

제3장

조사표 평가 방법론에 관한 최근 동향

박영실 · 박현정

제1절 연구배경

조사표는 응답자로부터 응답을 얻어내는 일차적인 도구로써 자료의 품질향상에 직접적인 영향을 미칠 수 있다는 점에서 잠재력이 가장 큰 분야이기도 하다. 그러나 지금까지 조사표 영역은 조사와 관련된 오차 중 가장 이해하기 어려운 분야로 여겨지면서 조사방법론 분야의 다른 영역에 비해 큰 관심을 받지 못하였다. 조사표 설계 및 평가 등과 관련하여 견고한 이론적 틀의 마련 또한 비교적 최근에야 이루어지기 시작하였다(Sykes and Morton-Williams, 1987; Oksenberg, 1991; Demaio et al., 1993; Beatty and Willis, 2007).

그런데, 1980년대 후반 인지적인 접근법(cognitive perspective)이 조사 질문과 관련된 문제점을 파악하고 수정하는 기법으로 등장하면서 전환점이 마련되었다. 인지적 접근법은 사람들이 어떠한 언어 이해, 기억 처리, 판단 과정을 거쳐 질문에 반응 하는지를 응답 오차(response error)의 원인을 들어 설명하고 있다는 점에서 오차의 원인을 통계학적으로만 설명하려는 이전까지의 움직임과는 확연한 차별성을 보이고 있다(송현주 외, 2007). 이러한 흐름은 1983년 미국에서 심리학자, 사회학자, 인류학자, 조사방법론자 등이 참석한 가운데 개최된 소규모 컨퍼런스에서 출발하였다. 당시 컨퍼런스를 통해 본격적으로 조사방법론 분야에 인지심리학을 도입하자는 의견이 모아졌으며(Dippo et al., 1994), 1988년 미국 의회에서는 노동통계국(Bureau of Labor Statistics) 등에 인지실험실(cognitive laboratory) 설립을 위한 편드를 제공하면서 본격적인 연구가 이루어졌다. 이후 미국의 연방통계기관(Federal Statistical Agency)을 중심으로 인지실험실을 포함한 조사표 설계 부서가 조직되었으며, 인지적인 접근법을 적용해서 조사표를 평가하는 일은 조사표 설계 시 빼놓을 수 없는 과정으



로 자리 잡게 되었다(Willis, 1999).¹⁾ 이는 캐나다, 호주, 네덜란드 등 여타의 해외 통계작성 기관에서도 마찬가지였다(Hewett and Farrell, 2008; Houle, 2009; Snijkers, 2002).

이와 비교한다면, 우리나라에서 조사표의 설계에 대한 관심은 여러 측면에서 매우 미흡하다. 이론적인 측면에서 볼 때, 조사방법론 영역에서 조사표 설계 분야는 독자적인 영역을 구축하고 있는 학문적 성격 보다는 하나의 기법으로만 다루어지고 있다(김수택 외, 2009). 실무적인 차원에서 크게 다르지 않다. 인지적 접근법의 도입 이래로 조사표 설계 방법에 대한 개선이 많이 있어왔지만, 이에 대한 도입은 답보 상태에 머물러 있다. 통계청의 경우를 보면, 조사표 설계는 해당 조사를 수행하고 있는 부서 고유의 업무이며, 새로운 조사가 신설될 경우에 전문가 검토 및 시험조사를 통한 조사표 평가 이외의 다른 방법론에 대한 적용은 거의 이루어지지 않고 있다. 이러한 상황은 여타의 통계작성 기관에서도 크게 다르지 않을 것으로 예상된다.

조사표 설계가 자료품질향상에 미치는 영향력을 생각한다면 이 분야에 대한 관심을 더 이상 미루어서는 안 된다는 진단 하에, 본 연구에서는 인지적 접근법을 중심으로 조사표 평가 방법론 전반에 대해 검토해보고자 한다. 특히 최근 그 연구가 활발해지고 있는, 응답자의 질문 이해 과정을 파악함으로써 응답 오차를 진단할 수 있는 인지적인 접근법에 대해서 살펴보는 것은 매우 중요한 작업이다. 아울러, 미국을 비롯한 주요 국가의 통계청에 설립된 조사표 설계 관련 조직의 역할 등을 살펴봄으로써 각 기관에서의 조사표 설계 영역의 위치를 가늠해 보고 향후 우리나라 통계작성기관이 나아갈 방향에 대한 시사점을 도출해보고자 한다.

제2절 조사표 평가 방법론

가장 널리 알려져 있는 조사표 평가 방법은 실제 조사와 같이 훈련을 받은 조사원이 면접을 실시한 후 조사와 관련된 자신의 경험과 의견을 보고하는 현장조사(field test) 방법이다.²⁾ 그런데 이 방법은 조사원이나 연구자의 직관에 의존하는 경우가 많으며 문제를 찾아낼 뿐 원인을 파악하는 데에는 어려움이 있다는 한계가 인식되면서 대안적인 방법들

- 1) 미국 국립보건통계센터(National Center for Health Statistics)에서는 미국 연방통계기관에서 실시하고 있는 인지면접 테스트 결과물에 대한 데이터베이스인 Q-Bank를 운영하고 있는데, 해를 거듭할수록 그 축적량이 증가하고 있다. 1990년대까지만 해도 연간 1~2건에 해당하던 것이 2000년대에는 많은 해에는 14건 정도로 나타났는데, 이 데이터베이스에 수록되지 않은 것까지 감안한다면, 인지면접 테스트에 대한 수요를 예상해 볼 수 있다(www.nchnh.gov/qbank/home.aspx).
- 2) 본조사(main survey)에 앞서 조사 전반에 관한 평가를 위해 실시하는 현장조사 방법은 그 규모나 목적 등에 따라서 사전조사(pretest), 시범예행조사(dress rehearsal), 예비조사(pilot survey) 등으로 불리는데, 여기에서는 현장조사로 통칭하고자 한다.

이 개발되고 있다(Presser et al., 2004). 이 장에서는 조사표 평가 방법으로 활용되고 있는 주요 방법들을 살펴보고자 한다. 포커스 그룹(focus group), 전문가 검토(experts review), 인지면접(cognitive interviewing), 사용성 평가(usability test), 행동코딩(behavior coding), 응답자 및 조사원보고(respondent or interviewer debriefing), 현장실험(field experiment) 등이 주로 활용되고 있으며, 이는 실험실에서 진행되는 방법과 현장에서 진행되는 방법으로 구분해 볼 수 있다.³⁾

1. 실험실에서의 조사표 평가 방법론

실험실 안에서 비교적 소규모 참여자를 대상으로 이루어지는 조사표 평가 방법으로는 포커스 그룹, 전문가 검토, 인지면접, 사용성 평가 등이 있다. 포커스 그룹 및 전문가 검토는 우리나라의 통계작성기관에서도 많이 활용되고 있는 반면에 인지면접 및 사용성 평가의 활용은 매우 제한적이다.

가. 포커스 그룹

포커스 그룹은 사회자(moderator)에 의해 진행되는 소규모 참여자 간 토론으로 조사 주제에 관해 대상 모집단(target population)의 이해도를 파악하기 위한 평가방법이다(Groves et al., 2009). 포커스 그룹을 통해서 잠재적인 응답자가 조사 주제에 관해 무엇을 알고 있는지와 그들이 자신들의 사고를 어떻게 구조화하는지를 파악할 수 있다. 또한 응답자가 주제를 논의할 때에 사용하는 용어가 무엇이며, 그 용어를 어떻게 이해하고 있는지 조사할 수 있다. 이러한 점에서 포커스 그룹은 조사표 개발 초기 단계에 매우 유용한 방법으로 통한다(Census Bureau, 2003).

