

제3장

모니터링 시스템을 이용한 조사원 업무성과 측정에 관한 연구

박영실

제1절 서론

조사원의 업무성과(interviewer performance)를 측정하는 일은 통계작성기관에서 매우 중요한 부분을 차지한다. 조사원은 자료수집과정에서 중요한 대행자(agent)로, 이들의 모든 행동이 자료의 품질과 직결되어 있기 때문에 통계작성과정과 그 산출물의 품질을 평가함에 있어서 조사원의 업무성과는 반드시 측정되어야 한다. 조사원의 업무성과는 조사비용(survey costs)과도 관련된다. 조사원 수당은 조사비용 중 단일항목으로는 가장 큰 부분을 차지하는 것으로, 조사업무와 관련된 생산성을 향상시키기 위한 작은 노력도 전체 예산에 영향을 미치게 된다. 따라서 조사기관들은 조사비용을 현실적으로 추정하기 위해서 조사원의 업무성과를 측정해야 한다(Tarnai and Moore, 2006). 업무성과는 또한 슈퍼바이저와 조사원 사이의 커뮤니케이션을 위한 기초자료로 활용된다. 조사원의 업무성과를 근간으로 하여 조사원에게 업무를 재할당하기도 하고, 보상을 하기도 하며, 재교육을 시킬 수도 있다.

그런데 조사원 업무성과 측정이 갖는 이러한 목적에도 불구하고, 업무성과 측정대상이나 측정방법 및 측정지표 등에 대해서는 잘 알려져 있지 않고 그 중요성 또한 과소평가되어 온 경향이 있다. 이는 조사원 업무성과가 연구의 대상이라기보다는 주로 행정적인 관심(administrative

concern)으로 여겨져 왔으며, 업무성과를 측정하는 일이 필요한 일이라고는 생각하지만 성가신 부담으로 작용해 왔기 때문이다(Fellegi and Brackstone, 1999). 하지만 좀 더 실질적인 이유는 다음의 두 가지로 요약될 수 있다.

첫째, 조사원들이 수행하는 조사업무에 대한 직접적인 관찰이 어렵다는 점이다. 업무성과를 제대로 측정하기 위해서는 조사원을 관찰하는 일이 선행되어야 하는데, 이는 현실적으로 쉽지 않기 때문에 조사원에 대한 연구의 대부분은 수집된 자료를 바탕으로 조사원 효과(interviewer effect)를 추정하는 데 치중되어 온 경향이 있다(Groves, 2004). 둘째, 조사원 업무성과를 측정하는 데 있어서 보편적으로 타당하다고 판단되는 객관적인 기준을 마련하는 것이 어렵다(Kennickell, 2006). 조사원 개인마다 학습능력, 조사업무에 대한 기대치 등이 다양할 뿐만 아니라 조사원이 할당받은 업무와 그 업무가 놓여 있는 상황이 다르기 때문에 조사원의 업무성과를 객관적으로 측정하기란 쉽지 않다. 특히 국가통계기관처럼 다양한 특성을 지닌 다수의 조사원이 다양한 지역에 걸쳐서 조사업무를 수행하는 경우, 이들의 업무성과를 측정하여 비교하는 것은 어려운 일이다.

그러나 과학기술의 발전과 함께 조사업무에 컴퓨터 등이 보조적으로 활용되면서 조사원의 업무, 특히 면접과정을 모니터링하는 일이 좀 더 수월해졌다. 모니터링을 통해서 조사원이 자료의 품질에 미치는 영향을 통제하는 일이 어느 정도는 가능해졌고(Fowler, 1991), 그 과정에서 부수적으로 각종 정보 수집이 용이해짐에 따라 조사원 업무성과 측정에 대한 관심이 증가하고 있다.

이에 본 연구에서는 우리나라 통계청의 조사원 업무성과 측정과 관련한 현황을 검토한 후, 앞으로의 개선방향에 대한 제언을 해 보고자 한다. 앞서 문헌고찰 및 사례들을 통해서 조사원의 업무성과 측정을 위한 개념적인 틀을 살펴보고자 한다. 조사원의 업무성과 측정은 크게 세 부분으로 구분해서 생각해 볼 수 있다. 어떠한 업무를 측정할 것인가, 그 업무들을 어떠한 방법과 기준으로 측정할 것인가가 그것이다. 본 연구는 조사원 업무성과 측정을 위한 기초연구로서 이 중에서도 측정대상과 측정방법을 중심으로 살펴볼 것이며, 구체적인 측정지표에 대한 연구는 추후의 과제에서 진행하고자 한다.

제2절 조사원 업무성과 측정을 위한 개념적인 틀

1. 측정 대상

조사원 업무성과는 조사원이 수행하는 업무를 근거로 하여 측정되어야 한다. 조사원의 업무가 어디까지인가에 대해서는 통계작성기관 및 조사의 목적, 자료수집방법 등에 따라서 달라질 수 있지만, 일반적으로 자료수집방법의 영향을 많이 받는다. 전통적으로 가장 많이 사용하고 있는 면접조사방식에서 조사원의 업무범위를 살펴보면 다음과 같다.

지역할당표본추출방법을 사용하고 있는 가구조사의 경우를 예로 들어보면, 조사원은 먼저 추출된 지역에 방문하여 주소를 기록, 표집 틀(sampling frame)을 구축한다. 다음으로 응답가능(eligible)한 가구(혹은 개인)를 기록하고 추출방법에 따라서 응답자를 선택한 후, 이들을 방문하여 협조를 이끌어낸다. 조사원은 응답자에게 질문을 하고, 만일 이들이 응답을 제대로 하지 못할 경우에는 다시 질문을 한다거나 캐어묻기(probing) 등을 함으로써 응답자들이 그 의미를 잘 이해하고 해석하여 응답할 수 있도록 도와준다. 이외에도 조사원은 응답을 기록하고, 때로는 자료를 입력하여 중앙 조직에 전송하는 일을 하기도 한다(Groves et al., 2004). 면접조사인지, 컴퓨터를 보조적으로 이용한 조사인지 등에 따라서 위 업무 중 특정 부분의 업무가 제외되거나 혹은 추가되기도 하나, 조사원 업무에서 가장 중요한 것이 응답자를 접촉하고 그들의 협조를 얻어내는 것과 실제 면접을 진행하는 것이라는 데에는 누구나 동의할 것이다. 그러므로 조사원의 업무성과 측정은 이 업무를 중심으로 이루어져야 할 것이다.

각 업무에서의 성과 측정은 산출물(output)과 과정(process)에서의 평가라는 두 차원에서 진행되어야 한다(Loosveldt, Carton, and Billiet, 2004). 전통적인 관점에서의 업무성과는 산출물에 대한 최종 품질로 평가되어 왔으나, 최근에 품질에 대한 개념이 좀더 포괄적(comprehensive)으로 바뀌면서 산출물과 산출물이 도출되는 과정을 포함하는 품질에 대한 평가

작업이 동시에 시도되고 있다(Saeboe, Byfuglien, and Johannessen, 2003). 과정에 대한 평가는 조사원이 각 업무에 대한 교육이나 준비를 얼마나 철저히 하였으며, 실제 과정에서 이것을 얼마나 준수하였는지를 살펴본다. 또한 조사 과정에서 응답자를 접촉하고 협조를 얻기까지의 절차에 대한 평가와 면접을 진행하는 조사원 행동에 대한 평가가 이루어진다. 산출물에 대한 평가는 총조사오차(total survey error) 중 비표본오차와 밀접하게 관련되어 있다.1)

〈표 3-1〉 조사원 업무성과 측정대상

구 분	응답자 접촉 및 협조 유도	면 접
과정 평가	교육 및 준비 준수 및 피드백 접촉 절차 평가	교육 및 준비 준수 및 피드백 조사원 면접행동 평가
산출물 평가	무응답 오차	측정오차

자료: Loosveldt, Carton, and Billiet(2004).

응답자 접촉 및 협조 유도 업무의 산출물 평가는 무응답 오차(nonresponse error)에 대한 평가를, 면접 업무는 측정오차(measurement error)에 대한 평가를 중심으로 조사원 업무성과가 측정될 수 있다(Loosveldt, Carton, and Billiet, 2004)(표 3-1 참고).

가. 응답자 접촉 및 협조유도 업무 과정 평가

응답자 접촉 및 협조유도 과정에 대한 평가는 접촉절차, 즉 조사원이 표본단위를 접촉하기 위해서 어떠한 준비를 하고, 어떻게 접촉을 하여 설득하는지에 대한 정보를 바탕으로 이루어져야 한다. 이와 같은 과정에 대한 평가는 표본방문기록표(contact description form)를 통해서 할 수 있다. 본 연구에서는 AAPOR(American Association Public Opinion Research)

1) 총조사오차는 표본추출오차(sampling error)와 비표본추출오차(nonsampling error)로 구성된다. 표본추출오차란 모집단에 대한 완전한 센서스를 실시하는 대신에 단지 표본을 통해 결과를 산출함으로써 발생하는 오차를 말하며, 비표본추출오차란 자료수집 및 처리과정과 관련된 오차를 말한다(Biemer and Lyberg, 2003).

에서 개발한 표준화된 배치코드(standardized disposition code)를 소개하고자 한다.²⁾ 이 코드는 조사원이 조사업무를 수행하는 모든 과정에 대한 이력을 체계적으로 기록할 수 있는 지침을 제공해 준다.

코드는 최종배치코드(final disposition code)와 임시배치코드(temporary disposition code)의 두 종류로 구분된다. 최종배치코드는 모든 조사가 완료된 이후에 표본의 최종 상태를 보여주는 것이며, 임시배치코드는 해당 표본이 최종배치코드로 전환되기 전에 구체적으로 어떠한 시도를 했으며, 어떠한 결과들이 도출되었는가를 보여준다.

최종배치코드는 ① 면접을 완료한 경우(interview), ② 응답적격하나 면접을 완료하지 않은 경우(eligible, non-interview), ③ 면접을 완료하지는 않았으나 응답적격여부가 알려지지 않은 경우(unknown eligibility, non-interview), ④ 응답이 부적격한 경우(ineligible)의 네 가지로 구분된다. 이 코드는 세부적으로 일시적인 상태(temporary status)를 반영하는 코드를 포함하여야 한다. 임시배치코드는 최종배치코드 중 응답적격하나 면접을 완료하지 않은 경우(②)와 면접을 완료하지 않았으나 응답적격여부가 알려지지 않은 경우(③)에 고려될 수 있으며, 이 코드에는 면접이 완료되지 않은 다양한 이유가 포함된다(표 3-2 참고).

임시배치코드에서 최종배치코드로의 전환 과정에 대한 평가가 갖는 함의는 크다. 예를 들면 응답자에 의한 거부인지, 다른 가구원에 의한 거부인지, 면접을 피하기 위해 약속을 깨는 것인지 등에 대한 분석을 통해서 거부전환과정(refusal conversion process)을 파악할 수 있기 때문이다. 그리고 이러한 모든 과정들은 슈퍼바이저들에 의해 평가된다. 슈퍼바이저들은 조사원이 방문기록 노트를 잘 남겼는지(누가, 무엇을, 어디에서), 방문이력(visit history)을 효율적으로 검토하였는지, 시간을 효율적으로 사용하였는지 등을 평가한다.³⁾

2) AAPOR는 여러 통계작성기관들의 표본 방문 코드를 비교해 보고 비슷한 분류를 채택하고 있는 경우가 매우 드물다는 것을 발견, 표준화된 체계를 제안하였다.

