안 된다. 실제 연구에서는 원자료를 사용한 결과와 조정된 자료를 사용한 결과를 함께 제시하고, 서로 다른 조정 방법을 적용했을 때 주요 결론이 얼마나 달라지는지를 확인하는 것이 바람직하다. 즉, 응답 스타일 조정은 모든 연구에서 반드시 적용해야 하는 기계적 절차라기보다 연구 결과의 강건성을 확인하기 위한 민감도 분석의 하나로 사용하는 것이 적절하다.

비교문화 연구에서는 측정 불변성 검증과 응답 스타일 조정을 함께 고려해야 한다. 측정 불변성이 확보되었다고 해서 응답 스타일의 영향이 자동적으로 사라지는 것은 아니다. 반대로 측정 불변성이 충족되지 않았다고 해서 그 원인을 모두 응답 스타일로 돌릴 수도 없다. 척도의 바닥효과와 천장효과, 번역과 문항 의미의 차이, 문화별 척도 사용 방식, 사회적 바람직성 응답, 실제 문화적 합의가 모두 측정 불변성 결과에 영향을 줄 수 있기 때문이다(Fischer et al., 2023; Welzel et al., 2023). 따라서 국가 간 비교에서는 측정 불변성, 응답 스타일, 문항 기능, 외부 변인과의 관계를 함께 검토할 필요가 있다.

응답 스타일에 대처하는 가장 현실적인 방법은 사후 조정에만 의존하지 않고 설문과 척도를 설계하는 단계에서 문제를 줄이는 것이다. 문항은 간결하고 명확하게 작성하고, 가능하면 일반적인 동의 - 비동의 형식보다 문항의 내용에 직접 맞는 응답 범주를 사용하는 것이 좋다. 응답 범주에는 분명한 언어적 라벨을 제공하고, 응답자가 자신의 판단을 충분히 표현할 수 있도록 적절한 수의 범주를 마련해야 한다. 중간 응답을 줄이기 위해 중간 범주를 무조건 삭제하는 것은 바람직하지 않다. 실제로 중간적인 태도를 가진 응답자의 답을 왜곡할 수 있기 때문이다.

긍정 문항과 부정 문항을 함께 사용하는 방법도 신중하게 적용해야 한다. 이 방법은 ARS를 자동으로 없애는 방법이라기보다, 내용 요인과 묵인 응답 요인을 구분하기 위한 정보를 제공하는 설계 방법이다. 그러나 부정 문항은 정방향 문항보다 이해하기 어려울 수 있고, 두 문항이 실제로 같은 구성개념의 반대 방향을 측정하지 않을 수도 있다. 따라서 역문항이나 반대 방향 문항을 사용할 때는 예비조사

와 인지 면접을 통해 응답자가 문항을 어떻게 이해하는지 확인해야 한다. 또한 요인분석과 IRT 기반 문항 분석을 통해 문항과 응답 범주가 의도한 방식으로 기능하는지 점검할 필요가 있다.

최근에는 응답시간(response time: RT)을 이용하여 응답 스타일을 응답 과정 수준에서 이해하려는 연구도 증가하고 있다. Henninger와 Plieninger(2021)는 응답자가 자신이 평소 선호하는 응답 범주를 선택할 때 응답시간이 짧아지는 경향을 보여주었다. 이는 응답 스타일이 단순히 문항을 충분히 생각하지 않은 결과만은 아님을 시사한다. 특정 범주가 어떤 응답자에게 익숙하고 쉽게 선택할 수 있는 응답 전략이기 때문에 더 빠르게 선택될 수도 있다.

Lyu와 Bolt(2024)는 응답시간을 설명할 때 관심 특성과 응답 스타일을 함께 고려해야 한다고 보았다. 선택한 범주가 응답자의 특성 수준에서 볼 때 그럴듯할수록 응답시간이 짧았으며, 동시에 그 응답자가 평소 자주 사용하는 범주일수록 응답시간도 짧았다. 두 과정 중 하나만 고려하면 다른 과정의 효과가 잘못 추정될 수 있었다. Kim과 Bolt(2024)도 빠른 응답과 느린 응답에서 문항이 관심 특성과 응답 스타일을 반영하는 정도가 다를 수 있음을 보여주었다. 특히 빠른 응답에서는 응답 스타일과 관련된 변별력이 상대적으로 더 크게 나타났다. 따라서 응답시간은 부주의한 응답을 선별하기 위한 단순한 지표가 아니라, 관심 특성과 응답 스타일, 문항 처리 과정이 함께 반영된 과정 자료로 이해할 필요가 있다.