포커스 그룹이 다른 조사표 평가 방법론과 구별되는 가장 중요한 특징은 참여자간 의견이 그룹 간 상호작용(interaction)을 기반으로 한다는 것이다. 따라서 제한된 시간 안에 단일의 주제에 대해 많은 양의 정보를 얻을 수 있다(Willis, 2005). 그런데 상호작용과정 내에서 그룹 내 한 두 명이 토론 과정을 지배할 수도 있고, 다른 구성원의 기여를 제한할 수도 있다는 우려가 있다. 따라서 포커스 그룹에서는 무엇보다도 진행자의 역할이 중요하다. 진행자는 모든 참여자가 동등하게 토론에 참가할 수 있도록 해 주어야 하며, 토론의 주제에서 벗어나지 않도록 자연스럽게 다음 주제로 넘어갈 수 있도록 분위기를 조성해야 한다. 가장 큰 한계점으로 지적되고 있는 것은 토론의 참여자가 조사 모집단을 대표하지 않으므로 포커스 그룹을 통해서 도출된 인식이나 경험 분포를 일반화할 수 없

3) 조사표 평가 방법은 박영실·송호만(2009) 연구 결과의 일부분을 수정·보완하였다.

다는 점이다. 자료 또한 질적인 특성을 갖고 있어 계량화하기 어렵다. 하지만, 이러한 제한에도 불구하고 포커스 그룹은 조사표를 구성하기에 앞서서 조사 주제에 관한 정보를 파악하는 효율적인 방법으로 평가받고 있다(Groves et al., 2009).

나. 전문가 검토

전문가 검토는 조사표 설계 전문가(questionnaire design experts) 및 주제 전문가(subject matter experts)가 관련 분야의 지식을 근거로 조사표 상의 오차를 진단하는 방법이다. 주제 전문가가 해당 조사표를 통해서 조사의 목적에 부합하는 정보들이 수집되는지를 판단하는 역할을 수행한다면, 조사표 설계 전문가는 조사 질문 및 설계 그 자체를 검토함으로써 용어나 질문의 워딩, 질문의 구조, 응답범주, 질문순서, 조사표 흐름, 지시문 등에서의 문제를 발견하는 역할을 수행한다(Groves et al., 2009).

전문가 검토는 개별적으로 이루어지는 경우가 보통이나, 그룹 내 토론을 통해서 이루어지기도 한다. 그룹 내 토론을 통해서 전문가 검토가 이루어지는 경우 표준화된 코딩 지침을 바탕으로 각각의 질문을 평가하거나 포커스 그룹과 마찬가지로 좀 더 자유로운 방식으로 토론이 이루어지기도 한다(Willis, Schechter, and Whitaker, 1999).

전문가 검토는 조사표 문제를 식별하는데 있어서 가장 효과적이라는 의견이 있으나 그 결과 역시 전문가의 해석과 판단에 따라 달라지는 주관적인 의견이라는 단점을 갖고 있다. 최근에는 이러한 한계를 극복하고, 좀 더 타당성(comprehensive)있는 데이터를 얻기 위해 코딩 지침과 같은 정형화된 틀에 따라서 검토를 하는 것이 일반화되고 있다(Biemer and Lyberg, 2003). 질문에 관한 잠재적인 문제를 평가하기 위한 지침 개발 시도가 있어왔으며, 특히 레슬러와 포시쓰(Lessler and Forsyth, 1996)는 응답과정에서 인지적인 분석이 필요한 아래의 7개 범주를 제시한 바 있다(<표 3-1> 참고).

<표 3-1> 전문가 검토를 위한 코딩 범주

1. 읽기와 관련된 문제(Problems with Reading)
2. 지시문과 관련된 문제(Problems with Instructions)
3. 항목의 명확성과 관련된 문제(Problems with Item Clarity)
4. 가정과 관련된 문제(Problems with Assumptions)
5. 지식/기억과 관련된 문제(Problems with Knowledge/Memory)
6. 민감성/바이어스와 관련된 문제(Problems with Sensitivity/Bias)
7. 응답범주와 관련된 문제(Problems with Response Category)



다. 인지면접

현재 가장 많이 활용되고 있는 조사표 평가 방법은 인지면접으로, 이는 질문에 대한 응답자들의 이해 및 해석 과정에 대한 인지심리학적 접근에서 출발하였다. 일반적으로 응답자들은 질문이해(comprehension of the question) → 정보인출(retrieval from memory of relevant Information) → 판단(judgement) → 응답보고(response process) 순의 네 단계를 거쳐서 응답을 한다. 질문이해는 응답자가 질문을 해석하는 것이며, 정보인출은 응답자가 질문이 요구하는 정보를 기억으로부터 인출해내는 것, 판단은 생각해낸 정보를 통합하거나 요약하는 것이다. 이러한 과정을 거쳐 최종적으로 응답자는 응답 내용을 선택하여 질문에서 알고자 하는 대답을 하게 된다(Tourangeau, Rips and Rasinski., 2000).

응답형성의 각 단계에서 발생하는 오차를 진단하기 위해 면접원⁴⁾은 이해 및 해석과정을 파악할 수 있는 질문을 하고, 참여자는 자신의 응답과정에 대한 정보를 제공하는 방식으로 인지면접이 이루어진다(Hughes, 2004). 구체적으로 생각발화(think-aloud) 면접이나 캐어묻기(verbal probing) 등과 같이 응답의 전 과정을 펼쳐볼 수 있는 방법을 통해 조사 질문을 평가하는 방법이 사용된다(Demaio et al., 1993; Willis, 1999; Willis, 2005; Eurostat, 2006 등).

생각발화 면접은 참여자에게 응답과 관련된 정보를 인출하는데 사용된 절차에 대해서 묻고, 그 과정을 기록하는 것이다. 예컨대, 면접원은 참여자에게 “당신이 생각하는 것을 말씀해 주세요.” 혹은 “그것에 대해 좀 더 말씀해 주세요.” 등으로 질문함으로써 어떻게 응답에 도달하게 되었는지 전 과정을 듣게 된다. 캐어묻기는 질문-응답 과정 중 특정한 측면에 초점을 두어 참여자에게 질문을 하는 것이다. 면접원이 질문을 한 후 참여자가 응답을 하고 나면, 면접원은 다시 그 질문이나 응답 내용으로 돌아가 특정 용어의 의미를 무엇으로 생각했는지, 왜 그 선택지를 골랐는지, 준거기간을 어떻게 해석했는지 등을 묻는다. 생각발화 면접의 경우 참여자에게 개방형 질문을 하므로 면접원의 개입이 최소화되어 면접원에 의한 오차 발생 가능성이 적다는 장점이 있으나 참여자의 사고 능력에 따라서 얻어지는 정보의 양이 차이가 많다. 반면에 캐어묻기는 면접의 전 과정에 대한 통제가 가능하여, 참여자로부터 꽤 일관된 정보를 얻을 수 있다. 그러나, 캐어묻기 질문에 연구자의 의도가 반영됨으로써 인위적인 효과(artificiality)가 발생할 수 있음을 항상 주지해야 한다.