3) 조사과정에 대한 기록과 함께 표본대체(substitutions)와 대리응답(proxies) 또한 보고되어야 한다. 현장조사가 어려워짐에 따라 표본을 대체하거나 응답자가 아닌 다른 가구원이 대리로 응답하는 경우가 종종 발생한다. 표본대체와 관련하여서는 사용된 표본 대체규칙이 무엇인지—표본대체에는 일반적으로 두 가지 유형이 있는데 하나는 표본내(within) 대체이며, 다른 하나는 표본간(between) 대체이다. 어떤 표본이 대체되었는지, 대체 이전 표본 특성은 어떠한지 등을 기록해야 한다. 대리응답도 보고되어야 한다. 대리라 함은 원래 추출된 표본이 아닌 제3자가 대신 응답하는 것으로 가구조사

<표 3-2> 임시배치코드의 예

<p>■ 응답적격하나, 면접이 완료되지 않은 경우</p> <ul style="list-style-type: none">- 다른 가구원에 의한 약한 거부(soft refusal, other)- 다른 가구원에 의한 강한 거부(hard refusal, other)- 응답자에 의한 약한 거부(soft refusal, respondent)- 응답자에 의한 강한 거부(hard refusal, respondent) - 소개 도중에 조사 결렬(breakoff, during introduction)- 면접 시작 후 조사 결렬(breakoff, after interview started) - 응답자와의 구체적인 약속(definite appointment, respondent)- 다른 가구원과의 구체적인 약속(definite appointment, other)- 응답자와의 확실하지 않은 약속(unspecified appointment, respondent)- 다른 가구원과의 확실하지 않은 약속(unspecified appointment, other) - 기타, 슈퍼바이저와의 상의(other, referred to supervisor) <p>■ 면접을 완료하지는 않았지만, 응답적격여부가 알려지지 않은 경우</p> <ul style="list-style-type: none">- 기타, 슈퍼바이저와의 상의(other, referred to supervisor)

자료: AAPOR(2006).

나. 응답자 접촉 및 협조유도 업무 산출물 평가

응답자 접촉 및 협조유도 업무의 산출물에 대한 평가는 무응답 오차에 대한 분석을 통해서 이루어진다. 무응답 오차는 표본의 대표성과 관련하여 의미를 갖는다. 낮은 응답률이 자료의 품질이 미흡한 것으로 인식되는 경향이 있기 때문이다.⁴⁾ 이로 인해 통계작성기관에서는 예산의

의 경우에 이러한 유형의 응답이 종종 발생한다. 이 방법은 원칙적으로 제한되기는 하지만 실제 현장조사를 하다 보면, 응답을 얻을 수 없는 경우에 참값에 가장 근접한 응답을 얻어내기 위해 사용되는 방법이기도 하다. 중요한 것은 응답을 얻어냄에 있어서 대리인으로부터 정보를 얻었는가를 보고해야 한다는 것이다(AAPOR, 2006).

4) 이러한 이유로 인해서 어떤 연구자들은 자신들의 응답률을 인위적으로 부풀리고자 (artificially inflate) 하였으며, 이에 대한 위험성이 이미 오래 전에 경고된 바 있다(Kviz, 1977).

많은 부분을 응답률을 높이는 데 할애하고 있다. 조사원으로 하여금 응답거부표본에 더 많이 방문하도록 함으로써 응답 거부자를 전환시키거나 응답자 혹은 조사원에게 현금과 같은 인센티브를 제공함으로써 무응답을 감소시키고자 하는 노력을 하고 있다(Curtin, Presser, and Singer, 2002).

그런데 최근에는 무응답에 대한 입장이 변화되고 있다. ESS(the European Social Survey)에서는 이 조사에 참여하는 국가들에 대한 기대 응답률을 70% 정도로 보고 있다(Loosveldt, Carton, and Billiet, 2004). 이는 현장조사의 어려움으로 인해 일정정도의 무응답 발생을 허용하는 사회적인 분위기를 반영하기도 하며, 낮은 응답률이 반드시 낮은 자료의 품질은 아니라는 최근의 연구경향을 보여주는 것이기도 하다. 미국의 NORC(National Opinion Research Center, 이하 NORC)에서 CAPI(Computer Assisted Personal Interviewing, 이하 CAPI) 방식으로 수행되는 조사에서의 모니터링 평가양식 또한 이를 잘 보여준다. 과거에는 응답거부를 협조로 전환했는지, 그 과정에서의 거부전환이 잘 시도되었는지가 중요한 판단기준이었다면, 최근의 변경된 기준에서는 거부전환이 상황에 맞고 그 내용은 사실적이었는지, 그리고 불가피한 상황에서는 응답거부가 용인되었는지가 중심이 되었다. 변화된 기준을 통해서 알 수 있는 것은 기존에는 무응답이 줄어든 결과 자체가 중요하게 평가되었다면, 최근에는 응답을 거부한 상황에서 조사원이 응답을 얻기 위해 설득하는 과정이 얼마나 합리적인가를 평가하는 방향으로 전환되었음을 알 수 있다.

무응답 오차에 대한 정확한 평가를 위해서는 응답자 집단과 무응답자 집단의 비교가 함께 이루어져야 한다. 만일 응답자와 무응답자 간에 유의미한 차이가 발견되지 않는다면, 무응답 비율과 무관하게 무응답 오차는 발생하지 않을 것이다. 하지만 무응답자에 대한 정보는 일반적으로 접근하기가 어렵다. 그러므로 조사 설계 단계에서부터 가능한 한 많은 정보를 구하고자 노력을 해야 하는데 이는 방문기록표를 통해서 어느 정도 보완될 수 있다. 조사에 협조하지 않는 가구의 주거환경이나 생활상태 및 유형 등에 관한 정보를 방문기록표에 기입하도록 할 수 있다(Loosveldt, Carton, and Billiet, 2004). NORC에서 수행하고 있는 GSS(General Social Survey)에서는 방문과정에 대한 기록과 함께 면접을

완료하지 못한 이유⁵⁾를 적게 하고, 만일 응답자가 응답을 거부하였을 경우 거부한 구체적인 이유와 함께 거부자의 성별, 연령(추정), 기타 특이 사항 등을 기록하게 함으로써 이들의 특성을 파악하고자 시도하고 있다.

무응답의 정의에 따라서 응답률은 다양한 방식으로 계산될 수 있다. AAPOR에서는 완성된 면접(Complete Interview, 이하 I)과 부분 면접(Partial Interview, 이하 P), 거부(Refusal and Break-off, 이하 R), 비접촉(Non-Contact, 이하 NC), 기타(Other, 이하 O), 응답적격여부가 알려지지 않은 가구원(Unknown if household/occupied, 이하 HU), 응답적격여부가 알려지지 않은 기타 성원(Unknown, Other, 이하 UO), 결정되지 않은 응답적격여부 사례에 대한 추정량(Estimated Proportion of cases of unknown eligibility that are eligible, 이하 e)을 구분하여 응답률을 계산하고 있다. 응답률에 대한 정의는 완성된 면접의 수를 응답가능한 표본수로 나눈 것으로,⁶⁾ 부분면접을 어떻게 다룰 것인지, 결정되지 않은 응답적격여부 사례에 대한 추정량이 얼마인지 등에 따라서 여섯 가지 계산방식이 도출될 수 있다(표 3-3 참고).

$$RR(\text{response rate}) = C/E$$

여기에서 C 는 완성된 면접 수, E 는 응답가능한 표본 수

응답률에 대한 가장 전형적인 정의($RR1$)는 완성된 면접수를, 면접을 한 경우(완성된 면접+부분면접)와 면접을 하지 않은 경우(거절+비접촉+기타), 그리고 응답적격여부가 알려지지 않은 경우(가구원+기타 성원)를 합산한 값으로 나누어 준 것이다.

5) 면접을 완료하지 못한 이유로는 응답자가 응답을 거부해서, 응답자를 찾을 수가 없어서, 응답자가 응답할 상황이 아니어서(질병, 고령 등), 응답자가 시간이 없어서, 기타 등이 있다.

6) 응답률과 연관된 개념으로 완성률(completion rate)이 있다. 완성률은 전체 표본에서 완성된 면접의 비율을 의미한다. 분모에 응답가능한 표본과 함께 응답부적격 표본도 포함된다는 점에서 응답률과 차이를 보인다.

$$CR(\text{completion rate}) = C/n$$

여기에서 C 는 완성된 면접의 수, n 은 표본 크기

이 지표는 특정의 모집단으로부터 완성된 면접의 수를 획득하기 위해서 추출해야 하는 표집단위가 얼마인가를 결정하는 데 도움을 준다. 이는 현장운명을 계획하고, 일정을 잡고, 예산을 세우는 데 필수적인 정보이다(Kviz, 1977).

〈표 3-3〉 응답률 계산 공식

$$\begin{aligned}
 RR1 &= \frac{I}{(I+P) + (R+NC+O) + (UH+UO)} \\
 RR2 &= \frac{(I+P)}{(I+P) + (R+NC+O) + (UH+UO)} \\
 RR3 &= \frac{I}{(I+P) + (R+NC+O) + e(UH+UO)} \\
 RR4 &= \frac{(I+P)}{(I+P) + (R+NC+O) + e(UH+UO)} \\
 RR5 &= \frac{I}{(I+P) + (R+NC+O)} \\
 RR6 &= \frac{(I+P)}{(I+P) + (R+NC+O)}
 \end{aligned}$$

두 번째($RR2$)는 첫 번째 공식의 분모에 부분면접을 포함한 것이다. 세 번째($RR3$)는 응답적격여부가 알려지지 않은 사례에 그 정도가 얼마만큼인지를 추정할 값을 곱해줌으로써 응답률을 계산하는 것이며, 네 번째($RR4$)는 세 번째 산식의 분모에 부분면접을 포함한 것이다. 다섯 번째($RR5$)는 결정되지 않은 응답적격사례의 추정량을 0으로 가정한 것으로, 알려지지 않은 사례 중에서 응답적격한 사례가 없다는 것이 타당한 경우에 가능한 공식이다. 여섯 번째($RR6$)는 다섯 번째의 산식의 분모에 부분면접을 더해 준 것으로 응답률이 가장 높게 계산된다.

최근에는 응답률과 함께 협조율(cooperation rates), 거부율(refusal rates), 접촉률(contact rates) 등의 수치를 함께 보고하기도 한다. 이는 조사원 업무수행뿐만 아니라 현장운영의 전반적인 양상에 대해서도 파악할 수 있도록 해주는 지표이다.