앞으로의 연구에서는 몇 가지 과제가 중요하다. 첫째, 서로 다른 탐지 방법이 실제로 같은 응답 스타일을 측정하는지 확인할 필요가 있다. 둘째, 응답 스타일을 연속적인 특성, 잠재계층 또는 하위 의사결정 과정으로 보는 여러 모형을 체계적으로 비교해야 한다. 셋째, 문항 반응과 응답시간을 함께 분석하여 실제 특성에 따른 응답과 응답 범주 사용 습관을 보다 정확하게 구분할 필요가 있다. 넷째, 반복측정 자료와 비교문화 자료에서는 개인 수준과 상황 및 국가 수준의 요인을 함께 고려하는 다층적 연구가 필요하다. 마지막으로 응답 스타일을 조정한 점수가 외부 행동, 실제 수행 및 의사결정의 예측을 실제로 향상시키는지 검증해야 한다.

궁극적으로 타당한 자기 보고식 측정은 응답 스타일을 통계적으로 제거한 하나의 점수를 얻는 것만으로 이루어지지 않는다. 응답자가 문항을 어떻게 읽고 이해하는지, 어떤 정보를 떠올리는지, 어떤 판단을 내리고 이를 응답 범주에 어떻게 연결하는지를 함께 이해해야 한다. 따라서 응답 스타일에 대한 대처는 명확한 문항과 응답 범주의 설계, 적절한 자료 수집, 응답 과정에 대한 측정, 이론에 근거한 통계 모형, 그리고 조정 전후 결과의 비교를 포함하는 통합적인 과정이어야 한다. 이러한 관점에서 응답 스타일은 단순히 제거해야 할 오차가 아니라, 자기 보고식 측정이 어떤 조건에서 타당하게 이루어지는지를 이해하기 위한 중요한 연구 대상이다.

참고문헌

Adams, D. J., Bolt, D. M., Deng, S., Smith, S. S., & Baker, T. B. (2019). Using multidimensional item response theory to evaluate how response styles impact measurement. *British Journal of Mathematical*

and Statistical Psychology, *72*(3), 466-485. https://doi.org/10.1111/bmsp.12169

Austin, E. J., Deary, I. J., & Egan, V. (2006). Individual differences in response scale use: Mixed Rasch modelling of responses to NEO-FFI items. *Personality and Individual Differences*, *40*, 1235-1245. https://doi.org/10.1016/j.paid.2005.10.018

Ayidiya, S., & McClendon, M. J. (1990). Response effects in mail surveys. *Public Opinion Quarterly*, *54*, 229-247. https://doi.org/10.1086/269200

Baumgartner, H., & Steenkamp, J.-B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, *38*(2), 143-156. https://doi.org/10.1509/jmkr.38.2.143.18840

Benítez, I., He, J., Van de Vijver, F.J.R. and Padilla, J.-L. (2016). Linking extreme response style to response processes: A cross-cultural mixed methods approach. *International Journal of Psychology*, *51*, 464-473. https://doi.org/10.1002/ijop.12379

Bentler, P. M., Jackson, D. N., & Messick, S. (1971). Identification of content and style: A two-dimensional interpretation of acquiescence. *Psychological Bulletin*, *76*, 186-204. https://doi.org/10.1037/h0031474

Billiet, J. B., & Davidov, E. (2008). Testing the stability of an acquiescence style factor behind two interrelated substantive variables in a panel design. *Sociological Methods and Research*, *36*, 54-562. https://doi.org/10.1177/0049124107313901

Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling*, *7*, 608-628. https://doi.org/10.1207/S15328007SEM0704_5