이외에도 비네트(vignette)나 바꾸어 말하기(paraphrasing), 신뢰도 평가(confidence ratings) 기법 등도 사용된다. 비네트는 응답자에게 가상의 이야기를 제공한 후 이 경우

4) 현장조사에서의 조사원 및 응답자와 구분하기 위해 인지면접에서는 맥락에 따라 면접원과 참여자로 칭한다.



라면 어떻게 응답할 것인지를 묻는 방법이다. 드물게 발생하는 상황 등에 대해서 응답자들의 이해과정을 파악하기 위한 목적으로 주로 활용된다. 바꾸어 말하기는 질문의 이해과정에서 응답자가 실제로 사용하는 용어를 식별하기 위한 목적으로 응답자 자신의 말로 다시 표현해 줄 것을 요구하는 기법이며, 신뢰도 평가는 응답자 스스로 자신의 응답 내용을 얼마나 신뢰할 만한 수준인지 평가하게 하는 것이다.

인지면접은 조사 실무자들에게는 잘 보이지 않는 오차의 원천을 참여자로부터 직접 발견할 수 있다는 점에서 그 활용도가 증가하고 있으나, 실험실에서 실시되는 다른 방법론과 마찬가지로 소규모의 자발적인 표본을 대상으로 한 면접 결과를 모집단 전체에게 일반화할 수 없다는 것이 한계점으로 지적되고 있다. 그런데, 동일한 질문과 개념에 대해서 응답자들이 반복적으로 혼란을 보인다면 주요 문제들에 대해서는 소규모 면접을 통해서도 발견될 수 있으며, 또한 표본규모가 작기 때문에 여러 번의 인지면접(인지면접 → 문제진단 → 수정 → 인지면접)을 통해 반복적인 사전 검사가 가능하다는 것이 오히려 장점으로 꼽히기도 한다(Demaio et al., 1993). 그러나 최근에는 컴퓨터 기술을 접목하여 대규모의 인지면접 결과에 대한 경험적인 분석 기법을 개발하여 이러한 한계점을 극복하는 시도들이 있다. 대표적으로 미국 국립보건통계센터에서는 인지면접 자료 분석 소프트웨어인 Q-notes 등을 개발하여 보급하고 있다. 인지면접의 전 과정을 데이터베이스화하여 응답자별 혹은 질문별로 다양한 분석을 쉽게 할 수 있도록 하고 있다.

포커스 그룹 및 전문가 검토는 우리나라의 통계작성기관에서도 오랫동안 많이 활용되고 있는 방법론이지만, 조사표 설계에 인지면접을 활용하는 사례는 매우 드물다(박영실·송호만, 2009; 박영실·박현정, 2012; 박현정·박영실, 2012).

라. 사용성 평가

컴퓨터가 조사원 행동 및 성과 등에 미치는 효과에 대한 관심이 증가하기 시작한 1990년대 후반부터 컴퓨터를 기반으로 하는 조사표 설계에서도 사용성 이슈가 증가하기 시작하였다(Hansen and Couper, 2004). 사용성 평가란, 인지면접의 일종으로 응답자가 목적을 완수하기 위해서 시스템을 얼마나 효과적으로 사용하는가에 초점을 맞추고 있다. 일반적으로 다음의 세 가지 지표로 정의된다. 첫째는 효과성(effectiveness)이다. 효과성은 조사표에서 얼마나 많은 오차가 발생하는지, 이 오차가 얼마나 심각한지 등을 측정한다. 둘째는 응답자가 얼마나 빨리 과제를 수행하는지를 측정하는 효율성(efficiency)으로 정의되며, 셋째는 응답자가 얼마나 디자인에 만족하는지를 측정하는 만족도(satisfaction)로 정의된다.⁵⁾

5) 노동통계국 사용성 평가 담당자(Jean Fox)와의 면담 결과(2011.11.8.)에 의한 것이다.

연구자는 조사과정에서 응답자가 응답부담을 느끼는 지점을 파악한 후 장애요소를 제거하여 보다 효율적으로 조사표를 설계할 수 있다. 사용성 평가 시에는 주로 시선추적장치(eye-tracker)를 이용하는데, 이것은 실험에 참여하고 있는 응답자의 안구 운동(eye movement)과 안구 주시 패턴(eye fixations pattern)을 살펴볼 수 있는 유용한 도구이다. 시선추적장치를 통해서 주시 횟수, 주시 기간 등에 관한 객관적인 자료가 확보 가능한데, 질적인 자료를 생산하는 인지면접의 한계를 보완해 줄 수 있다는 측면에서 선호되는 방법이다. 그러나 그 자료를 해석함에 있어서는 주의가 요구된다. 예컨대, 이 자료를 통해서 사람들이 주시하지 않는 것이 무엇인지는 말할 수 있으나, 그렇다고 해서 사람들이 실제로 보고 있는 것을 말할 수 있는 것은 아니라는 점이다. 따라서 사용성 평가 결과에 대한 유의미한 해석을 위해서는 인지면접과 같은 다른 평가 방법이 결합될 필요가 있다 (Census Bureau, 2003).

인지면접 및 사용성 평가 장비로 구축된 인지실험실은 일반적으로 응답자와의 면접이 이루어지는 면접실과 이 과정을 살펴볼 수 있는 관찰실로 구성된다. 이 때 면접실과 관찰실은 일 방향 거울(one-way mirror)로 연결되어 면접 과정에 대한 간섭 없이 관찰이 가능하도록 설계되어 있다. 최근에는 비디오카메라로 면접 과정을 촬영하고 그 과정을 관찰실에 있는 컴퓨터 모니터를 통해서 실시간으로 관찰할 수 있도록 바뀌는 추세에 있다. [그림 3-1]은 사용성 평가 과정을 보여주고 있다. 그림 상단의 왼쪽이 면접실, 오른쪽이 관찰실이다. 실험 참여자는 컴퓨터 화면의 모니터에 나오는 지시문에 따라서 주어진 과제를 수행하고, 연구자는 관찰실에 있는 컴퓨터를 통해서 참여자의 과제수행 과정을 관찰할 수 있다. 그림 하단은 과제수행 결과를 보여준다. 왼쪽은 시선추적장치를 통해 응답자가 어느 쪽에 얼마나 오래 동안 주시했는지를 화면상에서 색상 변화로 표현해 주고 있는 것이며, 오른쪽은 분석결과표 형태로 제시한 것이다.

국내에서 사용성 평가가 조사표 설계 분야에 직접적으로 활용된 예를 찾기는 쉽지 않다. 일반적으로 홈페이지 개발이나 제품의 사용성과 관련하여 마케팅 리서치 등에서의 활용이 보고되고 있으며(한우석, 2009), 조사표 설계와 관련해서는 송현주 외(2007)를 제외하고 그 예를 찾아보기 쉽지 않다. 송현주 외(2007)는 응답자들이 조사표 앞부분에 제시되는 조사표 기입 요령과 조사 질문 중간에 삽입되는 조건부 지시문 등을 어떻게 읽는지 확인하기 위해 시선추적연구를 통해 살펴본 바 있다. 총 10명에 대해 실험을 수행했는데 많은 참가자들이 조사표 기입 요령을 많이 응시하지 않았으며, 전혀 응시하지 않은 참가자도 3명이나 되었다. 조건부 지시문에 대해서는 10명의 참가자 중 조건부 지시문을 제대로 따르지 않아 응답하지 말아야 할 문항에 응답오류를 보인 참가자들이 4명 있었고, 이 4명 참가자의 주시시간 평균은 오류를 보이지 않은 참가자들에 비해 짧은 경향이 있었다.