다. 면접업무 과정평가

면접업무에 대한 과정평가는 조사원이 교육받은 내용을 얼마나 잘

이행하는지를 중심으로 이루어진다. 통계작성기관은 조사원 오차를 최소화하기 위해 질문전달, 캐어묻기 행위, 응답자에 대한 피드백, 그리고 속도와 명확성 등에 관해서 교육을 한다(Mathiowetz and Cannell, 1980). 그리고 이러한 교육내용을 면접과정에서 잘 수행하는지를 평가한다. 질문전달의 경우 질문을 쓰인 그대로 읽는지 혹은 중요하지 않은 표현을 변경하는지, 캐어묻기의 경우 적절하게 잘 캐어묻는지, 피드백의 경우 피드백을 적절하게 전달하는지, 속도 조절의 경우 너무 빨리 혹은 느리게 읽는지, 항목 사이의 시간간격은 적절한지 등을 평가한다.

예를 들어, 면접시간 분석은 면접 과정을 평가할 수 있는 측정도구 중 하나가 될 수 있다. 만일 조사원들 사이에 면접시간의 차이가 난다면 이는 어떤 조사원은 면접을 너무 빠르게, 혹은 너무 느리게 진행하였기 때문으로 추정해 볼 수 있다. 그러나 이러한 분석은 오직 간접적인 측정일 뿐 직접적인 측정은 조사원과 응답자 간의 상호작용에 대한 평가, 특히 그 과정에서의 조사원 행동에 대한 평가를 통해서 이루어져야 한다 (Loosveldt, Carton, and Billiet, 2004).

라. 면접업무 산출물 평가

측정오차는 잘못된 기록에 의한 응답, 응답자, 조사원, 자료수집방법, 조사표 등 다양한 원인에 의해 발생할 수 있다. 이 중 조사원 행동에 주목하여 보면, 조사원은 표준화된 지침에서 벗어남으로써 조사원 오차를 발생시킨다. 예컨대, 조사원이 질문을 표현대로 읽지 않았거나, 적절하게 캐어묻기를 하지 않았거나 혹은 응답자에게 적당한 피드백을 제공하지 않았거나, 응답자의 질문에 정확히 답하지 않았거나, 응답들을 올바르게 기록하지 않아서 발생하는 숨겨진 오류들이 항상 내포되어 있을 수 있다(Bierner and Lyberg, 2003). 물론 이러한 효과가 조사원들 사이에서 상쇄되어 전체적으로는 작은 오차만을 낳을 수도 있으나 개별 조사원의 오차는 클 수 있으며, 이는 전반적으로 추정의 정확성에 심각한 영향을 미칠 수도 있다는 점을 간과해서는 안 된다(UN, 2005).

조사원 오차는 다수준분석(multi-level analysis)을 통해서 측정될 수 있다. 이 분석에서는 일반적으로 응답자의 특성이 1수준에서, 조사원의

특성이 2수준에서 고려된다. 조사원 사이의 차이가 응답자의 특성이 통제된 아래에서 분석됨으로써 조사원이 응답에 미치는 영향을 분석할 수 있으며(Pickery and Loosveldt, 2002 등), 최근에는 지역의 특성을 3수준에서 고려한 분석이 이루어짐으로써 조사원 효과와 독립적으로 지역이 미치는 효과를 검증하는 연구들이 시도되고 있기도 하다(Campanelli and O'Muircheartaigh, 1999 등).

지금까지 조사원 업무성과 측정을 위한 개념적인 틀에 대해서 살펴 보았다. 조사원 업무성과는 응답자 접촉 및 협조 유도 및 면접이라는 두 가지 업무를 대상으로 하여 그 과정과 산출물에 대한 평가를 중심으로 이루어진다. 조사원을 직접적으로 관찰하는 일이 어렵기 때문에 과정에 대한 평가보다는 연구의 대부분이 수집된 자료를 바탕으로 조사원 효과를 추정하는 데 치중되어 온 경향이 있어 왔다. 그런데 최근 들어 모니터링 기법이 발달하면서 조사원 업무성과 중 과정에 대한 평가를 가능케 해주는 다양한 정보의 수집이 용이해지고 있다.

2. 측정방법

조사원 업무성과 측정을 위해서는 조사원에 대한 모니터링 시스템 체계가 구축되어 있어야 한다. 모니터링은 특히 조사원 업무에서 과정에 대한 평가를 위한 핵심적인 틀이 된다. 모니터링의 방법은 모니터링의 시기나 조사방법 등에 따라서 구별될 수 있다. 면접을 수행하는 과정 중의 모니터링 방법으로는 통화모니터링이나 감독자 관측이, 면접을 완료한 이후에는 리인터뷰 조사(reinterview survey)와 재접촉 확인(verification recontacts) 방법 등이 있다(Biemer and Lyberg, 2003).

가. 면접과정 모니터링

면접과정 중의 모니터링 방법은 조사원이 질문을 할 때 조사표의 질문을 어느 정도 변경하는지, 입력 오류는 없는지, 전문적이지 않은 행동을 취하는지, 적절하지 않은 캐어문기를 하였는지, 피드백이 중립적이지 않은지 등에 대한 측정을 하는 것이다.

면접이 진행되는 과정 중에 수행되는 모니터링 시스템은 전화조사와 면접조사에 따라서 그 방법에 차이가 있다. 중앙집중식 전화조사에서 사용되는 통화모니터링은 조사원을 감독하고 조사원 오류를 관리하기 위해 널리 사용되는 방법이다. 조사원은 보통 면접을 시작할 때 응답자에게 조사원이 조사를 바르게 하는지 확인하기 위해 감독자가 통화내용을 들을 수 있다고 알리는 것으로 면접을 시작하며, 면접하는 내용 중 일부가 모니터링된다. 그렇지만 정확히 어느 부분에서 통화내용이 모니터링되는지는 응답자와 조사원 모두 알지 못한다. 모니터링 과정에서 조사원의 오류가 발생하는 즉시 면접이 중단되고 해당 조사원에 대한 교육이 실시된다. 따라서 적절한 모니터링은 조사원과 관련된 오류를 감소시켜 줄 수 있다(Couper, Holand, and Groves, 1992). 면접에 대한 통화모니터링의 전형적인 비율은 외국의 경우에는 매 10시간의 면접당 1시간 정도이며, 우리나라의 경우 통계작성기관 및 조사업무에 따라서 차이가 있기는 하지만 전체 조사 건수 중 최소 10~30% 가량을 무작위 추출하여 모니터링을 한다.

면접조사에서는 전통적으로 테이프 녹음과 감독자 관측이라는 두 가지 방법으로 면접관측을 실시해 오고 있다. 첫 번째 방법은 조사원이 면접의 일부나 전부를 테이프로 녹음하여 나중에 그 녹음테이프를 감독자가 검토할 수 있게 하는 것이며, 두 번째 방법은 감독자가 그에게 배정된 모두 또는 일부 조사원과 동행하여 면접지침 준수와 관련된 사항을 메모하면서 이들의 업무수행을 기록하고 평가하는 것이다(Lepkowski, Sadosky, and Weiss, 1998). 이러한 방법은 조사원의 주요 언행에 대해서 코드를 부여하고 분석하는 행동코딩(behavior coding) 방법의 발전과 궤를 같이한다. 최근에는 컴퓨터를 보조적으로 이용한 면접기법이 발전함에 따라서 면접관측기법이 확대될 것으로 기대하고 있다(Dijkstra, Ongena, and Smit, 2007).⁷⁾ 조사원이 휴대한 노트북 컴퓨터에 장착된 마이크를 이용해서 조사원과 응답자에게 드러나지 않도록 면접내용을 하드 디스크에 디지털 방식으로 녹음하는 것이 가능해졌기 때문이다

7) 면접관측법은 현실적인 적용가능성이라는 한계에 부딪혀 있다. 예컨대, 조사원과 응답자는 조사의 전체 과정이 관측되고 있다는 것을 알게 되는데 이런 상황이 조사에 방해요인으로 작용할 수 있다. 또한 감독자는 조사원을 관측하기 위해 때로는 상당히 먼 거리를 동행해야 하는 경우가 생기는데, 이는 국가 차원의 대형조사에서는 많은 비용과 시간을 필요로 한다(Bierner and Lyberg, 2003).

(Computer Audio Recorded Interviewing, 이하 CARI). 이것은 침묵 모니터링(silent monitoring)이라고 불리고 있으며, 미국 센서스국에서는 품질관리의 새로운 기법 중 하나로 CARI에 대한 시험조사를 실시하였다. 이 방법은 응답자나 조사원 모두에게 모니터링 부담(monitoring response burden)을 감소시켜 주며, 감독자 관측보다 객관적이고, 양적으로 또한 풍부한 접근을 할 수 있다는 장점을 지니는 것으로 나타났다(Census Bureau, 2007). 우리나라의 경우, 이러한 조사기법은 아직 확산되어 있지 않으나, 패널조사에서 이 방법을 이용하고 있는 한 통계작성기관의 설명에 의하면 아직까지는 면접 과정에 대한 내용분석이 진행되고 있지는 않은 수준이다.

나. 면접 후 모니터링

면접 후 모니터링은 조사원이 임의로 자료를 조작하였는지를 평가할 수 있는 틀을 제공해 줌으로써 조사원 업무성과를 측정하는 데 도움을 준다. 구체적으로, 면접 후 모니터링으로 많이 활용되고 있는 리인터뷰 조사와 재접촉 확인 방법에 대해서 살펴보도록 하자. 리인터뷰 조사는 응답자 중에서 비교적 작은 확률표본을 선택해서 최초 면접 후 짧은 기간 내에 그들과 접촉하여 최초 면접에 포함되었던 것과 같은 주제들에 대해 다시 면접하는 것을 말한다. 일반적으로 리인터뷰 조사는 두 가지 목적을 갖는다. 첫째는 현장조사 업무를 평가하는 것으로 조사원이 고의적으로 자료를 조작하거나 조사 지침을 잘못 습득하여 보충훈련이 필요한지를 평가하는 것이며, 둘째는 단순응답분산과 응답편의 등 조사모델에서의 오차요소를 추정하는 것이다. 따라서 리인터뷰 조사는 목적에 따라 표본추출방법과 조사원선정, 응답자선정, 조사표설계, 자료수정여부 결정, 조사방법 등을 고려해서 설계해야 한다. 이 중 조사원의 업무성과 평가를 위한 리인터뷰인 경우 2단계, 즉 조사원을 1차로, 응답자를 2차로 하여 표본을 선정할 수 있다(통계청, 2003). 질문은 크게 공통항목과 특성항목으로 구분된다. 공통항목은 조사현장의 실태를 파악하기 위해 모든 조사통계에 동일하게 적용되는 것으로 응답자명, 성별, 적용한 자료수집 방법, 지침 이외의 자료수집 방법 적용 이유, 조사 소요시간 등이다. 특성항목은 응답내용의 일치 정도를 점검하기 위한 것으로 통

계마다 다르게 선정되는데, 일반적으로 응답자에게 직접 묻지 않고는 정확한 조사가 불가능하다고 판단되는 2개의 조사항목을 선정한다. 미국 센서스국은 1950년대부터 대규모의 경상통계조사와 센서스에서 조사원의 업무수행을 평가하는데 이 방법을 사용해 오고 있다(Biemer and Forsman, 1992). 반면, 우리나라에서 리인터뷰 기법은 아직 일반적이지는 않은 편이다. 통계청의 경우 국가통계 품질관리의 일환으로 한시적으로 리인터뷰 조사를 도입한 바 있다.

