

1. 연구의 배경

- 1.1. 정책평가의 필요성
- 1.2. 통계 현황
- 1.3. 분석틀
- 1.4. 자료 수집과 현황

1. 연구의 배경

1.1. 정책평가의 필요성

정책 평가는 정책과정의 마지막 단계로서 환류과정을 통해서 정책과정이 원활하고 발전적으로 순환할 수 있도록 만들어주는 중요한 과정이다. 한국은 경제발전과 더불어 공공부문이 급속한 양적 성장을 이룩하였는데, 질적인 성장을 통하여 지속적인 발전을 유지하기 위해서는 정책평가가 중요하다. 또한 민주주의의 발전으로 인하여 다양한 집단이 정책 과정에 참여하고 있는데, 각 집단의 지원과 협조를 이끌어내기 위해서는 정책판단을 위한 자료가 필요하다. 더 이상 강압적인 강요나 굴복에 의해서는 정책을 원활하게 수행할 수 없다. 합리적인 근거와 자료를 가지고 대화와 타협을 통한 행정을 이룩하기 위해서는 정책에 대한 객관적인 평가가 필요하다.

한국에서는 1960년대부터 정부 업무에 대한 평가를 실시하기 시작했는데, 국회에 의한 국정감사, 감사원의 성과감사, 국무조정실의 정부업무평가로 크게 나누어볼 수 있다. 국정감사는 권력분립에 따른 견제의 원리로 이루어지는 외부평가라면, 감사원의 성과감사는 행정적인 관리통제 차원의 준외부적 평가이고, 국무조정실의 정부업무평가는 내부평가의 성격을 가진다고 할 수 있다(김현구, 2003).

신공공관리론이 IMF위기라는 파도를 타고 한국을 휩쓸게 되면서 한국 정부도 기존의 공급자 중심의 일방적 행정에서 고객중심의 정부로 거듭나기 위한 노력을 추진하고 있다. 이러한 개혁의 흐름에서 가장 중요시 되는 현상은 국민들이 정부업무에 과연 얼마나 만족하느냐에 관한 것이다. 따라서 정책평가에 일반국민들이 정책수혜자로서 이해관계자로서 참여하는 것은 바람직한 현상이라고 볼 수 있다. 그리고 실제로 많은 부처에서 정부정책과 관련하여 만족도 조사를 실시하고 있다. 정책평가를 시민의 입장에서 수행한다는 의미에서 바람직한 현상이기는 하지만 자칫 잘못된 조사로 인하여 조사결과가 왜곡된다면 오히려 정책에 치명적인 영향을 미칠 수 있다. 또한 만족도 조사를 수행하는 기관들이 통계적 왜곡을 통하여 만족도 조사결과를 특정한 방향으로 의도적 해석을 할 우려도 있다. 그러므로 민주주의의 측면에서 시민참여의 한 경로가 될 수 있는 만족도 조사가 정확하게 수행되어 실효성을 담보하기 위해서는 조사에 대한 통계적인 분석과 평가가 이루어져야 한다.

1.2. 통계 현황

한국은 기본적으로 분산형 통계제도를 사용하고 있다. 통계활동이 여러 기관에 분산되어 각 기관이 필요한 통계를 직접 작성하고 있기 때문이다. 그러나 이러한 통계의 작성은 모두 통계법에 따라야한다.

2016.12.13 (1시 기준) 통계법 제18조(또는 제20조)에 의거 승인받은 통계는 총 991종으로서 지정통계 94종, 일반통계 897종이다. 작성방법별로는 조사통계는 435종, 보고통계는 458종, 가공통계는 98종이다. 작성기관을 정부기관과 지정기관으로 구분하여 볼 때 정부기관에 의하여 작성되고 있는 통계는 815종(통계청은 60종)이며, 지정기관에서 작성하고 있는 통계는 176종이다(Kosis).

제18조(통계작성의 승인) ①통계작성기관의 장은 새로운 통계를 작성하고자 하는 경우에는 그 명칭, 종류, 목적, 조사대상, 조사방법, 통계표 서식, 조사사항의 성별구분 등 대통령령으로 정하는 사항에 관하여 미리 통계청장의 승인을 받아야 한다. 승인을 받은 사항을 변경하거나 승인을 받은 통계의 작성을 중지하고자 하는 경우에도 또한 같다. <개정 2016.1.27.>

②통계청장은 다음 각 호의 어느 하나에 해당하는 경우에는 제1항에 따른 승인을 하지 아니할 수 있다.

1. 이미 승인을 받은 다른 통계와 조사 또는 보고의 대상·목적 및 방법 등 그 내용이 동일 또는 유사하다고 인정되는 경우

2. 표본규모가 지나치게 작거나 검증된 통계작성기법을 사용하지 아니하여 통계의 신뢰성을 확보할 수 없다고 인정되는 경우

3. 조사 또는 보고의 대상 또는 목적 등이 특정 이익집단 또는 특정부문에 편중되거나 영리적인 목적으로 작성되는 등 공공의 이익을 목적으로 작성된다고 보기 어려운 경우

③통계청장은 제1항에 따른 승인을 한 때에는 이를 고시하여야 한다. 이 경우 승인을 한 통계의 명칭, 통계작성기관의 명칭 등 고시에 포함되어야 할 사항은 대통령령으로 정한다.