[그림 3-1] 사용성 평가 과정

2. 현장에서의 조사표 평가 방법론

본조사에 앞서 실시하는 현장 조사는 조사표 평가에 있어서 가장 오랫동안 활용되어 오고 있는 방법이다. 현장조사를 통해 조사원 및 응답자로부터 관련 경험에 대한 의견을 듣거나 수집된 자료의 응답분포를 검토하는 방법으로 질문 평가가 이루어진다. 그런데 최근에는 현장조사 과정에 좀 더 체계적인 방법론을 적용하여 조사표 평가를 실시하고 있다. 조사원 및 응답자로부터 보고를 듣는 방법 이외에도 훈련된 코더(coder)가 조사원 및 응답자의 행동을 관찰하여 코딩하는 행동코딩 방법이나 수정된 조사표 설계 효과를 살펴보기 위한 무작위화 실험 등이 그러하다. 이러한 방법들은 앞에서 살펴본 실험실 방법론에서 발생할 수 있는 인위적 효과(artificial effect) 등을 배제하고 실제 조사환경에서 조사표 설계를 검토한다는 측면에서 중요한 의미를 갖는다(Eurostat, 2006).

가. 행동코딩

행동코딩 방법은 현장조사 과정에서 조사원이 제대로 질문을 했는지, 이에 응답자는 적절한 반응을 했는지 등 조사원과 응답자간 상호작용을 관찰하여 코딩을 하는 기법이다. 처음에는 조사원의 성과를 평가하고 모니터링 하는데 사용되었으나, 최근에는 질문-응답 과정 분석을 통해서 조사표를 평가하는 목적으로 활용되고 있다(Oksenberg, 1991; Demaio et al., 1993). 구체적인 코딩 지침의 예시는 <표 3-2>와 같다. 조사원이 질문을 쓰인 대로 읽었는지 혹은 의미를 바꿔서 읽었는지, 응답자는 질문에 적당한 응답을 제공했는지 혹은 질문에 대한 설명을 요청했는지 등을 코딩한다. 이 지침에 따라서 조사원이 질문을 정확하게 읽는 횟수 혹은 응답자가 응답에 어려움을 겪는 횟수 등을 분석한다. <표 3-2>는 조사원과 응답자 사이에서 발견되는 핵심적인 행동 예시이며, 각 조사에 적합하게 코딩 지침들을 개발하여 사용할 수 있다.

<표 3-2> 행동코딩을 위한 코드 범주

코드범주	내용
조사원 질문 행동 (하나 선택)	<ol style="list-style-type: none"> 1. 쓰여진 대로 정확하게 읽음 2. 약간 변화주어 읽음 3. 의미를 바꿔서 읽음
응답자 응답 행동 (해당되는 것 모두 선택)	<ol style="list-style-type: none"> 1. 질문 읽는 것을 방해 2. 질문을 명확하게 해줄 것을 요구 3. 적절한 응답 제공 4. 정확성이 보장된 답 제공 5. 부정확한 응답 제공 6. “모르겠다”라고 응답 7. 응답 거절

출처 : Groves et al., 2009

행동코딩은 표준화된 측정도구를 이용한다는 점에서 앞서 살펴본 다른 질적인 방법론과 비교해 볼 때 좀 더 체계적이고, 객관적인 평가를 할 수 있다는 장점을 갖는다(Hughes, 2004; Willis, 2005). 단, 유의미한 결과를 도출하기 위해서는 일정 정도의 표본 규모가 확보되어야 하는데, 건너뛰기(skip) 형식 등에 따라서 그 규모가 달라진다.⁶⁾ 또

6) 윌리스(Willis, 2005)는 코드의 분포를 통해서 문제를 판단하기 위해서는 한 질문에 최소한 12 표본 이상은 되어야 한다고 설명하고 있다.



한, 응답자가 자발적인 참여자라기보다는 실제 조사환경에서의 응답자와 유사하고, 자료가 수집되는 과정 또한 실제 환경과 동일하므로 분석 결과에 대한 일반화가 가능하다는 것도 중요한 특징이다(Willis, 2005). 하지만, 행동코딩을 통해 어떠한 항목에 문제가 있는지를 파악할 수는 있으나 그 원인이 무엇인지를 진단하는 것은 가능하지 않다. 행동코딩은 조사원과 응답자 행동을 통해서만 문제가 진단될 수 있기 때문인데, 만일 응답자가 질문 자체를 잘못 해석해서 응답하는 경우 이를 조사원이 지각하지 못한다면 문제점이 숨겨질 가능성이 높다(Census Bureau, 2003).

나. 조사원 및 응답자 보고

조사원과 응답자 모두로부터 유용한 정보를 얻을 수 있는 방법은 보고이다. 조사원 보고는 현장조사 완료 후 조사원으로부터 질문의 워딩이나 구조 등의 문제에 관해서 듣는 방법으로 현장조사의 한 단계로 수행된다(Campanelli, Martin, Rothgeb, 1991). 일반적으로 조사원 그룹 보고와 평가 양식을 통한 보고 두 가지 형태가 있다. 그룹 보고는 조사를 수행한 조사원들이 질문의 워딩, 질문 순서, 면접의 전체적인 흐름 등에 대한 자신의 조사 경험을 말하는 방법으로 이루어진다. 조사원들의 경험은 향후 조사를 수행할 때 발생할 수 있는 문제점을 수정하는데 유용한 정보로 활용된다. 이 경우 많은 수의 조사원이 필요하지는 않으나, 다양한 조사경험과 능력을 갖춘 조사원들이 포함되도록 하는 것이 좋다. 표준화된 평가 양식을 이용하여 각각의 질문을 평가하기도 한다. 조사원이 질문이 표현된 대로 정확히 읽는데 어려움이 있었는지, 응답자들이 그 질문에 대한 응답을 얻거나 제공함에 있어서 어려움이 있었는지, 응답자들이 일관되게 이해하지 못한 용어나 개념이 있었는지 등에 대해 ‘문제가 전혀 없다’에서부터 ‘문제가 있다’까지로 응답을 하게 한다(Demaio et al., 1993). 비록 조사원 그룹 보고가 질적인 자료의 성격을 갖고 있으므로 객관화시키기가 어렵다는 한계를 갖고 있긴 하나, 이는 평가 양식을 통한 보고 방법을 통해 보완이 가능하다.