재접촉 확인은 통계작성기관에서 가장 널리 사용되는 방법으로 표본의 10% 정도를 추출한 후 하위 표본에 전화를 걸어서 적절한 응답자에게 면접을 수행하였는지를 확인함으로써 조사원의 임의작성여부를 확인하는 것이다. 이는 우리나라의 많은 민간 통계작성기관에서 자료의 품질관리나 조사원 업무성과를 파악하기 위해 흔히 사용되고 있는 방법이다. 외국과 달리 우리나라 통계작성기관에서는 조사된 모든 표본에 대해서 100% 접촉 확인을 시도하고 있으며, 만일 전화번호가 결번이거나 받지 않는 등의 이유로 통화가 되지 않을 경우에는 추가적인 방문확인조사를 실시한다. 혹은 오류로 간주하여 조사된 표본에서 삭제하는 경우도 있다. 물론 응답자가 고의로 전화번호를 잘못 알려주는 경우도 있을 수 있으나 이것은 인정되지 않으며, 응답자로부터 정확한 전화번호를 받는 것은 조사원이 이행해야 할 의무로 계약내용에 포함되어 있다. 이러한 방법들을 통해서 조사원 업무성과 측정을 위한 정보를 수집하고 이를 이용하여 각종 지표들을 만들 수 있다.

3. 측정내용

앞에서 살펴보았듯이 조사원 업무성과는 두 가지 주요한 업무에 대한 과정과 산출물을 중심으로 평가되어야 한다. 초기에 조사원의 업무성과는 완성된 업무량의 비중과 같은 비교적 간단한 도구로 측정되었으며, 이는 일반적으로 응답률이라는 지표로 계산되었다(Turkow, Bailey, and Stamper, 2000). 그러나 응답률 그 자체만으로 업무성과를 측정하는데에는 한계가 있어 왔다. 첫째, 지난 30여 년 동안 응답률에 대한 계산 방법은 거의 연구되지 않았으며 최근에는 AAPOR 등을 비롯한 기관에

서 그 기준을 정립하였다(Mccarty, 2003)⁸⁾. 이에 따르면 응답률 계산방법은 무응답을 어떻게 처리하는지에 따라서 정도가 달라질 수 있기 때문에 각 통계작성기관에서 응답률을 제시할 때에는 그 기준을 보여주는 것이 필요하다. 둘째, 조사환경이 점차 어려워지는 상황에서 좋은 업무 성과로서 높은 응답률을 강조하는 것은 조사원으로 하여금 조사완성에만 집착하도록 하여 수집되는 자료를 부실하게 만들 우려가 있다. 셋째, 고립된 단일지표는 잘못된 해석을 낳을 수 있는 위험이 있다. 두 번째 이유와 연관지어서 생각해 본다면, 높은 응답률이 반드시 양질의 자료를 의미하는 것이 아니기 때문에 조사원의 업무성과를 측정함에 있어서 다양한 지표를 동시에 고려하는 안목이 필요하다. 특히 산출물에 대한 평가뿐만 아니라 과정에 대한 평가가 동시에 이루어져야 하므로, 응답률과 함께 이를 보완해 줄 수 있는 다양한 지표의 개발을 통해서 조사원들의 업무성과를 측정해야 할 필요성이 제기되었다.

Tarnai & Moore(2006)는 161개 통계작성기관들을 대상으로 전화조사원의 업무성과 측정내용을 조사하였다.⁹⁾ 물론 전화조사원과 면접조사원의 업무성과 측정내용에는 차이가 있을 수 있으나 최근에는 점차 유사한 측정도구를 사용하는 경향이 나타나고 있다(Durand, 2005). 조사원 업무성과 측정에 관한 초기의 연구에서는 중앙집중식 시설에서 이루어지는 전화조사원을 대상으로 한 경우가 많이 있다. 왜냐하면 표본(전화번호)은 개별 조사원의 특성이 아닌 접근가능성(availability)에 따라서 무작위적으로 할당되므로 모든 조사원은 유사한 업무를 수행하고 있다고 가정되기 때문이다. 조사원 업무성과를 측정할 때 가장 어려운 점 중의 하나가 조사원이 할당받은 업무와 그 업무가 놓여 있는 상황이 다르기 때문에 업무성과를 객관적으로 측정하는 것이 쉽지 않다는 점이다. 이

8) 응답률의 정의 및 측정방법 등에 대한 합의를 찾기까지 많은 시간과 노력이 들었다. 응답률에 대한 개념을 정립하기 위한 초기의 노력은 CASRO(the Council of American Survey Research Organization)에서 발행된 "On the Definition of Response Rates"에서 출발하였으며, 이를 기반으로 하여 AAPOR가 몇 차례의 개정을 거듭한 끝에 응답률과 함께 여러 지표를 계산하는 표준화된 산식을 개발하였다(AAPOR, 2006).

9) 조사대상으로는 AAPOR 회원국과 2005 International Field Directors Technology Conference에 참여한 490개 기관이며, 이 중 161개 기관이 응답하였다. 기관의 특성을 보면, 학술기관이 34.8%, 국가기관이 4.3%, 비영리기관이 9.3%, 상업기관이 44.7%였다.

런 면에서 볼 때 중앙집중식 시설의 전화조사에서는 환경적인 요인에 대한 통제가 가능하여 업무성과를 측정함에 있어서 상대적으로 수월하다(표 3-4 참고).

〈표 3-4〉 조사원 업무성과 측정내용과 활용정도

(단위: %)

조사원 업무성과 측정내용	통계작성기관에서의 활용정도
전화시도 횟수	93.5
완성된 조사수	97.6
거부된 조사수	89.3
응답부적격 조사수	73.3
업무 시간	92.6
면접 시간	90.0
근무태도	75.2
슈퍼바이저의 평가	81.1
모니터링 점수	66.4
항목무응답 수	35.2

자료: Tarnai & Moore(2006).

Tarnai & Moore(2006)의 조사 결과, 조사원 업무성과를 측정함에 있어서 응답률을 계산하는 데 필요한 완성된 조사표의 수 이외에도 다양한 지표들이 사용되고 있는 것으로 나타났다.

측정내용으로는 전화시도 횟수, 완성된 조사수, 거부된 조사수, 응답 부적격 조사수 등이 있다. 이러한 정보는 응답률, 협조율, 거부율, 접촉률 등을 계산할 수 있는 근거자료로서 응답자 접촉 및 협조유도 업무 과정에 대한 평가 지표들로 많이 활용되고 있다. 좀 더 구체적으로 호주 통계청의 조사업무성과 매트릭스(survey performance metrics)에서는 다양한 응답률을 그 지표로 사용하고 있었다. 가장 보편적으로 사용되는 것은 최종 응답률이며, 이와 함께 마감기간 내의 응답률, 자료수집방법에 따른 응답률 등이 구분되어서 측정되고 있었다. 영국 통계청 가구조사

의 경우 응답률 및 면접시간과 함께 응답자 접촉 및 협조유도 업무에 대한 과정평가에서 핵심적인 항목 중의 하나인 방문패턴을 주요한 측정 지표의 하나로 활용하고 있는 것으로 나타났다(Full, Haworth, and Stephens, 2001).

또한 전체 업무시간, 면접시간 등이 고려되었으며, 근무태도, 슈퍼바이저의 평가, 모니터링 점수 및 항목무응답 수 등도 업무성과를 측정하는 지표로 활용되고 있다. 이는 최근의 모니터링 기법의 발달과 밀접히 연관되어 있다. 예를 들면 면접시간의 경우, CATI의 키스트록 파일이나 로그파일 내의 정보를 통해서 수집 가능한 것이다.

조사원 업무성과를 측정하는 여러 가지 측정내용 중에서도 완성된 조사수(97.6%), 전화시도 횟수(93.5%), 업무시간(92.6%) 등을 활용하는 기관이 가장 많았으며, 조사원의 면접과정에 대한 모니터링을 통해 측정가능한 지표인 근무태도, 슈퍼바이저의 평가, 모니터링 점수는 활용도 면에서는 그보다는 약간 낮은 것으로 드러났으나 여전히 많은 통계작성 기관에서 업무성과를 측정하기 위한 정보로 활용하고 있었다.

제3절 우리나라 통계청의 조사원 업무성과 측정 현황

1. 조사원 제도 개요

통계청에서는 일반직, 별정직, 계약직, 도급직, 파트타임직 등 다양한 직급을 가진 조사원이 현장조사를 수행하고 있다. 2006년 12월 1일 현재 통계청 지방사무소 조사원 현황을 살펴보면, 도급직 조사원이 35.9%로 가장 많은 비중을 차지하고 있으며, 일반직 조사원이 31.4%인 것으로 나타났다. 이 중 일반직, 별정직, 계약직은 정규직으로, 도급직과 파트타임직은 임시직으로 분류될 수 있다(표 3-5 참고).

도급직과 파트타임직의 성격에 대해서 잠깐 언급해 보면 다음과 같다. 통계청에서는 2003년부터 표본개편으로 인해 증가한 가구 및 사업체 부문의 조사업무를 정상적으로 수행하기 위해 경상조사 임시조사원을 채용하였다. 그러나 임시조사원 활용이 지속적으로 증가하여 이들에

〈표 3-5〉 직급별 조사원 현황

(단위: 명, %)

계	일반직	별정직	계약직	도급직	파트타임직
1,485	466	291	158	533	37
(100.0)	(31.4)	(19.6)	(10.6)	(35.9)	(2.5)

대한 탄력적인 운영이 요구됨에 따라 2005년부터 경상조사임시조사원을 도급임시조사원으로 운영하고 있다. 도급조사원의 업무수행은 기존의 임시조사원과 크게 다르지 않으나, 계약의 형태에서 차이를 보인다. 도급조사원은 자신에게 배정된 조사업무 및 조사관련 업무를 완성할 것을 약정하고 지방통계청장이 업무수행 결과에 대해 보수를 지급할 것을 약정하는 방식으로 계약을 하는 일종의 사업자 성격을 띤다. 이들은 특별히 지정하는 일시 이외는 자유의사에 따라서 출퇴근을 하는 등 독립성과 자주성이 보장되어 있다는 점에서 경상조사 임시조사원과는 차이가 있다. 한편, 파트타임조사원은 정규조사원의 결원, 출산휴가 퇴직예정, 장기병가, 육아휴직 및 도급조사원의 출산휴가 등에 따른 대처방안으로 지원되고 있다.

조사원의 업무량은 직급별로 다르게 배정되어 있다. 우선 도급조사원의 경우에는 지방통계청을 통하여 실시하는 각종 통계조사의 현장조사, 조사표의 내용검토 및 자료입력, 모집단 관리 등 통계조사 관련 업무를 맡고 있다. 정규조사원의 경우에도 통계조사와 관련해서는 도급조사원과 유사한 업무량을 배정받고 있으며, 여기에 부가적으로 각종 전산업무와 행정업무, 그리고 조사팀제 운영 원칙 등에 따라서 도급조사원이 수행해 온 조사표에 대한 내용검토를 추가로 실시하고 있다.