Bishop, G. F. (1987). Experiments with the middle response alternative in survey questions. *Public Opinion Quarterly*, *51*(2), 220-232. ttps://doi.org/10.1086/269030

Bishop, G. F., Oldendick, R. W., Tuchfarber, A. J., & Bennett, S. E. (1980). Pseudo-opinions on public affairs. *Public Opinion Quarterly*, *44*(2), 198-209. https://doi.org/10.1086/268584

Böckenholt, U. (2012). Modeling multiple response processes in judgment and choice. *Psychological Methods, 17*(4), 665-678. https://doi.org/10.1037/a0028111

Böckenholt, U. (2017). Measuring response styles in Likert items. *Psychological Methods*, *22*(1), 69-83. https://doi.org/10.1037/met0000106

Böckenholt, U., & Meiser, T. (2017). Response style analysis with threshold and multi-process IRT models: A review and tutorial. British Journal of Mathematical and Statistical Psychology, 70(1), 159-181. https://doi.org/10.1111/bmsp.12086

Bolt, D. M., & Johnson, T. R. (2009). Addressing score bias and differential item functioning due to individual differences in response style. *Applied Psychological Measurement, 33*(5), 335-352. https://doi.org/10.1177/0146621608329891

Bolt, D. M., Lu, Y., & Kim, J. S. (2014). Measurement and control of response styles using anchoring vignettes: A model-based approach. *Psychological Methods, 19*(4), 528-541. https://doi.org/10.1037/met0000016

Bolt, D. M., & Newton, J. R. (2011). Multiscale measurement of extreme response style. *Educational and Psychological Measurement, 71*(5), 814-833. https://doi.org/10.1177/0013164410388411

Bolt, D. M., & Meng, L. (2025). IRT-based response style models and related methodology: Review and commentary. *British Journal of Mathematical and Statistical Psychology*. Advance online publication. https://doi.org/10.1111/bmsp.70006

Bonjeer, T., & Vonkova, H. (2024). Relationships between response styles and the Hofstede and GLOBE dimensions of culture in a sample of adolescents from 33 countries. *Cross-Cultural Research*, *58*(2-3), 180-207. https://doi.org/10.1177/10693971231203759

Cabooter, E. F. K. (2010). *The impact of situational and dispositional variables on response styles with respect to attitude measures*. Doctoral dissertation, Ghent University, Ghent, Belgium.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin,* 56(2), 81-105. https://doi.org/10.1037/h0046016

Chen, C., Lee, S. Y., & Stevenson, H. W. (1995). Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science*, *6*(3), 170-175. https://doi.org/10.1111/j.1467-9280.1995.tb00327.x

Clarke, I. (2000). Extreme response style in cross-cultural research: An empirical investigation. *Journal of Social Behavior & Personality*, *15*(1), 137-152.

Couch, A., & Keniston, K. (1960). Yea-sayers and nay-sayers: Agreeing response set as a personality variable. *Journal of Abnormal and Social Psychology*, *60*, 151-172. https://doi.org/10.1037/h0040372

Crandall, J. (1973). Sex differences in extreme response style: Differences in frequency of use of extreme positive and negative ratings. *Journal of Social Psychology*, *89*, 281-293. https://doi.org/10.1080/00224545.1973.9922601

De Boeck, P., & Partchev, I. (2012). IRTrees: Tree-based item response models of the GLMM family. *Journal of Statistical Software*, 48(Code Snippet 1), 1-28. https://doi.org/10.18637/jss.v048.c01

De Jong, M. G., Steenkamp, J. B., Fox, J. P., & Baumgartner, H. (2008). Using item response theory to measure extreme response style in marketing research: A global investigation. *Journal of Marketing Research*, *45*, 104-115. https://doi.org/10.1509/jmkr.45.1.104

Eid, M., & Diener, E. (1999). Intraindividual variability in affect: Reliability, validity, and personality correlates. *Journal of Personality and Social Psychology*, *76*(4), 662-676. https://doi.org/10.1037/0022-3514.76.4.662

Eid, M., & Rauber, M. (2000). Detecting measurement invariance in organizational surveys. *European Journal of Psychological Assessment*, *16*, 20-30. https://doi.org/10.1027//1015-5759.16.1.20