응답자의 해석에 관한 정보를 도출하기 위해 응답자 보고를 실시하기도 한다. 응답자 보고도 조사원 보고와 마찬가지로 그룹 보고와 조사표를 통한 보고 두 가지 형태로 이루어진다. 그룹 보고 형태로 진행할 경우 어느 한 두 명의 응답자에 의해 토론이 좌우될 경우 발생할 수 있는 그룹 편차를 줄이기 위해 둘 혹은 그 이상의 그룹 간 토의를 동시에 진행하여 그 결과를 종합하는 방식으로 평가를 하기도 한다(Biemer and Lyberg, 2003). 조사 완료 이후에 응답 항목의 이해와 관련된 별도의 조사표를 배부해서 응답자로부터 다시 응답을 얻어내는 방법으로 보고가 이루어지기도 한다. 이런 이유로 일부에서는 응답분석조사(response analysis surveys) 또는 내용평가(content evaluations)라고 부르

기도 한다. 단, 연구자는 응답자 보고를 위한 조사표를 작성하기 이전에 조사표 문제에 대해 정확하게 인식하는 것이 필요하다. 예를 들면 특정 항목을 응답할 때 'a'라는 아이템을 포함해서 응답했는지 그렇지 않은지 등을 물어볼 수 있다. 이러한 정보 등을 통해 응답자가 특정 항목에 대해 왜 잘못 이해했는지를 판단하는데 유용한 정보를 얻을 수 있다(Census Bureau, 2003). 조사원 보고나 응답자 보고는 조사표 평가를 하기 위해 독립적으로 수행되기 보다는 다른 방법과 연동되어서 수행될 때 더욱 효과적이다(Bierner and Lyberg, 2003). 예컨대 특정 질문에 대해 행동코딩을 이용하여 문제를 진단했다면 그 원인을 응답자 보고를 통해서 파악할 수 있다.

다. 무작위화 실험

수정된 조사표가 기존의 조사표에 비해 개선되었는지를 평가하기 위해 무작위화 실험 방법을 사용한다. 이는 확률 추출된 표본을 다시 무작위로 하위 집단으로 나눈 후 각 집단에 서로 다른 버전의 조사표를 이용하여 자료를 수집하는 방법이다. 여기에서 핵심적인 것은 실험 설계 시 다른 모든 조건은 동일한 상태에서 검증하고자 하는 부분에만 차이를 주는 것이다. 이로써 응답 내용의 차이가 표집오차나 표본의 특성 등으로 인해 나타나는 것이 아니라 조사표에서 나타나는 것임을 확신할 수 있어야 한다. 이를 위해 적절한 표본 규모와 그 표본 내에서의 무작위 할당이 중요하다(Demaio et al., 1993).

수집된 자료를 이용하여 조사표 간 항목 무응답률과 응답분포 등을 비교한다. 특정 항목에 대해 무응답이 많다는 것은 그 만큼 응답자들이 어려움을 겪는다는 것을 의미하는 것이므로 질문을 평가하는데 매우 유용한 분석도구이다. 무작위화 실험에서 질문의 위당이나 질문의 순서에 따라 응답분포가 달라지는지 또한 주요한 분석 지표이다(Demaio et al., 1993). 그런데, 서로 다른 버전의 조사표에서 다른 응답분포가 발생하므로 실험만으로 어떤 버전이 다른 버전에 비해서 좋은지를 결정하지는 못한다. 이러한 이유로 인해서, 조사 결과를 비교할 수 있는 외부의 자료, 예컨대 행정자료 등과 비교를 하기도 한다(Groves et al., 2009). 한편, 자료 분석 시 내용검토(editing)와 무응답대체(imputation) 이전의 원 자료(raw data)를 사용해야 한다는 것을 주의해야 한다. 그 이유는 내용검토와 무응답대체가 질문의 문제점을 숨길 수 있기 때문이다.

무작위화 실험은 일반적으로 조사표 설계의 최종 단계에서 수행된다. 미국 노동통계국에서는 경제활동인구조사(Current Population Survey, 이하 CPS) 조사표를 개선하는데 인지면접 및 행동코딩, 무작위화 실험을 실시한 바 있다. 인지면접을 통해 CPS에서 사용



하고 있는 주요 개념들에 대한 문제를 진단한 후 새로운 버전의 조사표를 발전시켰으며, 대안 조사표를 평가하기 위해 무작위화 실험을 수행하였다. (Dippo et al., 1994). 이 조사 자료를 바탕으로 항목별 응답분포분석을 통해 새로운 질문 버전 혹은 구조가 노동력 추정치에 영향을 미칠 것으로 예상되는 응답항목에서 차이를 발생시켰는지 살펴보았다. 응답분포 분석을 통해 첫째, 동일한 정보를 얻기 위해서 서로 다른 질문전략이 발생시키는 효과, 둘째 질문의 구조가 발생시키는 차별적인 효과, 셋째 대안적인 질문이 더 많은 목표 모집단(targeted universe)을 포괄하는지 등을 확인할 수 있었다고 평가하였다(Demaio et al., 1993). 센서스국과 노동통계국은 이러한 절차를 거쳐 새롭게 수정된 질문으로 1994년 1월에 조사를 시작하였다.

지금까지 조사표 설계 및 평가와 관련된 다양한 방법론을 검토해 보았다. 조사표 설계에 있어서 조사표 평가 방법 간에 차이가 있는지를 살펴본 몇몇 연구들에서 흥미로운 결과들을 도출하였다(Presser and Blair, 1994; Rothgeb et al., 2001; Hughes, 2004). 연구 결과에 따르면 각각의 평가 방법은 조사항목이 갖고 있는 문제점을 성공적으로 지적해 냈으나, 어느 한 방법이 일관적으로 모든 문제를 식별해 내지는 못하는 것으로 나타났으며, 이들은 개별적으로 서로 다른 수준에서 질문의 문제점을 발견하는데 기여하였다. 각각의 방법은 고유의 장단점을 갖고 있으므로, 평가의 목적, 시간, 비용 등의 조건 하에 특정 방법을 선택하되, 다양한 방법을 고루 포함하는 반복적인 검사를 통해 조사표를 검토할 필요성이 있음을 시사한다.

제3절 해외 통계작성기관의 조사표 설계 조직 현황

앞서 언급했듯이 조사방법론 영역에 인지적인 접근법이 도입되면서 미국을 비롯한 주요 국가의 통계작성기관에는 조사표 설계와 관련된 독자적인 조직이 구축되기 시작하였다. 1980년대 후반 이후 조사방법론 연구 조직이 표본추출 등을 아우르는 수리적인 영역과 조사표 설계를 중심으로 하는 사회과학 영역으로 분리되기 시작하였으며, 이후 인지실험을 포함한 조사표 설계가 사회과학 영역 내에서 핵심적인 업무로 자리 잡았다. 여기에서는 미국의 주요 통계작성기관을 비롯하여 캐나다, 호주, 네덜란드 통계청 조사표 설계 관련 부서의 현황을 살펴봄으로써 향후 우리나라의 통계작성기관이 나아갈 방향에 대한 시사점을 얻고자 한다.

1. 미국⁷⁾

가. 센서스국

센서스국(Census Bureau)에서 조사표 설계 관련 업무를 수행하는 곳은 조사측정센터(Center for Survey Measurement, 이하 CSM)이며, 이 센터는 언어및측정연구(Language and Measurement Research), 가구조사조사표사전조사(Questionnaire Pretest for Household Survey), 인적요인및사용성연구(Human Factors and Usability Research)의 3개 부서로 구성되어 있다.

CSM에는 2011년 현재 약 22명의 직원이 있으며, 이 중 가구조사조사표사전조사 부서에는 7명이 근무하고 있다. 주로 심리학, 사회학, 컴퓨터공학 등의 전공을 배경으로 하고 있는데 조사측정센터의 전신이라고 할 수 있는 통계연구센터(Statistical Research Center)의 주 전공자가 통계학인 것과는 대비되는 모습이다. 실험 대상자 모집 및 관리 등을 위해 인지실험실 및 사용성 실험실에는 각각 1명이 상시적으로 고용되어 있다.