2. 측정방법

통계청에서 조사원의 업무성과 측정을 위해 실시하고 있는 모니터링 시스템에는 세 종류가 있다. 가장 많이 활용되는 방법은 내용검토이다. 이 방법은 조사원의 월간조사업무량 및 관련업무량의 10% 이상을 선정하여 성실히 수행하였는지 내용검토를 한 후, 업무별 가감점 기준표에

따라서 개인별 평가를 실시하는 것이다. 부실조사가 발견될 경우 조사원의 업무 전체에 대해서 재점검을 하고 해당 조사원에 대해서는 현지 확인을 실시한다.¹⁰⁾

통계청에서는 또한 사업체 및 가구부문 경상조사 담당조사원을 대상으로 1인당 연간 1회 이상 자체 현지 확인점검을 실시하고 있다. 이는 현장조사를 담당하는 조사원에게 긴장감을 고조시키고 부실한 자료가 수집되는 것을 사전에 방지하여 통계품질을 향상시키기 위해 고안된 것이다. 그러나 개인별로 연간 1회 정도 점검을 받을 뿐이며, 한번 점검을 받은 직원은 이후에는 점검대상자에서 제외되기 때문에 이 제도가 조사원에게 실제적인 통제력을 발휘하기는 쉽지 않다. 확인점검시 사업체 부문은 사업체관리를 포함한 전체내용, 가구부문도 가구관리실태를 포함한 전체내용을 동시에 점검하기 때문에 효율적이고 실질적인 점검이 이루어지지 못하며, 오히려 조사담당자 및 응답자의 부담만 가중된다는 평가를 받고 있다.

또 다른 모니터링 방법으로는 리인터뷰 조사가 있다. 통계청에서는 품질관리의 일환으로 2002년부터 3년간 리인터뷰 조사를 실시한 바 있다. 이 방법은 응답자를 대상으로 점검을 실시함으로써 현장조사의 오류 유형과 발생 원인에 대한 실증자료를 축적하여 조사품질을 개선시키고자 하는 목적을 갖고 있다. 이러한 기능이 작동되면 조사원의 조사수행 성실도가 높아지고 현장조사 단계에서 품질관리에 대한 중요성이 인식되어 지속적으로 조사업무 개선을 유도할 수 있다(통계청, 2007). 통계청에서는 2002년부터 3년간 전화점검을 실시하였다. 연도별로 보면 2002년에 광공업동태조사 등 8종, 2003년에 건설업통계조사 등 9종, 2004년에 소비자물가조사 등 7종을 점검하였다. 매년 서로 다른 조사에 대해서 리인터뷰를 수행하고 있으므로 앞의 현장 확인점검과 마찬가지로 리인터뷰를 한번 실시한 경우 특정 시점까지 다시 조사를 하지 않는다는 한계를 가지고 있다.

통계청의 모니터링 시스템의 특징을 요약해 보면 다음과 같다. 첫째, 모니터링을 실시하는 시기가 부정기적이다. 내용검토 방법을 제외하고

10) 도급조사원 운영지침 제정 통계청훈령 제126호(2005.12.27).

현지 확인점검이나 리인터뷰 조사의 경우 점검의 주기가 부정기적이며 매우 길다는 특징을 갖고 있다. 따라서 정기적으로 조사원의 업무성과를 측정할 수 없으며, 모니터링시스템이 이들의 업무성과를 향상시키는 데에도 영향을 미치기가 어렵다. 둘째, 응답자로부터의 점검 시스템이 취약하다. 응답자에게 직접 면접결과를 확인하는 시스템은 조사원의 부적절한 행동을 사전에 예방할 수 있으나, 통계청 모니터링 시스템 중 이 부분을 보완해 줄 수 있는 제도는 부족하다. 셋째, 조사원이 면접과정에서 지침을 얼마나 잘 이행하는가에 대한 평가를 할 수 있는 과정에 대한 모니터링 시스템이 부재하며, 따라서 조사원이 발생시킬 수 있는 숨겨진 오차에 대한 평가를 하는 것이 쉽지 않다.

이러한 한계는 통계청 조사체계의 특성과 연관되어 있다. 통계청에서는 다양한 유형의 조사가 실시된다. 경제활동인구조사, 소비자물가조사, 광공업동태조사 등 매월 혹은 분기별로 실시하는 경상조사와 인구주택총조사, 광업제조업통계조사 등의 특별조사나 연간조사 등이 있다. 경상조사의 경우 연동표본을 이용함으로써 고정된 응답자에게 매번 동일한 질문을 한다는 특징을 갖고 있다. 따라서 이러한 조사체계 하에서 조사원의 성실한 조사수행여부를 응답자로부터 점검하거나 면접과정에서 모니터링을 하는 것은 응답자에게 조사원에 대한 불신을 심어줄 개연성이 있다. 그러므로 통계청의 이러한 특성을 반영한 모니터링 시스템에 대한 개선이 요구된다.

3. 측정내용과 사례분석

주요 통계작성기관에서 조사원 업무성과 측정을 위해서 완성된 조사수나 거부된 조사수 및 면접시간 등을 많이 활용하고 있는 것과 다르게, 통계청에서는 조사표에 대한 내용검토를 한 후 작성된 개인별 업무수행평가서와 사후심사평가서 두 가지 양식을 사용하고 있다.

개인별 업무수행평가서는 업무별 가감점 기준에 따라서 작성된다. 업무별 가감점 기준표에는 조사유형별 착오유형에 따른 감점점수가 나와 있는데, 모든 조사에서 공통적인 감점항목은 다음과 같다. 조사표 임의 또는 허위조사, 사업체(가구) 3개월 이상 미방문조사, 분류착오(산업,

가구, 활동상태 등), 단위착오, 사업체(가구) 변동보고 지연(3개월이상), 표본사업체(가구) 대체지연(3개월 이상), 조사표 입력 착오 및 누락, 조사표 입력 지연, 기타 경미한 사항 등이며, 이 중 조사표 임의 또는 허위 조사에 대한 감점점수가 10점으로 가장 높으며, 3개월 이상 미방문 조사의 경우는 7점, 분류착오는 5점이다. 이와 함께 각 조사별로 구체적인 감점기준이 제시되어 있다. 가구조사 중 가계조사의 경우 가구실태항목 착오, 품목 누락, 은행자동이체품목 누락, 월간 가계수지 총괄집계 착오, 가계수지항목 분류부호 착오, 주요품목 금액착오 여부 등에 따라 감점이 된다. 사업체 조사 중 광공업 동태조사의 경우에는 조사품목 누락, 조사품목 포괄범위 조사착오, 주요조사항목착오(생산, 출하, 내수, 수출, 재고, 등), 동일공장 재투입여부 조사착오, 위탁·수탁여부 조사착오 등이 주요 감점 기준이다. 한편, 현장조사 업무와 관련해서 지방통계청(사무소)장 이상 표창을 받은 경우와 업무개선안을 제출하여 팀장(과장)으로부터 인정받은 경우 등은 가점을 받게 된다.

사후심사평가서는 조사원의 조사업무에 대한 이해도, 업무수행태도 및 성실성, 지시사항 이행여부의 세 가지 영역으로 구성되어 있으며, 출장소장, 담당팀장, 담당지도공무원에 의해 평가된다. 각 영역은 5점척도로 측정되며, 세 영역의 점수를 합하여 100점 만점(4:3:3)을 기준으로 80점 이상을 받았을 경우 우수(A등급), 51~79점인 경우에는 보통(B등급), 50점 이하인 경우는 미흡(C등급)한 것으로 결정된다. 이 결과는 추후에 재계약을 할 때 중요하게 작용을 하는데, 만일 최근에 C등급을 받은 경우는 사후평가일로부터 1년 이내에는 조사원 채용을 위한 사전심사평가 대상에서 제외된다.

통계청의 조사원 업무성과 측정내용은 앞에서의 개념적인 틀에 비추어 볼 때 다음과 같은 한계를 갖는다. 첫째, 조사원의 두 가지 주요 업무 중 응답자 접촉 및 협조유도 업무에 대한 평가보다는 면접 업무에 대한 평가를 중심으로 업무성과가 측정되고 있다. 둘째, 과정에 대한 평가보다 산출물에 대한 평가가 이루어지고 있다. 면접과정에서의 조사원 행동에 대한 평가는 크게 질문전달, 프로빙 행위, 응답자에 대한 피드백, 그리고 속도와 명확성 등에 대해서 이루어지는데, 통계청에서는 면접이 완료된 이후에 조사표를 바탕으로 착오사항을 점검한다. 특히 개인별

업무수행평가서의 경우 내용검토 결과 밖으로 표출되는 문제점만을 가지고 평가를 하므로, 면접과정 중에 발생하는 눈에 보이지 않는 오류에 대해서는 지침을 제공해 주지 못한다.

조사원 업무성과 측정내용이 갖고 있는 한계점을 통해서 통계청의 현황에 대한 개선방향을 제안할 수 있을 것이다. 좀 더 구체적인 한계를 파악해 보기 위해 모니터링 시스템 활동 사례에 대한 분석을 해 보고자 한다. 분석은 크게 두 가지 자료를 이용하였다. 첫째는 개인별 업무수행평가서와 조사원 사후심사평가서에 대한 분석이며, 다른 하나는 리인터뷰 조사 결과에 대한 분석이다. 리인터뷰는 통계청에서 조사원의 업무성과를 측정하는 데 활용되지는 않았지만, 이후에 업무성과를 측정함에 있어서 유용한 정보를 제공해 줄 수 있다는 점에서 분석을 시도해 보았다.

가. 사례 분석 1: 업무수행평가서와 사후심사평가서 분석

먼저, 조사원의 개인별 업무수행평가서 분석을 통해서 개선방향에 대한 함의를 찾아보도록 하자. <표 3-6>은 2007년 현재 한 지방통계청에서 가구조사를 수행하고 있는 4명의 정규조사원과 4명의 도급조사원이 지난 4개월 동안 수행한 업무수행평가서 결과 분포를 보여준다.

조사원간 점수분포를 보면, 그 편차가 매우 작음을 알 수 있다. 가점된 경우는 매우 드물며, 거의 모든 조사원의 조사표 상에서 착오사항이 발견되지 않았거나 한두 건의 착오만이 나타났다. 착오사항이 없는 경우는 정규조사원이 도급조사원에 비해서 더 많았으며, 이런 경향은 매월 비슷하게 나타났다. 이러한 결과는 감점대상으로 채택한 항목이 과연 조사원 업무성과를 적절히 측정할 수 있는 기준역할을 하고 있는가에 대한 문제를 제기한다.