Eid, M., Lischetzke, T., Nussbeck, F. W., & Trierweiler, L. I. (2003). Separating trait

effects from trait-specific method effects in multitrait-multimethod models: A multiple-indicator CT-C(M-1) model. Psychological Methods, 8(1), 38-60. https://doi.org/10.1037/1082-989X.8.1.38

Fischer, R., Karl, J. A., Fontaine, J. R. J., & Poortinga, Y. H. (2023). Evidence of validity does not rule out systematic bias: A commentary on nomological noise and cross-cultural invariance. *Sociological Methods & Research*, *52*(3), 1420-1437. https://doi.org.libproxy.sogang.ac.kr/10.1177/00491241221091756

Fukudome, K., & Takeda, T. (2024). The influence of cognitive reflection on consistency of responses between reversed and direct items. *Personality and Individual Differences*, *30*, 112811. https://doi.org/10.1016/j.paid.2024.112811

Gelfand, M. J., Szabo, E., Fischlmayr, I., & et al. (2011). Differences between tight and loose cultures: A 33-nation study. *Science*, *332*(6033), 1100-1104. https://doi.org/10.1126/science.1197754

Gibbons, J. L., Zelner, J. A., & Reduk, D. J. (1999). Effects of language and meaningfulness on the use of extreme response styles by Spanish-English bilinguals. *Cross-Cultural Research*, *33*, 369-381. https://doi.org/10.1177/106939719903300404

Greenleaf, E. A. (1992a). Measuring extreme response style. *Public Opinion Quarterly*, *56*, 328-350. https://doi.org/10.1086/269326

Greenleaf, E. A. (1992b). Improving rating-scale measures by detecting and correcting bias components in some response styles. *Journal of Marketing Research, 29*(2), 176-188. https://doi.org/10.2307/3172568

Grimm, S. D., & Church, A. T. (1999). A cross-cultural study of response biases in personality measures. *Journal of Research in Personality*, *33*(4), 415-441. https://doi.org/10.1006/jrpe.1999.2256

Hamilton, D. (1968). Personality attributes associated with extreme response style. *Psychological Bulletin*, *69*, 192-203. https://doi.org/10.1037/h0025606

Harzing, A.-W. (2006). Response styles in cross-national survey research: A 26-country study. *International Journal of Cross Cultural Management*, *6*(2), 243-266. https://doi.org/10.1177/1470595806066332

He, J., & Van de Vijver, F. J. R. (2016). Response styles in factual items: Personal, contextual and cultural correlates. *International Journal of Psychology*, *51*, 445-452. https://doi.org/10.1002/ijop.12263

He, J., & Van de Vijver, F. J. R. (2019). Assessment of the general response style: A cross-cultural validation. *Acta de investigación psicológica*, *9*(3), 14-24. https://doi.org/10.22201/fpsi.20074719e.2019.3.320

He, J., Van de Vijver, F. J. R., Fetvadjiev, V. H., Espinosa, A. C. D., Adams, B., Alonso-Arbiol, I., Aydinli Karakulak, A., Buzea, C., Dimitrova, R., Fortin, A., Hapunda, G., Ma, S., Sargautyte, R., Sim, S., Schachner, M. K., Suryani, A., Zeinoun, P., & Zhang, R. (2017). On enhancing the cross cultural comparability

of Likert scale personality and value measures: A comparison of common procedures. *European Journal of Personality*, *31*(6), 642-657. https://doi.org/10.1002/per.2132

Henninger, M., & Plieninger, H. (2021). Different styles, different times: How response times can inform our knowledge about the response process in rating scale measurement. *Assessment*, 28(5), 1301-1319. https://doi.org/10.1177/1073191119900003

Henninger, M., Vanhasbroeck, N., & Tuerlinckx, F. (2025). Affect dynamics or response bias? The relationship between extreme response style and affect dynamics in a controlled experiment. *Psychological Assessment*, *37*(11), 639-655. https://doi.org/10.1037/pas0001370

Hofstede, G., & McCrae, R. R. (2004). Personality and culture revisited: Linking traits and dimensions of culture. *Cross-Cultural Research*, *38*(1), 52-88. https://doi.org/10.1177/1069397103259443

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (Eds.). (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Sage Publications.