CSM의 핵심 업무는 신규 조사표 개발 및 기존 조사표 개선에 대한 조사표 사전평가 작업이다. 센서스국에서는 신규 또는 기존 항목을 수정하기에 앞서 인지실험등을 통한 테스트를 하는 것을 원칙으로 하고 있으며 이와 관련된 프로젝트를 자체 또는 외부와의 공동연구 등을 통해 진행하고 있다. 내부적인 프로젝트 뿐 아니라 외부기관에서 요청한 프로젝트 또한 수행하고 있는데 연 평균 15건 정도이다.

주로 이용하는 조사표 평가 방법은 포커스 그룹, 인지면접, 행동코딩 등이다. 인지면접 참여자는 각종 언론매체 또는 이메일 등을 통해 모집하고 참여자에게는 예산관리처(OMB: Office of Management and Budget) 규정에 따라 40달러가 지급된다.⁸⁾ 일반적으로 면접은 실험실에서 진행되지만 경제통계조사 부문은 조사 대상 사업체를 방문하여 수행되기도 한다. 행동코딩도 자주 사용하는 방법 중의 하나이다. 면접과정을 직접 관찰하거나 혹은 녹음내용을 듣고 분류하는 방법 둘 다 가능한데, 최근 컴퓨터 지원 녹음 면접(CARI, Computer Assisted Recording Interview) 방식에 따라서 녹음된 면접 내용을 행동코딩에 이용하는 방법이 선호되고 있다.⁹⁾

7) 미국 내 주요 통계작성기관의 조사표 설계 관련 부서 현황은 2011년 11월7일~10일 간 진행된 면담 결과를 토대로 작성된 것이다. 각 기관별 면담자는 다음과 같다.

- 센서스국: T. DeMaio, (2011.11.7.)
- 노동통계국: W. Mockovak, B. Kopp, S. Fricker, J. Fox, J. Edgar (2011.11.8.)
- 국립보건통계센터: K. Miller, A. Maitland (2011.11.10.)

8) 포커스 그룹 참여자에게는 75달러가 지급된다.

9) 컴퓨터를 활용하는 면접조사에서 면접과정의 일부를 무작위로 녹음하도록 하는 프로그램이 컴퓨터에 장착되어 있으며, 녹음된 내용의 일부를 행동코딩 분석에 활용한다. 분석된 내용은 또한 조사 관리자에게 제공되어 조사원 행동을 모니터링 하는데 사용된다.



자료 공표 및 분석에 초점이 맞추어져 있던 사용성 평가 또한 최근에는 조사 항목을 이해하거나 기술적인 문제를 식별하는데 활용하고 있다. 응답자의 시선이 어디에 오랫동안 멈춰져 있는지, 조사표 내에서 위아래, 좌우로 얼마나 이동하는지 등을 시선추적장치를 통해 확인한다. 일부의 컴퓨터에는 얼굴표정 감지기를 부착하여 조사표에 반응하는 응답자들의 표정을 관찰하기도 한다.

나. 노동통계국

노동통계국 내에는 조사방법론 전반에 대한 연구 부서라고 할 수 있는 조사방법연구부(Office of Survey Methods Research, 이하 OSMR)가 있으며, OSMR은 행동과학연구센터(Behavioral Science Research Center, BSRC)와 수리통계연구센터(Mathematical Statistical Research Center)로 구분되는데 이 중 BSRC에서 조사표 관련 업무를 맡고 있다. 총 22명의 OSMR 직원 중 BSRC에 근무하는 직원은 7명이며 이들은 심리학, 사회학, 사용성 관련 전공자이다.

조사표 설계 및 평가를 위해 BSRC에서 수행하고 있는 주요 방법은 전문가 검토, 포커스 그룹, 인지면접, 응답자 및 조사원 보고, 행동코딩, 무작위화 실험 등이다. 이들이 수행하는 연간 프로젝트의 수는 약 5개 정도이다. 시간활용조사(Time Use Survey) 등 노동통계국에서 주관하는 조사에 관한 연구 중심이며, 외부기관의 프로젝트는 수용하지 않고 있다. 조사표 설계 방법론 그 자체를 평가하는 것도 중요한 목적으로 두고 있기 때문에 이론적인 단계를 준수하여 프로젝트를 수행한다. 예컨대, 행동코딩 결과의 신뢰성 확보를 위해 까다로운 검증절차를 밟고 있다. 단일 조사에 적용되는 코딩 체계의 종류가 수십 개에 이를 정도로 매우 복잡하며, 이 코딩의 정확성을 위해서 코딩 요원 및 녹취자들에 대한 훈련을 철저히 하고 있다. 특히, 이중코딩(double coding)을 통해 코딩 요원 간 신뢰도 테스트를 하고 있는 점은 특기할 만한 사항이다. 인지실험실 운영과 관련해서는 응답자를 모집·관리하는 1명의 정규직 직원이 고용되어 있다. 일반인 대상 조사를 평가하는 경우에는 노동통계국의 데이터베이스를, 장애인 등 특수 계층 대상 조사를 평가하는 경우 광고나 전단지 등을 이용하여 참여자를 별도로 모집한다. 참여자에게는 OMB 규정에 따라 40달러를, 응답자 접근성을 고려하여 외부에서 면접을 할 경우 25달러를 지급한다.

다. 국립보건통계센터

국립보건통계센터의 조사표설계연구실(Questionnaire Design Research Laboratory, 이하 QDRL)은 1988년 설립 당시 5명의 직원으로 출발하였지만 현재 그 규모가 약 20명 정도

로 증가했다. 이 중 10명은 조사 질문 평가 업무를, 10명은 실험실 운영 업무를 담당(데이터베이스 구축, 행정업무, 실험참여자 모집 등)하고 있으며, 대부분이 사회학을 배경으로 하고 있다.

QDRL은 인지면접 방법론을 연구하고 해당 방법론을 실무에 적용하는 것을 핵심업무로 하고 있으며 포커스그룹이나 행동코딩과 같은 다른 방법론을 전혀 사용하지 않고 있다는 점에서 센서스국 및 노동통계국과 차이점을 보이고 있다. 이와 관련하여 인지면접 자료를 관리하고 분석하는 소프트웨어를 개발하여 보급하고 있다. QDRL에서는 약 5년 전부터 인지면접 자료 분석을 위한 소프트웨어인 Q-notes와 Q-video, Q-bank 개발에 주력하고 있다. 이는 인지면접 자료에 대한 분석(analysis), 투명성(transparency), 가산성(countability), 비교성(comparability) 등의 과학적 접근법이 강조됨에 따른 것이다. Q-notes는 인지면접 자료 분석프로그램이고, Q-video는 인지면접 비디오 자료, Q-bank는 관련 연구 결과를 축적하고 있는 데이터베이스이다. 초기에는 자료 분석 시 인지면접 녹음자료만을 사용하였으나, 현재는 녹음 자료와 비디오 자료를 통합하였으며, 이 자료를 분석 프로그램(Q-note)에 링크시킴으로써 결론에 어떻게 도달했는지 전 과정에 대한 추적이 가능하게 되었다.