조사원 사후심사평가서도 비슷한 문제를 내포하고 있다. [그림 3-1]과 [그림 3-2]는 2006년 통계청에서 실시한 한 연간조사에서 업무를 수행한 경험이 있는 임시조사원들의 사후심사평가서 결과를 보여준다. A지역의 경우(그림 3-1 참고) 조사원의 98.1%가 우수하다는 평가를 받았으며, 보통이라는 평가를 받은 경우는 1.9%, 미흡하다는 경우는 전혀 없었다. 반면에 B지역(그림 3-2 참고)은 A지역과 달리 점수 분포가

〈표 3-6〉 조사원 업무수행평가결과 분포

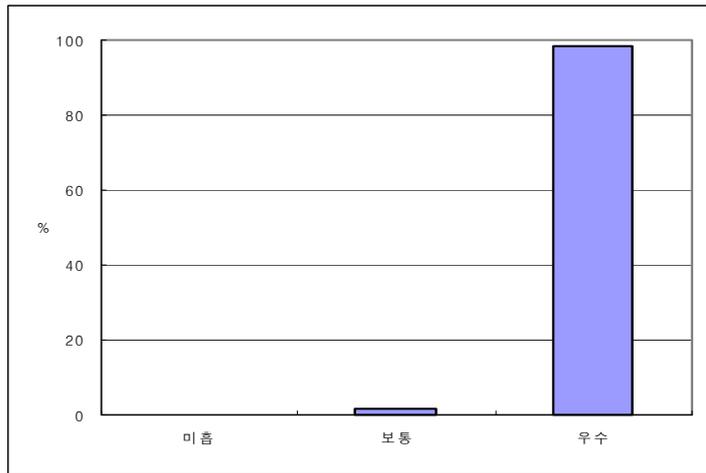
구분		1월	2월	3월	4월	평균
정규직	조사원A	100.0	100.0	100.0	100.0	100.0
	조사원B	100.0	100.0	100.0	100.0	100.0
	조사원C	100.0	100.0	100.0	100.0	100.0
	조사원D	97.5	100.0	99.5	101.5	99.6
도급직	조사원E	98.0	98.0	100.0	100.0	99.0
	조사원F	100.0	100.0	100.0	98.0	99.5
	조사원G	98.0	98.0	99.0	99.0	98.5
	조사원H	94.0	99.0	93.0	100.5	96.6

종모양의 정규분포 형태를 띠고 있는데, 우수하다는 평가를 받은 조사원은 전체의 33.3%이며, 보통이라는 평가를 받은 조사원은 58.3%, 그리고 미흡하다는 평가를 받은 조사원은 8.4%였다.

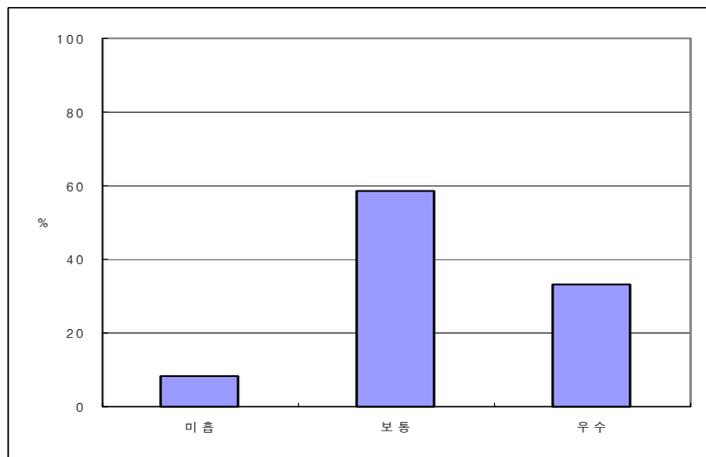
이러한 분포는 사후심사평가 척도가 조사원간 변별성을 제공하지 못함을 보여준다. 물론 A지역의 경우 모든 조사원이 우수할 수 있다는 가정을 전혀 배제할 수 없으나, 이는 동시에 척도가 조사원간 차이를 반영하지 못한다는 것을 의미할 수도 있다. 또 다른 한편으로, 평가자 집단이 한 조사원에 대한 인상을 공유하고 평가를 내렸을 가능성도 있다. 실제로 이러한 평가방법에 대해 평가자의 주관이 개입될 소지가 있다는 것은 조사연구의 초기부터 지적된 바 있다. 당시에 조사원들을 평가하기 위해서 ‘탁월(excellent), 평균 이상(above average), 미흡(poor)’과 같은 형용사가 포함되어 있는 체크리스트를 활용하였는데, 이러한 측정은 개인 평가자 내부에 마련된 기준에 의존하는 경향이 있어서 주관적으로 치우칠 경향이 있다는 비판을 받았다(Sudman, 1966).

슈퍼바이저들은 종종 모니터링의 목적에 충족되는 객관적인 근거를 사용한다기보다는 조사원과의 상호작용에 관한 일반적인 인상을 기록하는데, 이는 차별(discriminations) 혹은 편애(favoritism) 등을 반영할 수

[그림 3-1] A지역 임시조사원의 사후심사평가 결과 분포



[그림 3-2] B지역 임시조사원의 사후심사평가 결과 분포



있다는 것이다(Couper et al., 1992). 이런 비판의 타당성은 A지역과 B지역 평가자의 점수분포를 비교해 볼 경우 어느 정도의 설득력을 얻을 수 있다. A지역의 경우 한 조사원에 대해서 모두 3명의 평가자가 평가를 내렸는데 이들 간의 분산이 0인 것으로 나타났다. 즉, 세 평가자가 동일

한 점수를 준 경우가 94.0%이었다. 반면에 B지역의 경우에는 5명의 평가자가 평가를 하였으며, 한 조사원에 대해서 평가자들 간에 동일한 점수를 준 경우는 없는 것으로 나타났다. 이러한 현상은 평가 척도뿐만 아니라 평가 척도를 운영하는 평가자들의 태도 또한 업무성과 측정에서 중요한 부분임을 보여준다. NORC에서는 평가자들의 신뢰성 있는 측정을 위해서 측정연습(calibration exercise)을 실시한다. 이것은 일종의 가상 모니터링 시스템이다. 조사원이 면접하는 과정을 보고 모니터링 평가양식에 따라서 평가를 하도록 한 후, 동료 평가자와 비교해 봄으로써 척도의 신뢰도를 체크하는 방법을 활용하고 있다.

나. 사례분석 2: 리인터뷰 자료

다음으로 리인터뷰 자료에 대한 분석을 통해서 통계청의 조사원 업무성과 측정에 대한 시사점을 얻고자 한다. 조사원 업무성과를 측정함에 있어서 지금까지는 응답자 접촉 과정이나 면접 행동 등 조사원 개인의 행동에 대한 평가가 주를 이루었다. 그러나 개인의 업무수행을 평가함에 있어서도 조사원이 처한 환경의 중요성이 강조되고 있다. 리인터뷰 자료를 통해서 조사원의 업무성과를 측정할 이후에 그 결과를 제시하는 방법 중의 하나는 정확성을 측정하는 문항을 이용하여 계산한 불일치율의 정도를 조사원별로 나열하여 비교하는 것이었다. 그런데 이 방법은 문제점을 갖는다(Stokes, 1989). 할당된 업무량은 모든 조사원에게 동일하게 어렵지 않기 때문이다. 어떤 가구는 다른 가구에 비해서 조사하기가 훨씬 더 어려운데 이런 유형의 가구를 포함하는 업무지역을 할당받은 조사원은 비록 그들이 조사절차를 잘 따른다고 할지라도 쉬운 지역을 할당받은 조사원에 비해서 불일치율이 더 높게 나타날 것은 자명한 일이다. 조사원 업무량의 크기 또한 지역별로 다양하기 때문에 관찰된 불일치율만을 가지고 조사원이 수행한 업무에 대해서 평가하기란 쉽지 않다. 이는 업무성과 측정시 환경적 특성의 중요성을 암시해 준다.

통계청에서 2003년과 2004년에 수행한 8개 조사¹¹⁾의 리인터뷰 자료

11) 분석대상은 8개 조사이나 자료의 한계에 따라서 일부 조사가 분석에서 제외된 경우도 있다.

를 분석하여 조사원 업무성과 측정에서 환경적 특성이 갖는 효과를 살펴보고자 한다. 여기에서는 환경적 특성으로 자료수집방법과 지역효과의 중요성을 살펴보도록 하겠다. 조사방법론 분야에서 중요한 이슈 중의 하나는 자료수집방법에 따라서 자료의 품질에 미치는 효과가 달라지는가(mode effects)에 관한 것이다. 이에 관한 연구는 컴퓨터를 보조적으로 활용하는 기법(CAPI, CARI 등)의 발전과 함께 종이조사표를 이용한 자료수집방법과의 비교를 통해서 더욱 활발히 진행되고 있다(Fuchs, Couper, and Hansen, 2000). 자료수집방법의 효과는 아직 결론적이지 않은 상태에서 최근에는 응답률을 증가시키고 자료수집 비용을 감소시킨다는 기대 하에 혼합방식(mixed methods)이 많이 허용되고 있는 추세이다.¹²⁾

이러한 경향은 우리나라에서도 크게 다르지 않다. 일반적으로 면접을 원칙으로 하는 조사에서 면접이 아닌 다른 방법으로 자료 수집을 하는 경우가 있다. 그러나 한 조사에서 다양한 자료수집방법이 혼용되고 있다는 사실은 잘 알려져 있지 않다. 왜냐하면 조사 지침에 명기된 자료수집방법과 다른 방법을 사용할 경우, 조사원이 원칙을 어겼다는 것으로 인식되므로 조사원은 그 사실을 밝히길 꺼리는 경향이 있기 때문이다. 통계청에서 실시한 리인터뷰 조사에서는 조사원이 어떤 방법을 이용했는지를 응답자에게 직접 묻고 있어 혼합자료수집방법의 현황을 살펴볼 수 있는 기회를 제공한다(표 3-7 참고).

대부분의 조사가 면접과 함께 팩스나 이메일 등으로 완료되었으며, 특히 F 조사나 H 조사의 경우에는 50% 이상이 지정된 자료수집방법 이외에 다양한 방법이 사용되었다. 다른 자료수집방법이 사용된 비율의 차이는 조사의 성격과 관계 있는 것으로 여겨진다. 팩스를 이용한 조사의 비율이 높은 F 조사나 H 조사의 경우 사업체 조사인 반면에 면접의 비율이 높은 A 조사의 경우 가구 조사이다.

12) 이에 따라서 미국의 RTI(Research Triangle Institute)에서는 단일조사에서 다양한 자료 방식을 채택하고 모드효과를 분석적으로 측정하기 위한 연구를 시도하고 있기도 하다.