Johnson, T., Kulesa, P., Cho, Y. I., & Shavitt, S. (2005). The relation between culture and response styles: Evidence from 19 countries. *Journal of Cross-Cultural Psychology, 36*(2), 264-277. https://doi.org/10.1177/0022022104272905

Kam, C. C. S. (2023). Why do regular and reversed items load on separate factors? Response difficulty vs. item extremity. *Educational and Psychological Measurement*, *83*(6), 1085-1112. https://doi.org/10.1177/00131644221143972

Kam, C. C. S., Meyer, J. P., & Sun, S. (2021). Why do people agree with both regular and reversed items? A logical response perspective. *Assessment*, *28*(4), 1110-1124. https://doi.org/10.1177/10731911211001931

Kankaraš, M., Vermunt, J. K., & Moors, G. (2011). Measurement equivalence of ordinal items: A comparison of factor analytic, item response theory, and latent class approaches. Sociological Methods & Research, 40(2), 279-310. https://doi.org/10.1177/0049124111405301

Kemmelmeier, M. (2016). Cultural differences in survey responding: Issues and insights in the study of response biases. *International Journal of Psychology*, *51*(6), 439-444. https://doi.org/10.1002/ijop.12386

Khorramdel, L., & von Davier, M. (2014). Measuring response styles across the Big Five: A multiscale extension of an approach using multinomial processing trees. *Multivariate Behavioral Research, 49*(2), 161-177. https://doi.org/10.1080/00273171.2013.866536

Kieruj, N. D., & Moors, G. (2010). Variations in response style behavior by response scale format in attitude research. *International Journal of Public Opinion Research, 22*(3), 320-342. https://doi.org/10.1093/ijpor/edq001

Kieruj, N. D., & Moors, G. (2013). Response style behavior: Question format dependent or personal style? *Quality & Quantity*, *47*, 193-211. https://doi.org/10.1007/s11135-011-9511-4

Kiesler, S., & Sproull, L. S. (1986). Response effects in the electronic survey. *Public Opinion Quarterly*, *50*(3), 402-413. https://doi.org/10.1086/268992

Kim, N., & Bolt, D. M. (2024). Evaluating psychometric differences between fast versus slow responses on rating scale items. *Journal of Educational and Behavioral Statistics*, *49*(4), 565-594. https://doi.org/10.3102/10769986231195260

Knowles, E. S., & Condon, C. A. (1999). Why people say ''yes'': A dual-process theory of acquiescence. *Journal of Personality and Social Psychology*, *77*, 379-386. https://doi.org/10.1037/0022-3514.77.2.379

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5*(3), 213-236. https://doi.org/10.1002/acp.2350050305

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, *50*, 537-567. https://doi.org/10.1146/annurev.psych.50.1.537

Krosnick, J. A., Narayan, S., & Smith, W. R. (1996). Satisficing in surveys: Initial evidence. *New Directions for Evaluation*, *70*, 29-44. https://doi.org/10.1002/ev.1033

Kuppens, P., Van Mechelen, I., Nezlek, J. B., Dossche, D., & Timmermans, T. (2007). Individual differences in core affect variability and their relationship to personality and psychological adjustment. *Emotion*, *7*(2), 262-274. https://doi.org/10.1037/1528-3542.7.2.262

Lau, A. L. D., Cummins, R. A., & Mcpherson, W. (2005). An investigation into the cross-cultural equivalence of the Personal Wellbeing Index. *Social Indicator Research*, *72*, 403-430. https://doi.org/10.1007/s11205-004-0561-z

Lee, S., Alvarado-Leiton, F., Yu, W., Davis, R., & Johnson, T. P. (2022). Developing a short screener for acquiescent respondents. *Research in Social and Administrative Pharmacy*, *18*(5), 2817-2829. https://doi.org/10.1016/j.sapharm.2021.06.012

Light, C. S., Zax, M., & Gardiner, D. H. (1965). Relationship of age, sex, and intelligence level to extreme response style. *Journal of Personality and Social Psychology*, *2*, 907-909. https://doi.org/10.1037/h0022746

Lyu, W., & Bolt, D. M. (2024). Predicting response time on self-report rating scale assessments of noncognitive constructs. *Behavior Research Methods*, *56*, 1123-1139. https://doi.org/10.3758/s13428-023-02073-w

Marín, G., Gamba, R. J., & Marín, B. V. (1992). Extreme response style and Acquiescence among Hispanics: The role of acculturation and education. *Journal of Cross-Cultural Psychology*, *23*(4), 498-509.