2. 호주 및 캐나다 등

가. 호주 통계청

호주 통계청의 조사표 설계 관련 업무는 자료수집방법론부서(Data Collection Methodology, 이하 DCM)에서 수행하고 있다. 방법론·자료관리과(MDMD, Methodology and Data Management Division)의 하위 조직으로, 가구조사와 사업체조사 부서에서 별도로 운영되던 지원 조직을 하나로 통합하여 2007년 8월에 새로이 발족하였다. 과거에는 표본추출, 추정 등과 같은 방법론 지식과 DCM에서 요구되는 지식이 구분되지 않았으나, 조사표 설계 분야 지식의 전문화가 인정되면서 독립적인 분과로 인정받아 별도의 승진체계에 따라 운영되고 있다. 2008년 기준으로 약 7명이 근무하고 있으며, 대부분 심리학 혹은 사회학 전공자이다(Hewett and Farrell, 2008).

DCM의 주요 업무는 조사표 검토, 조사표 설계 및 평가에 관한 문헌연구, 실험 및 관찰 연구, 그리고 조사표 평가를 위한 훈련과 지원 등이다. 이와 함께 조사표에 대한 표준 지침을 연구하고 개발한다. 관련 매뉴얼로 조사표 설계 표준 매뉴얼(Forms Design Standards Manual), 조사표 개발 및 평가 매뉴얼(Forms Development and Evaluation Manual)이 있는데 이는 장기적인 관점에서 호주 통계청의 조사 업무 수행체계에 지대한



영향을 미친다는 점에서 중요하다. 특히 조사표 설계 표준 매뉴얼은 호주 통계청 사업체 조사의 형식에 대한 일관된 가이드라인을 제공하고 있다. 조사표 설계 및 평가 매뉴얼은 조사표 평가 방법 및 수행에 대한 개선을 목적으로 하는 것인데, 이를 위해 DCM에서는 연간 3~4회, 매회 수일동안 평가에 직접 참여하고 있으며, 새롭게 추가될 내용 등이 발견될 경우 숙련된 전문가의 검토를 거쳐서 매뉴얼에 반영한다(Hewett and Farrell, 2008).

미국과 마찬가지로 조사표 평가 과정에 인지면접, 포커스 그룹 등의 인지적 접근법과 함께 전문가 검토, 행동코딩, 조사원 보고 등을 적용하고 있다. 인지면접 방법은 보편화된 지 얼마되지 않았으나 호주 통계청에서 선호되는 방법으로 자리잡았다. 특히, CATI(Computer Assisted Telephone Interviewing) 인터페이스의 사용성 평가를 자주 수행하고 있다. 컴퓨터 화면 상의 디자인 뿐 아니라 질문에 대한 응답자의 이해정도를 파악하기 위해 시선추적연구 등을 실시하여 조사표에 관한 제언을 하도록 권고하고 있다(ABS, 2001).

나. 캐나다 및 네덜란드 통계청 등

조사표설계연구센터(Questionnaire Design Resource Centre, 이하 QDRC)는 캐나다 통계청에서 조사표 설계 업무를 담당하는 기관으로 1986년에 최초로 설립되었다. 이듬해인 1987년에 ‘1991 센서스 조사표 개발’의 일환으로 인지면접과 포커스 그룹이 처음으로 도입되었으며, 1990년대 중반 이후 그 영역을 인구사회 분야는 물론 경제 및 농업 방면까지 확대하여 조사표 설계 및 평가 업무를 수행해오고 있다. 2004년 이래로 QDRC에서 수행되는 연간 조사표 평가 프로젝트 중 거의 대부분이 농업 혹은 사업체 조사에 관한 것이기도 하다. 오늘날 캐나다 통계청의 조사표 대부분은 인지면접과 포커스 그룹을 활용하여 평가하고 있으며, 지난 18년간 인지면접을 통해 평가된 조사의 수가 약 300개에 달한다(Houle, 2009; Lawrence, 2009).

QDRC의 정체성이 확립된 시점은 1992년이다. 그 이전에는 캐나다 통계청에서 조사표 설계와 관련된 평가 업무가 주로 민간리서치전문기관에 의해 수행되었으나, 1992년을 기점으로 QDRC의 핵심 업무가 되었다. QDRC 초기에는 당시 민간리서치기관에서 주로 활용하고 있던 포커스 그룹 중심으로 평가가 이루어졌으나, 이후 인지면접이 도입되면서 이 분야에 대한 업무가 활발히 진행되었다. 1994년에 마련된 ‘조사표 설계 및 평가에 관한 정책(Policy on the Review and Testing of Questionnaire at Statistics Canada)’이라는 규정이 2002년 개정되면서, 조사표의 신규 개발이나 개편 시에는 반드시 QDRC의 검토를 거치도록 하고 있다(Houle, 2009). 최근에는 CATI 및 CAWI(Computer Assisted Web Interviewing)의 개발과 평가 방법에 관한 연구를 중점과제로 추진하고 있다. 아직은

웹기반 조사에 대한 사용성 평가 경험은 많지 않지만 이 분야에 대한 견고한 기법을 개발하는 것 또한 캐나다 통계청의 과제로 설정하고 있다(Lawrence, 2011).

한편, 네덜란드 통계청은 캐나다 통계청의 QDRC를 벤치마킹하여 1992년에 조사표 실험실(The Questionnaire Laboratory, 이하 QL)을 설립하였다. QL은 설립 당시에는 통계방법론부서(Department of Statistical Methods)에 위치하였으나, 1994년 말에 자료수집방법론부서(Department of Data Collection Methodology)로, 2000년에 다시 통계방법론 부서로 이동하는 등의 개편과정을 거쳤다. 1994년 당시 직원은 9명이었는데, 조사표 설계 및 평가와 관련된 연구팀에 7명이, 그리고 실험참여자 채용, 블레이즈 시스템하의 조사표 준비 등과 관련된 행정팀에 2명이 배치되었다. QL이 설립된 이래 처음 5년 동안 연간 5~10개 정도의 프로젝트를 수행하였으며 이들 프로젝트의 대다수는 네덜란드 통계청 내부의 수요에 의한 것이었다(Snijkers, 2002).

제4절 결 론

조사표 설계는 자료의 품질을 향상시키는데 직접적으로 영향을 미치는 중요한 분야임에도 불구하고 우리나라에서는 지금까지 이 분야가 연구 대상에서 배제되어온 경향이 있다. 1980년대 말 이후로 인지적인 접근 방법이 조사방법론 영역에 소개되면서 해외 주요 통계작성기관에서 이에 대한 연구를 활발히 진행해 오고 있는 것과는 매우 대조적인 모습이다. 이에 본 논문에서는 조사표 설계 시 활용되고 있는 주요 평가 방법과 해외 통계작성기관의 조사표 설계 관련 조직 현황 등에 대해 살펴봄으로써 향후 우리나라 주요 통계작성기관이 나아갈 방향 등에 대한 시사점을 찾고자 하였다.