〈표 3-7〉 혼합자료수집방법 현황

(단위: 대상처수, %)

	면접	전화	팩스	이메일	기타	미확인	전체
A 조사	1,095 (92.9)	60 (5.1)	-	-	-	23 (2.0)	1,178 (100.0)
B 조사	177 (83.5)	13 (6.1)	18 (8.5)	1 (0.5)	1 (0.5)	2 (0.9)	212 (100.0)
C 조사	601 (80.1)	52 (6.9)	77 (10.3)	-	-	20 (2.7)	750 (100.0)
D 조사	394 (87.4)	43 (9.5)	9 (2.0)	-	-	5 (1.1)	451 (100.0)
E 조사	401 (80.2)	36 (7.2)	53 (10.6)	-	-	10 (2.0)	500 (100.0)
F 조사	87 (50.0)	34 (19.5)	33 (19.0)	19 (10.9)	-	1 (0.0)	174 (100.0)
G 조사	397 (66.6)	160 (26.8)	-	-	-	39 (6.5)	596 (100.0)
H 조사	49 (43.2)	7 (6.2)	48 (42.5)	9 (8.0)	-	-	113 (100.0)

면접 또한 조사원이 직접 기입을 했는지, 응답자가 직접 기입을 했는지에 따라서 차이를 보였는데, 응답자 기입이 차지하는 비율이 조사원 기입에 비해서 높은 부분을 차지하였다(표 3-8 참고). 특히 A, F, H 조사의 경우에 응답자 기입 비율이 각각 74.2%, 64.3%, 83.7%로 높게 나타났다. 이는 조사원이 묻고, 응답자가 답하는 전형적인 면접의 형식과는 차이를 보이는 것이다.

자료수집방법은 조사원 업무성과를 측정하는 데 다음과 같은 영향을 미칠 수 있다. 첫째는 면접시간의 차이를 가져온다. 면접시간은 조사원이 질문을 읽고 답하는 과정에서 조사원이 적절한 시간과 속도를 유지했는가를 평가하기 위한 기초 자료로 활용된다. 또한 조사원 수당을 면접시간에 따라서 제공할 경우 자료수집방법의 차이는 수당제공을 위한 중요한 근거가 된다. 다음의 <표 3-9>를 보면 면접을 한 경우가 전화를 이용한 경우에 비해서 오래 걸린 것으로 나타났다.

〈표 3-8〉 면접조사에서 응답기입주체의 비교

(단위: 대상처수, %)

구분	A 조사	C 조사	D 조사	E 조사	F 조사	H조사
응답자 기입	813 (74.2)	243 (40.4)	137 (34.8)	161 (40.1)	56 (64.4)	41 (83.7)
조사원 기입	230 (21.0)	288 (47.9)	207 (52.5)	170 (42.4)	21 (24.1)	7 (14.3)
조사원 보충기입	52 (4.7)	70 (11.7)	50 (12.7)	70 (17.5)	10 (11.5)	1 (2.0)
전체	1,095 (100.0)	601 (100.0)	394 (100.0)	401 (100.0)	87 (100.0)	49 (100.0)

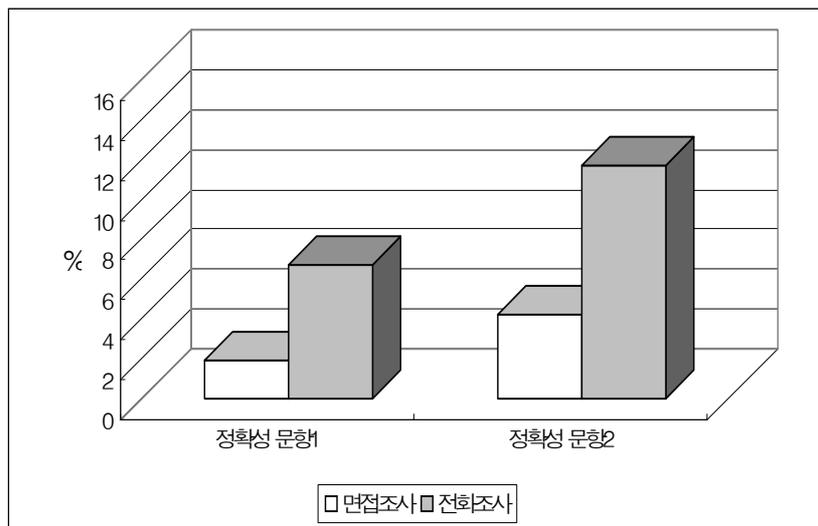
〈표 3-9〉 자료수집방법에 따른 면접시간 차이 비교

(단위: %)

구분		10분 미만	10~20분 미만	20분 이상
B 조사	면접	22.0	25.4	52.6
	전화	69.3	15.4	15.4
C 조사	면접	39.8	32.9	27.3
	전화	76.9	11.5	11.5
D 조사	면접	18.5	24.9	56.6
	전화	83.8	7.0	9.3
E 조사	면접	27.7	27.4	44.9
	전화	77.8	16.7	5.6
F 조사	면접	75.9	17.1	7.0
	전화	98.8	0.6	0.6
G 조사	면접	32.7	28.6	38.8
	전화	100.0	0.0	0.0
H 조사	면접	21.6	36.4	42.1
	전화	71.9	12.5	15.6

둘째, 자료수집 방법에 따라서 불일치율에도 차이가 나타났다. 리인터뷰에서는 정확성을 체크하기 위해 조사의 특성에 따라서 조사원이 대상처에 알아보지 않고는 임의로 작성하기 곤란하다고 판단되는 항목을 물어본다. 여기에서 서로 다른 두 시점에서 물어본 두 문항간의 일치여부가 자료수집방법에 따라서 차이가 있는가를 보았다. [그림 3-3]을 보면 두 가지 정확성 문항 모두에서 전화조사를 한 경우가 면접조사를 한 경우에 비해서 기존 문항과 정확성 검증 문항 사이에 불일치율이 높았으며 이는 통계적으로 유의미한 것으로 나타났다. 다양한 자료수집방법의 혼합이 조사원 업무성과 측정에서 갖는 의미는 무엇인가? 단일조사 내에서 다양한 자료수집방법에 따른 모니터링 측정기준을 마련해야 한다. 면접과정에 대한 모니터링은 질문전달(속도와 명확성), 프로빙 행위, 응답자에 대한 피드백 등에 중점이 두어져 있는데, 실제 현장에서 면접이나 전화 이외에 다른 조사방법이 사용되고, 면접조사 중에서도 응답자 기입이 차지하는 비중이 크다는 것은 면접모니터링에 대한 지침에서 질문전달과 관련된 속도나 명확성이 평가에서 차지하는 비중이 조정되어야 함을 보여준다.

[그림 3-3] 자료수집방법에 따른 불일치 비율의 차이 비교: A 조사



그렇다면 이러한 조사방법의 차이가 나타난 이유는 무엇인가? 만일 조사원 개인의 선택이 아닌 환경적 요인이라면 조사원 업무성과 측정시 이에 대한 고려를 해야 할 것이다. 본 연구에서는 도시규모에 따른 자료 수집방법의 차이가 통계적으로 유의미한가를 분석하여 보았다(표 3-10 참고).

〈표 3-10〉 도시규모에 따른 자료수집방법의 차이 비교: A 조사

(단위: 대상처수, %)

구분	면접조사	전화조사	미확인	전체
특별시	129 (90.8)	10 (7.0)	3 (2.1)	142 (100.0)
광역시	391 (93.5)	19 (4.5)	8 (1.9)	418 (100.0)
일반시도	575 (93.0)	31 (5.0)	12 (1.9)	618 (100.0)

도시규모는 행정구역에 따라서 특별시, 광역시, 일반시도로 구분하였으며, 분석결과, 특별시에서 면접조사가 이루어지는 비율(90.8%)은 광역시나 일반시도에서 면접조사가 이루어지는 비율(93.5%, 93.0%)에 비해서 낮음을 알 수 있다. 이러한 차이는 통계적으로 유의미하지는 않았으나 지역규모를 읍·면·동별로 구분하여 분석하였을 경우, 시골지역과 도시지역의 효과가 좀 더 명확히 나타날 것으로 기대해 볼 수 있다. 자료의 한계상 진전된 분석이 이루어지지 못하였으며 대안적으로 면접조사 방식 내에서 응답기업 주체가 누구인가를 살펴보았다. 시골지역보다는 도시지역에서 응답자 스스로 기입하고자 하는 추세가 나타날 것으로 예측되며(표 3-11 참고), 분석결과는 이를 지지해 주었다.

특별시와 광역시에서는 응답자가 직접 기입하는 비율이 85.3%, 80.3%이며, 일반시도에서는 이보다 낮은 67.7%이다. 반대로 조사원 기입 비율은 일반시도(28.0%)가 특별시(11.6%)나 광역시(13.8%)에 비해서 높게 나타났으며 이는 통계적으로 유의미하였다. 이러한 분석결과는 자료수집방법을 선택함에 있어서 지역적 특성이 작용함을 보여주는 것이

다. 특별시인 경우가 광역시나 일반시도인 경우에 비해서 면접을 통한 자료수집 비율이 낮았으며, 특별시나 광역시인 경우가 일반시도인 경우에 비해서 응답자 기입비율이 조사원 기입비율보다 높은 것으로 나타났다. 이러한 분석결과는 자료수집방법을 선택함에 있어서 지역적 특성이 작용함을 보여주는 것이다. 이는 조사원 업무성과 측정시 환경적 특성의 중요성을 간접적으로 지지해 준다.

〈표 3-11〉 지역에 따른 응답기입주체의 차이 비교: A 조사

(단위: 대상처수, %)

구분	응답자 기입	조사원 보충기입	조사원 기입	전체
특별시	110 (85.3)	4 (3.1)	15 (11.6)	129 (100.0)
광역시	314 (80.3)	23 (5.9)	54 (13.8)	391 (100.0)
일반시도	389 (67.7)	25 (4.3)	161 (28.0)	575 (100.0)

제4절 결론

지금까지 조사원 업무성과 측정을 위한 기초연구의 일환으로 조사원 업무성과 측정을 위한 개념적인 틀과 우리나라 통계청의 업무성과 측정 현황을 살펴보고, 앞으로의 개선방향에 대해서 언급해 보았다.

조사원 업무성과는 조사원의 주요 업무인 응답자 접촉 및 협조유도와 면접이라는 두 가지 업무를 대상으로 과정과 그 산출물에 대한 평가를 중심으로 이루어진다. 그동안 조사원 업무성과는 응답률, 거절률, 무응답률, 조사표 검토 등 산출물 중심으로 평가되었다. 그런 점에서 과정에 대한 평가가 보완되어야 하며 이를 위해서는 모니터링 시스템의 구축이 동시에 이루어져야 한다. 기존의 성과평가 틀에서도 과정 중심의 평가방법이 있었다. 관찰, 녹음 및 감독자의 주관적 평가 등을 실제로 활용해 왔는데, 최근 과학기술의 발달과 함께 도입된 모니터링 시스템

은 이런 평가방법에 비해서 좀 더 객관적인 정보를 풍부하게 제공해 줄 수 있다는 장점이 있다. 예를 들면, PDA나 CATI 등의 키스트록 파일이나 로그파일 내의 면접시간, 문항별 반응시간, 응답 대상자 선정과정 등의 정보들을 업무성과를 측정하는 데 활용할 수 있을 것이다. 그러나 통계청의 조사원 업무성과 측정현황에서는 응답자 접촉 및 협조유도와 면접 과정에 대한 평가가 매우 부족한 것으로 나타났다. 또한 현재 활용하고 있는 업무성과 측정지표의 신뢰도 및 타당도 문제가 제기되었다. 이러한 현황분석을 통해서 다음과 같은 제안을 할 수 있다.