Maydeu-Olivares, A., & Coffman, D. L. (2006). Random intercept item factor analysis. *Psychological Methods, 11*(4), 344-362. https://doi.org/10.1037/1082-989x.11.4.344

Meisenberg, G., & Williams, A. (2008). Are acquiescent and extreme response styles related to low intelligence and education? *Personality and Individual Differences*, *44*, 1539-1550. https://doi.org/10.1016/j.paid.2008.01.010

Moors, G. (2003). Diagnosing response style behavior by means of a latent-class factor approach: Socio-demographic correlates of gender role attitudes and perceptions of ethnic discrimination reexamined. Quality & Quantity, 37, 277-302. https://doi.org/10.1023/A:1024472110002

Moors, G. (2008). Exploring the effect of a middle response category on response style in attitude measurement. *Quality & Quantity*, *42*, 779-794. https://doi.org/10.1007/s11135-006-9067-x

Moors, G. (2012). The effect of response style bias on the measurement of transformational, transaction, and laissez-faire leadership. *European Journal of Work and Organizational Psychology*, *21*, 271-298. https://doi.org/10.1080/1359432X.2010.550680

Moors, G., Kieruj, N. D., & Vermunt, J. K. (2014). The effect of labeling and numbering of response scales on the likelihood of response bias. *Sociological Methodology, 44*(1), 369-399. https://doi.org/10.1177/0081175013516114

Morren, M., Gelissen, J. P., & Vermunt, J. K. (2011). Dealing with extreme response style in cross-cultural research: A restricted latent class factor analysis approach. *Sociological Methodology*, *41*(1), 13-47. https://doi.org/10.1111/j.1467-9531.2011.01238.x

Naemi, B. D., Beal, D. J., & Payne, S. C. (2009). Personality predictors of extreme response style. *Journal of Personality*, 77(1), 261-286. https://doi.org/10.1111/j.1467-6494.2008.00545.x

Nezlek, J. B. (2017). A practical guide to understanding reliability in studies of within-person variability. *Journal of Research in Personality*, *69*, 149-155. https://doi.org/10.1016/j.jrp.2016.06.003

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231-259. https://doi.org/10.1037/0033-295X.84.3.231

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

Rammstedt, B., Danner, D., & Bosnjak, M. (2017). Acquiescence response styles: A multilevel model of individual and country differences. *Personality and Individual Differences*, *107*, 190-194. https://doi.org/10.1016/j.paid.2016.11.038

Roberts, C., Gilbert, E., Allum, N., & Eisner, L. (2019). Satisficing in surveys: A systematic review of the literature. *Public Opinion Quarterly*, *83*(3), 598-626. https://doi.org/10.1093/poq/nfz035

Rost, J., Carstensen, C., & von Davier, M. (1997). Applying the mixed Rasch model to personality questionnaires. In J. Rost & R. Langeheine (Eds.), *Applications of latent trait and latent class models in the social sciences* (pp. 324-332). Munster, Germany: Waxmann.

Rowland, M. L., Naidoo, S., Abdulkadir, R., Moraru, R., Huang, B., & Pau, A. (2010).

Perceptions of intimidation and bullying in dental schools: A multi-national study. *International Dental Journal*, *60*, 106-112. https://doi.org/10.1922/IDJ_2327Rowland07

Schriesheim, C. A., & Hill, K. D. (1981). Controlling acquiescence response bias by item reversals: The effect on questionnaire validity. *Educational and Psychological Measurement, 41*(4), 1101-1114. https://doi.org/10.1177/001316448104100420

Schwarz, N., Knauper, B., Hippler, H. J., Noelleneumann, E., & Clark, L. (1991). Rating scales - Numeric values may change the meaning of scale labels. *Public Opinion Quarterly, 55*(4), 570-582. https://doi.org/10.1086/269282

Shrout, P. E., & Lane, S. P. (2012). Psychometrics. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 302-320). The Guilford Press.