최근에는 연구자 또는 조사원 입장에서 조사표를 평가하던 전통적인 방법에서 벗어나 응답자를 중심으로 조사 질문에 대한 응답자의 이해과정을 파악하여 조사표를 평가하는 인지적 접근법이 강조되고 있다. 현장조사를 통한 행동보고, 조사원 보고 등에서 벗어나 포커스 그룹, 인지면접, 사용성 평가 등 응답자의 사고과정을 이해하는 조사표 평가 방법론이 많이 활용되고 있다. 조사과정에서 조사원 및 응답자 모두 자료의 품질에 영향을 미치는 주요한 오류의 원천이므로 조사표 평가 시 양쪽 입장을 충분히 검토해야 한다는 입장을 반영한 것이다. 실제로 미국, 캐나다, 호주 통계청 등은 인적·물적 기반을 갖춘 상태에서 조사표 설계와 관련된 업무를 체계적으로 수행하고 있으며, 이러한 역사가 약 25년에 걸쳐 지속되고 있다.

반면, 우리나라는 조사표 설계 및 평가 방법론의 실무 적용이 매우 더딘 편이다. 인

지면접이나 사용성 평가 등을 직접적으로 적용한 사례는 드물며, 적용한 경우에도 아직 시험 단계라 할 수 있겠다. 마케팅 리서치 분야에서 광고효과나 제품의 유용성 등을 평가하는 데에 사용성 평가가 일부 활용되고 있는 것과 달리, 조사표 설계 분야에 대한 활용도는 매우 낮다. 외국에서 전자조사표의 개발 등에 사용성 평가가 중요한 방법론으로 부각되고 있는 점을 비추어 본다면, 인터넷 조사의 적용률이 점차 증가하고 있는 우리나라에서 사용성 평가 방법론의 도입은 자료품질 향상에 견인차 역할을 할 것으로 기대된다. 조사표 설계가 조사에서의 오차를 줄여주는 데 있어서 잠재력이 큰 분야임에도 불구하고 우리나라의 경우 외국에 비해 조사표 설계 관련 분야에 관한 관심이 부족하여 그 발전 정도가 많이 지체되어 있으며 이를 극복할 수 있는 적절한 벤치마킹 모델이 필요한 상황에서 공식통계기관으로서의 통계청의 역할이 무엇보다 중요하다고 하겠다.



참고문헌

- 김수택 · 김영원 · 류제복 · 박진우 · 변종석 · 이기성 · 이해용 · 이홍철 · 최경호 · 한근식 · 홍기학.
2004. 조사방법의 이해. 교우사.
- 송현주 · 손영우 · 황명진 · 박영실. 2007. 인지실험을 통한 조사표 개발 방법 연구. 통계개발원.
- 박영실 · 송호만. 2009. 근로형태별 부가조사표 개선방안: 인지적 접근법의 적용. 통계개발원 (미간행).
- 박영실 · 박현정. 2012. 2015 인총 표본조사표 설계. 통계개발원 (발간예정).
- 박현정 · 박영실. 2012. 자원봉사 규모와 실태의 측정방안. 통계개발원 (발간예정).
- 한우석. 2009. 사용성 평가(Usability Testing) 조사 소개. 리서치 봄호: 36-41.
- Australian Bureau of Statistics. 2001. Pretesting in Survey Development: An Australian Bureau of Statistics Perspective. Australian Bureau of Statistics.
- Beaty, P. C. and G. B. Willis. 2007. "Research Synthesis: The Practice of Cognitive Interviewing." *Public Opinion Quarterly* 71(2): 287-311.
- Biemer, P. P. and Lyberg, L. E. 2003. *Introduction to Survey Quality*. Hoboken, New Jersey: John Wiley & Sons.
- Campanelli, P. C., Martin, E. A., and Rothgeb, J. M., 1991. "The Use of Respondent and Interviewer Debriefing Studies as a Way to Study Response Error in Survey Data." *The Statistician* 40: 253-264.
- Census Bureau. 2003. Census Bureau Standard: Pretesting Questionnaires and Related Materials for Surveys and Censuses. Census Bureau.
- Lawrence, D. 2009. "Questionnaire Testing for Business Surveys at Statistics Canada: Practice and Considerations." *Presented at QUEST workshop*.
- Lawrence, D. 2011. "Developing Electronic Questionnaires at Statistics Canada: Issues and Challenges for Questionnaire Design." *Presented at QUEST workshop*.
- DeMaio, T., N. Mathiowetz, J. Rothgeb, M. Beach, and S. Durant. 1993. Protocol for Pretesting Demographic Surveys at the Census Bureau. Census Bureau.
- Dippo, C., A., Polivka, K., Creighton, D., Kostanich, and J. Rothgeb. 1994. Redesigning a Questionnaire for Computer-Assisted Data Collection: the Current Population Survey Experience. Bureau of Labor Statistics.
- Eurostat. 2006. *Handbook of Recommended Practices for Questionnaire Development and Testing in the European Statistical System*. Eurostat.
- Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., and Tourangeau, R, 2009, *Survey Methodology*, John Wiley & Sons, Inc.
- Hansen, S. E., and M. P. Couper. 2004. "Usability Testing to Evaluate Computer-Assisted Instruments" In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, E. Singer (Eds.). *Methods for Testing and Evaluating Survey Questionnaires*. Willey Series in Survey Methodology.

- Hewett, K. and E. Farrell, 2008, Organization of Data Collection Methodology Services in the Australian Bureau of Statistics, Proceedings of Statistics Canada Symposium.
- Houle, P. 2009. "Reducing the Number of Cognitive Interviews by Adding Other Cognitive Methods of Testing." *Presented at Statistics Canada's International Symposium.*
- Hughes, K. A. 2004, Comparing Pretesting Methods: Cognitive Interviews, Respondent Debriefing, and Behavior Coding. Census Bureau.
- Lessler, B. H. Forsyth, J. 1996. A Coding System for Appraising Questionnaires. In Answering Questions: Methodology for Determining Cognitive and Communicative Processes. In N. Schwarz, S. Sudman. *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research.* Sanfrancisco (California): Jossey-Bass.
- Oksenberg, L., C. Cannell, and G. Kalton, 1991, New Strategies for Pretesting Survey Questions. *Journal of Official Statistics* 7(3): 349-365.
- Presser, S. and J. Blair, 1994, Survey Pretesting: Do Different Methods Produce Different Results? In P. V. Mardsden (Ed.). *Sociological Methodology*, Cambridge, MA: Blackwell.
- Presser, S., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., Rothgeb, J. M., & Singer, E. 2004. "Methods for Testing and Evaluating Survey Questions." *Public Opinion Quarterly* 68: 109-130.
- Rothgeb, J., G. Willis, and B. Forsyth, 2001, Questionnaire Pretesting Methods: Do Different Techniques and Different Organizations Produce Similar Results. *Presented at the Annual Meeting of the American Association for Public Opinion Research.*
- Snijkers, G. 2002. *Cognitive Laboratory Experiences: On Pretesting Computerized Questionnaires and Data Quality.* Statistics Netherlands
- Sykes, W. and J. Morton-Williams 1987 "Evaluating Survey Questions." *Journal of Official Statistics* 3(2): 191-207.
- Tourangeau, R., L. J. Rips., and K. Rasinski. 2000. *The Psychology of Survey Response.* Cambridge.
- Willis, G. B. 1999. "Cognitive Interviewing: A How To Guide". *Presented at the Meeting of the American Statistical Association.*
- Willis, G., S. Schechter, and K. Whitaker, 1999. "A Comparison of Cognitive Interviewing, Expert Review, and Behavior Coding: What Do They Tell Us?" *Presented at the Meeting of the American Statistical Association.*
- Willis, G., 2005, *Cognitive Interviewing: A Tool for Improving Questionnaire Design.* Sage Publications.