제20조(통계작성의 협의) ①통계작성기관의 장은 다른 법률에 따라 통계를 작성하는 경우 제18조제1항에 따라 승인을받아야 하는 사항 중 그 법률에서 정하지 아니한 사항에 관하여는 미리 통계청장과 협의하여야 한다. 협의를 거친 사항을 변경하거나 협의를 거친 통계의 작성을 중지하고자 하는 경우에도 또한 같다.

②제1항에 따른 협의를 거친 경우에는 제18조제1항에 따른 승인을 받은 것으로 본다.

이와 같이 통계청장의 승인을 받고 통계법의 적용을 받는 통계들은 통계법 제9조에 따른 정기통계품질진단, 동법 제10조에 의한 수시통계품질진단, 동법 제11조에 의한 자체통계품질진단을 받는다.

제9조(정기통계품질진단) ①통계청장은 정부의 각종 정책 수립·평가 또는 다른 통계의 작성 등에 널리 활용되는 통계로서 대통령령으로 정하는 주요 통계의 작성 및 보급의 제반 과정에 대하여 10년의 범위 안에서 주기적으로 통계품질진단(이하 "정기통계품질진단"이라 한다)을 실시하여야 한다. <개정 2012.12.18.>

②통계청장은 다음 각 호의 사항이 포함된 정기통계품질진단계획을 수립하여 정기통계품질진단을 실시하는 해의 2월 말일까지 통계작성기관의 장에게 통보하여야 한다.

1. 정기통계품질진단의 대상 통계
2. 정기통계품질진단의 시기 및 방법
3. 정기통계품질진단 결과의 활용계획
4. 그 밖에 정기통계품질진단의 실시에 관하여 필요한 사항

③정기통계품질진단은 제35조에 따른 자료제출요구, 통계작성기관에 대한 방문 확인, 통계 응답자에 대한 현지 확인 등의 방법으로 실시한다.

④통계청장은 제10조에 따른 수시통계품질진단을 실시 중이거나 대통령령으로 정하는 기간 내에 실시하였던 통계와 제11조에 따른 자체통계품질진단을 한 통계 중 그 품질이 우수하다고 판단되는 통계에 대하여는 정기통계품질진단을 면제할 수 있다.

⑤통계청장은 정기통계품질진단을 완료한 때에는 그 결과를 지체 없이 관계 통계작성기관의 장에게 통보하여야 한다.

⑥정기통계품질진단계획의 수립, 정기통계품질진단의 주기, 방법 및 절차 등에 관하여 필요한 사항은 대통령령으로 정한다.<개정 2012.12.18.>

제10조(수시통계품질진단) ①통계청장은 제11조에 따른 자체통계품질진단을 실시하지 아니하거나 품질이 저하되었다고 믿을만한 상당한 이유가 있는 통계에 대하여는 수시로 통계품질진단(이하 "수시통계품질진단"이라 한다)을 실시할 수 있다.

②통계청장은 수시통계품질진단을 실시하는 때에는 해당 통계작성기관에 미리 수시통계품질진단의 사유·시기 및 방법 등을 통보하여야 한다.

③제9조제3항 및 제5항은 수시통계품질진단에 관하여 준용한다.

④수시통계품질진단의 방법 및 절차 등에 관하여 필요한 사항은 대통령령으로 정한다.

제11조(자체통계품질진단) ①통계작성기관의 장은 소관 통계에 관하여 매년 통계품질진단(이하 "자체통계품질진단"이라 한다)을 실시하여야 한다. 다만, 작성주기가 1년을 초과하는 통계의 경우에는 그 통계를 작성하는 해 또는 그 다음 해에 실시할 수 있다.

②통계작성기관의 장은 소관 통계에 대하여 정기통계품질진단 또는 수시통계품질진단을 받은 때에는 그 연도의 자체통계품질진단을 실시하지 아니할 수 있다.

③통계작성기관의 장은 자체통계품질진단을 실시한 해의 12월 31일까지 그 결과를 통계청장에게 제출하여야 한다.

④자체통계품질진단의 방법 및 절차 등에 관하여 필요한 사항은 대통령령으로 정한다.

그러나 통계법 제3조 제1항에 의해서 기관이 내부적으로 사용할 목적으로 작성하는 수량적 정보는 통계법의 적용을 받지 않는다. 이와 같이 통계자료를 제외한 수량적 정보를 포함하여 공공기관이 직무상 작성·취득하여 관리하고 있는 문서·대장 및 도면과 데이터베이스 등의 전산자료를 행정자료라고 한다. 통계법 적용 대상이 아닌 수량적 정보는 통계법 시행령 제2조에 열거되어 있다.