Smith, P. B. (2004). Acquiescent response bias as an aspect of cultural communication style. *Journal of Cross-Cultural Psychology*, *35*(1), 50-61. https://doi.org/10.1177/0022022103260380

Sorokowski, P., & Kowal, M. (2024). Reply to: Differences in response-scale usage are ubiquitous in cross-country comparisons and a potential driver of elusive relationships. *Scientific Reports*, *14*, 10951. https://doi.org/10.1038/s41598-024-60468-x

Sorokowski, P., Kowal, M., Sternberg, R. J. et al. (2023). Modernization, collectivism, and gender equality predict love experiences in 45 countries. *Scientific Reports*, *13*, 773. https://doi.org/10.1038/s41598-022-26663-4

Stantcheva, S. (2023). How to run surveys: A guide to creating your own identifying variation and revealing the invisible. *Annual Review of Economics*, *15*, 205-234. https://doi.org/10.1146/annurev-economics-091622-010157

Stening, B. W., & Everett, J. E. (1984). Response styles in a cross-cultural managerial study. *The Journal of Social Psychology*, *122*(2), 151-156. https://doi.org/10.1080/00224545.1984.9713475

Swain, S. D., Weathers, D., & Niedrich, R. W. (2008). Assessing three sources of misresponse to reversed Likert items. Journal of Marketing Research, 45(1), 116-131. https://doi.org/10.1509/jmkr.45.1.116

Trippas, D., Pennycook, G., Verde, M. F., & Handley, S. J. (2015). Better but still biased: Analytic cognitive style and belief bias. *Thinking & Reasoning*, *21*(4), 431-445. https://doi.org/10.1080/13546783.2015.1016450

Ulitzsch, E., Henninger, M., & Meiser, T. (2024). Differences in response-scale usage are ubiquitous in cross-country comparisons and a potential driver of elusive relationships. *Scientific Reports*, *14*, 10890. https://doi.org/10.1038/s41598-024-60465-0

Uz, I. (2015). The index of cultural tightness and looseness among 68 countries. *Journal of Cross-Cultural Psychology*, 46(3), 319-335. https://doi.org/10.1177/0022022114563611

van Herk, H., Poortinga, Y. H., & Verhallen, T. M. M. (2004). Response styles in rating scales: Evidence of method bias in data from six EU countries. *Journal of Cross-Cultural*

Psychology, 35(3), 346-360. https://doi.org/10.1177/0022022104264126

Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, *25*(2), 195-217. https://doi.org/10.1093/ijpor/eds021

Venaik, S., Midgley, D. F., & Christopoulos, D. (2021). Do within-subject standardized indices of societal culture distort reality? An illustration with the national Tightness culture scale. *Journal of World Business*, *56*(5), 101242. https://doi.org/10.1016/j.jwb.2021.101242

Webster, H. (1958). Correcting personality scales for response sets or suppression effects. *Psychological Bulletin, 55*(1), 62-64. https://doi.org/10.1037/H0048031

Weijters, B., & Baumgartner, H. (2012). Misresponse to reversed and negated items in surveys: A review. *Journal of Marketing Research*, *49*(5), 737-747. https://doi.org/10.1509/jmr.11.0368

Weijters, B., Baumgartner, H., & Schillewaert, N. (2013). Reversed item bias: An integrative model. *Psychological Methods*, 18(3), 320-334. https://doi.org/10.1037/a0032121

Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, *27*(3), 236-247. https://doi.org/10.1016/j.ijresmar.2010.02.004

Weijters, B., Geuens, M., & Schillewaert, N. (2009). The proximity effect: The role of inter-item distance on reverse-item bias. *International Journal of Research in Marketing*, *26*, 2-12. https://doi.org/10.1016/j.ijresmar.2008.09.003