첫째, 조사원이 면접을 진행하는 과정을 지속적으로 관리하고 감독할 수 있는 모니터링 시스템의 구축이 필요하다. 물론 지속적이고도 체계적인 모니터링 시스템을 구축하기 위해서는 컴퓨터 등 각종 장비가 도입되어야 할 것이며, 이는 많은 예산과 직결되는 부분이므로 당장 실현은 어려울 수 있다. 그러나 모니터링 시스템의 확립을 위해서 단계적인 도입은 생각해 볼 수 있을 것이다. 올 8월부터 울산출장소에서 CATI를 이용한 가구조사가 실시될 예정이며, 경제활동인구조사의 경우 PDA를 통한 조사를 시행하고 있는데 이 조사들에 대해서 부분적으로나마 면접과정에 대한 모니터링을 실시해 볼 수 있을 것이다.

과정 중심의 모니터링으로 조사원과 응답자 간의 면접과정에 대한 정보뿐만 아니라 응답자를 접촉해서 협조를 유도해 가는 과정에 대한 정보도 필요하다. 모니터링 시스템의 일부분으로서 조사원들의 모든 방문기록(방문일자, 방문시간, 방문결과 등)을 관리할 수 있는 트래킹 시스템의 체계를 확립할 필요가 있다. 트래킹 시스템은 조사의 전반적인 진행 상황을 한눈에 파악할 수 있게 해주며, 조사원별·조사가구별·조사지역별 진행 상황에 대한 분석을 통해서 업무성과가 현저히 낮은 곳에 대한 신속한 지원 체계를 마련해 줄 수 있다. 조사 관리자는 방문기록을 보고 응답 거부자를 설득하기 위한 적절한 시간이 언제인지, 업무패턴이 비효율적이지는 않은지 등을 파악, 조사원이 해야 될 다음 단계의 일을 빨리 결정할 수 있도록 해준다. 동시에 전형적이지 않은 패턴의 코드를 가진 조사원은 없는지, 완성된 조사표의 이력이 비현실적인 패턴을 가지고 있는지는 않은지 등을 통해서 허위조사에 대한 잠재적인 지표를 발견할 수도 있다.

둘째, 이러한 모니터링 아래에서 조사원들의 업무성과 측정을 위한 객관적인 지표를 만들 수 있을 것이다. 기존 측정지표의 내용과 점수체계에 대한 개선이 필요하며, 모니터링 시스템의 도입이라는 조건 아래에서 응답자 접촉 및 협조유도과정, 면접과정에 대한 측정지표의 개발이 필요하다. 또한 이 과정에서 리인터뷰 자료의 분석을 통해서 나타났듯이, 조사지역 및 자료수집방법 등 환경적인 특성을 고려해야 할 것이다. 측정지표를 개발함에 있어서는 그 지표가 조사원간 변별력을 띠어야 하며, 조사유형 및 조사지역간에 비교 가능해야 한다.

모니터링 시스템 및 업무성과 측정 시스템의 마련은 조사원에 대한 감시체제라기보다는 조사원과 조직 사이의 신뢰형성을 위한 기초 작업으로 이해되어야 한다. 계약과 거래, 나아가 모든 사회적 접촉에는 일정 정도 상대에 대한 신뢰가 개재되어 있기 마련인데, 우리 사회에서는 제도에 대한 신뢰가 미약하며(이재혁, 2006), 이는 조직의 투명성, 효율성 등에서 여러 문제점 등을 노출시켰다. 제도적인 신뢰를 조직적인 차원에서 생각해 본다면, 조직을 운영하는 시스템에 대한 신뢰라고 할 수 있다. 모니터링 시스템의 확립은 조사원들에 대한 감시체제라기보다는 자신의 업무를 충실히 수행하는 조사원들에 대한 제대로 된 평가를 위한 기초 시스템이다. 이런 시스템의 확립은 조사산출물이 나오기까지의 전 과정에 대한 종합적인 평가를 가능케 함으로써 업무성과 측정을 하는데 도움을 줄 수 있다.

이러한 개선을 통해서 다음과 같은 효과를 기대할 수 있다. 첫째는 자료의 품질개선이다. 모니터링 시스템을 통해서 조사원으로 인해 발생 가능한 오차를 어느 정도는 통제할 수 있다. 또한 조사원의 업무성과가 낮은 경우에 그 원인이 구체적으로 어디에 있는지 분석이 가능하고, 이에 따른 적절한 대안 마련을 통해서 자료의 품질을 높일 수 있을 것이다. 둘째, 업무성과가 높은 조사원에게 업무량을 재배치함으로써 전체 조사일정을 준수하고 추가 비용이 발생하는 것을 줄임으로써 효율적인 업무진행을 할 수 있을 것으로 기대된다.

마지막으로, 이 연구가 갖는 한계와 의의를 적시하고자 한다. 이 연구에서는 조사원 업무성과평가를 위한 구체적인 측정지표들이 제한적으로만 제시되었으며, 다양한 지표들을 합산한 종합적인 평가지표를 만

들 수 있는 방안은 추후의 과제로 남겨 놓았다는 한계점을 가지고 있다. 뿐만 아니라 통계청에서 활용하고 있는 모니터링 시스템의 분석 자료가 특정 지방사무소의 자료로 제한적이어서 전체로 일반화하여 설명하는데 어려움이 있다. 그럼에도 불구하고 이 연구는 자료수집과정에서 비표본오차의 원인으로 조사원과 이들의 성과평가의 중요성을 언급하였으며, 그동안 산출물 중심으로 이루어졌던 성과평가에서 과정평가를 포함한 종합적인 평가가 되는 방향으로 전환되어야 함을 강조하였다는 데에 의의가 있다.

참고문헌

- 이재혁(2006), “신뢰와 시민사회: 한미 비교연구”, 「한국사회학」, 40(5), pp.61-98.
- 통계청(2003), 「알기 쉬운 현장조사 모니터링의 이론과 실제 - 통계청 품질관리팀의 적용사례를 중심으로」.
- _____(2007), 「통계품질관리 이렇게 합니다」.
- American Association for Public Opinion Research(2006), *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*, 4th edition, Lenexa, Kansas: AAPOR.
- Biemer, P. and G. Forsman(1992), “On the Quality of Reinterview Data with Application to the Current Population Survey”, *Journal of the American Statistical Association*, 87, pp.915-923.
- Biemer, P. and L. Lyberg(2003), *Introduction to Survey Quality*. Wiley-Interscience.
- Campanelli, P. and C. O’Muircheartaigh(1999), “Interviewers, Interviewer Continuity, and Panel Survey Nonresponse”, *Quality and Quantity*, 33, pp.59-76.
- Census Bureau(2007), “Use of Hand Held Computers in the United States 2010 Decennial Census of Population and Housing: Lessons Learned So Far”, *UNECE/Eurostat meeting on Population and Housing Censuses*, Census Bureau.
- Couper, M., L. Holland, and R. Groves(1992), “Developing Systematic Procedures for Monitoring in a Centralized Telephone Facility”, *Journal of Official Statistics*, 8, pp.63-76.
- Curtin, R., S. Presser, and E. Singer(2002), “The Effects of Response Rate Changes on the Index of Consumer Sentiment”, *Public Opinion Quarterly*, 64, pp.413-428.

- Dijkstra, W., P. Yfke, P. Ongena, and J. Smit(2007), "Monitoring Interviewer Performance by Means of Behavior Coding in a Large Scale Cross National Survey", *Proceedings of the Annual Meeting of the American Association for Public Opinion Research*, Anaheim, CA.
- Durand, C.(2005), "Measuring Interviewer Performance in Telephone Surveys", *Quality and Quantity*, 39, pp.763-778.
- Fellegi, I. and G. Brackstone(1999), "Monitoring the Performance of a National Statistical Institute(NSI)", *Statistical Journal of the United Nations ECE*, 16, pp.251-265.
- Fowler, F.(1991), "Reducing Interviewer-Related Error through Interviewer Training, Supervision, and Other Means", pp.259-275 in P. Biemer and Others(eds.), *Measurement Error in Surveys*, New York: John Wiley and Sons.
- Fuchs, M., M. Couper, and S. Hansen(2000), "Technology Effects: Do CAPI or PAPI Interviews Take Longer?", *Journal of Official Statistics*, 16(3), pp.273-286.
- Full, S., M. Haworth, and A. Stephens(2001), "Quality Measurement and Reporting in the UK Office for National Statistics", *Proceedings of Statistics Canada Symposium*, Statistics Canada.
- Groves, R.(2004), *Survey Errors and Survey Costs*, Wiley Interscience.
- Groves, R., J. Fowler, Jr., P. Couper, J. Lepkowski, E. Singer, and R. Tourangeau(2004), *Survey Methodology*, Wiley Interscience.
- Kennickell, A.(2006) "Who's Asking? Interviewers, Their Incentives, and Data Quality in Field Surveys", *Proceedings of the 29th General Conference of the International Association for Research in Income and Wealth*, International Association for Research in Income and Wealth.
- Kviz, F.(1977), "Toward a Standard Definition of Response Rate." *Public Opinion Quarterly* 41 : pp.265-267.

- Lepkowski, J., S. Sadosky, and P. Weiss(1998), "Mode, Behavior, and Data Recording Error," pp.367-388 in M. Couper, R. Baker, J. Bethlehem, C. Clark, J. Martin, W. Nicholls II, and J. O'Reilly (eds.), *Computer Assisted Survey Information Collection*, Wiley, New York.
- Loosveldt, G., A. Carton, and J. Billiet(2004), "Assessment of Survey Data Quality: A Pragmatic Approach Focused on Interviewer Tasks", *International Journal of Market Research*, 46(1), pp.65-82.
- Mathiowetz, N. and C. Cannell(1980), "Coding Interviewer Behavior as a Method of Evaluating Performance", *Proceedings of the Section on Survey Research Methods*, American Statistical Association.
- Mccarty, C.(2003), "Differences in Response Rates Using Most Recent Versus Final Dispositional in Telephone Surveys", *Public Opinion Quarterly*, 67, pp.396-406.
- Pickery, J. and G. Loosveldt(2002), "A Multilevel Analysis of Interviewer Effects on Various Components of Unit Nonresponse", *Quality and Quantity*, 36, pp.427-437.
- Saeboe, H., J. Byfuglien, and R. Johannessen(2003), "Quality Issues at Statistics Norway", *Journal of Official Statistics*, 19(3), pp.287-303.
- Stokes, L.(1989), "Evaluation of the Interviewer Quality Control Procedure for the Post-Enumeration Survey", *Proceedings of the Survey Research Methods Section*, American Statistical Association.
- Sudman, S.(1966), "Quantifying Interviewer Quality", *Public Opinion Quarterly*, pp.664-667.
- Tarnai, J. and D. Moore(2006), "Measuring and Improving Telephone Interviewer Performance and Productivity", Presented at the 2nd International Conference on Telephone Survey Methodology, Abt associates Inc.
- Thurkow, N., S. Bailey, and M. Stamper(2000), "The Effects of Group and Individual Monetary on Productivity of Telephone

Interviewers", *Journal of Organizational Behavior Management*,
20(2), pp.3-25.

UN(2005), *Household Sample Surveys in Developing and Transition
Countries*, UN.