Weijters, B., Geuens, M., & Schillewaert, N. (2010a). The stability of individual response styles. *Psychological Methods*, *15*, 96-110. http://doi.org/10.1037/a0018721

Weijters, B., Geuens, M., & Schillewaert, N. (2010b). The individual consistency of acquiescence and extreme response style in self-report questionnaires. *Applied Psychological Measurement*, *34*, 105-121. https://doi.org/10.1177/0146621609338593

Weijters, B., Schillewaert, N., & Geuens, M. (2008). Assessing response styles across modes of data collection. *Journal of the Academy of Marketing Science, 36*, 409-422. https://doi.org/10.1007/s11747-007-0077-6

Welkenhuysen-Gybels, J., Billiet, J., & Cambré, B. (2003). Adjustment for acquiescence in the assessment of the construct equivalence of Likert-type score items. *Journal of Cross-Cultural Psychology*, *34*(6), 702-722. https://doi.org/10.1177/0022022103257070

Welzel, C., Brunkert, L., Kruse, S., & Inglehart, R. F. (2023). Non-invariance? An overstated problem with misconceived causes. *Sociological Methods & Research*, *52*(3), 1368-1400. https://doi.org/10.1177/00491241219955 21

Wetzel, E., Böhnke, J. R., & Brown, A. (2016). Response biases. In F. T. L. Leong, D. Bartram, F. M. Cheung, K. F. Geisinger, & D. Iliescu (Eds.), *The ITC international*

handbook of testing and assessment (pp. 349-363). Oxford University Press. https://doi.org/10.1093/med:psych/9780199356942.003.0024

Wetzel, E., Böhnke, J. R., & Rose, N. (2016). A simulation study on methods of correcting for the effects of extreme response style. *Educational and Psychological Measurement*, *76*(2), 304-324. https://doi.org/10.1177/0013164415591848

Wetzel, E., & Carstensen, C. H. (2017). Multidimensional modeling of traits and response styles. *European Journal of Psychological Assessment*, *33*(5), 352-364. https://doi.org/10.1027/1015-5759/a000291

Wong, N., Rindfleisch, A., & Burroughs, J. E. (2003). Do reverse-worded items confound measures in cross-cultural consumer research? The case of the material values scale. *Journal of Consumer Research*, *30*, 72-91. https://doi.org/10.1086/374697

Zuckerman, M., & Norton, J. (1961). Response set and content factors in the California F-scale and the parental attitude research instrument. *Journal of Social Psychology*, *53*, 199-210. https://doi.org/10.1080/00224545.1961.9922118

1차원고접수 : 2026. 02. 27
2차원고접수 : 2026. 07. 20
최종게재결정 : 2026. 08. 22

한국심리학회지: 일반
Korean Journal of Psychology: General
2026, Vol. 45, No. 4, 643-686
http://dx.doi.org/10.22257/kjp.2026.9.45.4.643

A Literature Review on Response Styles in Surveys and Psychological Assessment: Causes, Consequences, Detection, Adjustment, and Prevention

Jae Yoon Chang
Department of Psychology
Sogang University

Taehun Lee
School of Psychology
Korea University

Self-report questionnaires are widely used in social surveys and psychological assessment to measure individuals' attitudes, personality traits, values, emotions, and behavioral tendencies. However, self-report data rely on the assumption that respondents answer items based on their substantive content. When respondents tend to answer in particular ways regardless of item content, the validity of measurement may be compromised. One representative phenomenon of this problem is response style. Response styles can be understood not as random error, but as systematic measurement error arising from various factors, including individual differences, item format, scale construction, survey context, and cultural norms. This study reviews previous research on major response styles observed in self-report measurement. Specifically, it examines key types of response styles, including extreme response style, midpoint response style, and acquiescent and disacquiescent response styles. It also discusses the causes of response styles, methods for their detection, statistical correction procedures, and preventive strategies at the stages of scale design and survey administration. In addition, the study reviews the influence of response styles in the context of cross-cultural research. Finally, this study presents practical guidelines that researchers and practitioners should consider in identifying and managing response styles.

Key words : social survey, psychological assessment, response bias, response style, acquiescent responding, extreme responding, midpoint responding, detection, adjustment, prevention, cross-cultural research