제3조(정의) 이 법에서 사용하는 용어의 정의는 다음과 같다. <개정 2009.4.1., 2012.12.18., 2016.1.27.> 1. "통계"란 통계작성기관이 정부정책의 수립·평가 또는 경제·사회현상의 연구·분석 등에 활용할 목적으로 산업·물가·인구·주택·문화·환경 등 특정의 집단이나 대상 등에 관하여 직접 또는 다른 기관이나 법인 또는 단체 등(이하 "기관등"이라 한다)에 위임·위탁하여 작성하는 수량적 정보를 말한다. 다만, 통계작성기관이 내부적으로 사용할 목적으로 작성하는 수량적 정보 등 대통령령으로 정하는 수량적 정보를 제외한다. 2. "지정통계"란 제17조에 따라 통계청장이 지정·고시하는 통계를 말한다. 3. "통계작성기관"이란 중앙행정기관·지방자치단체 및 제15조에 따라 지정을 받은 통계작성지정기관을 말한다. 4. "통계자료"란 통계작성기관이 통계의 작성을 위하여 수집·취득 또는 사용한 자료(데이터베이스 등 전산자료를 포함한다)를 말한다. 5. "공공기관"이란 중앙행정기관, 지방자치단체 및 다음 각 목의 기관을 말한다. 가. 「공공기관의 운영에 관한 법률」에 따른 공공기관 나. 「지방공기업법」에 따른 지방공사 및 지방공단 다. 「유아교육법」, 「초·중등교육법」, 「고등교육법」, 그 밖의 다른 법률에 따라 설립된 각종 학교 라. 특별법에 따라 설립된 특수법인 마. 「신용정보의 이용 및 보호에 관한 법률」 제25조에 따라 금융위원회의 허가를 받은 종합신용정보집중기관 6. "통계종사자"란 통계작성기관에서 통계의 작성 또는 보급에 관한 사무에 종사하는 사람을 말한다. 7. "행정자료"란 통계자료를 제외한 것으로서 공공기관이 직무상 작성·취득하여 관리하고 있는 문서·대장 및 도면과 데이터베이스 등 전산자료를 말한다.
--

이에 따르면 각 부처에서 실시하는 대부분의 만족도 조사는 통계법의 적용을 받지 않는 경우가 많다. 따라서 만족도 조사 수행을 위한 적절한 기준의 제시와 평가가 필요하다. 이것이 만족도 조사를 통해서 정책평가의 유용성을 확보하는 첫걸음이 될 것이다.

통계법 시행령

제2조(법 적용 대상이 아닌 수량적 정보) 「통계법」(이하 "법"이라 한다) 제3조제1호 단서에
서 "대통령령으로 정하는 수량적 정보"란 다음 각 호의 수량적 정보를 말한다. <개정
2013.3.18.>

1. 통계작성기관이 대외적인 공개를 목적으로 하지 아니하고 업무 추진을 위하여 내부적으로 사용할 목적으로 작성하는 수량적 정보
2. 통계작성기관이 통계를 원활하게 작성하기 위한 사전 준비 또는 사후 확인과정에서 통계작성대상이나 절차 또는 방법 등의 적합성 및 타당성, 오차의 발생여부 등을 확인·점검하기 위하여 시험적으로 작성하는 수량적 정보
3. 통계작성기관에 소속된 직원이 개인적인 학술연구의 목적으로 연구 논문이나 보고서 등에 수록하기 위하여 작성하는 수량적 정보
4. 통계작성기관이 일상적인 업무수행의 추진 및 관리·감독을 위하여 하부조직, 소속기관, 산하기관 또는 관계기관으로부터 보고받거나 제출받은 현황, 실적 등의 자료를 집계하여 작성하는 수량적 정보
5. 통계작성기관이 그 소속 직원이나 회원, 이해관계자, 서비스 이용자 등을 대상으로 업무 추진성이나 계획에 관한 만족도 등 주관적인 인식이나 의식 또는 의견을 조사하여 작성하는 수량적 정보
6. 그 밖에 통계작성기관이 정부정책의 수립·평가나 경제·사회현상의 연구·분석 등 사회공공의 이익을 목적으로 작성한다고 보기 어려운 수량적 정보

1.3. 분석틀

오늘날 수많은 통계분석 결과들이 쏟아져 나오고 있다. 이를 통해 사회현상에 대한 이해의 폭과 깊이가 넓어졌지만, 일부에서는 분석결과에 대해 의심의 눈초리를 보내고 있다. 의심의 가장 큰 이유는 자신이 느끼는 ‘체감통계’와 실제 결과가 일치하지 않기 때문이다. 2016년 10월에 열린 통계청에 대한 국정감사에서는 국민 체감도를 외면한 통계발표에 대해 질타가 쏟아지기도 했다.

그렇다면 과연 어떠한 통계가 정확한 결과일까? 사회과학에서는 연구의 신뢰성과 타당성을 기준으로 연구결과를 평가한다. 현재 연구결과에 대한 신뢰성과 타당성을 평가하는 다수의 기법이 개발되어 있다. 그러나 이에 비하여 사전에 연구의 신뢰성과 타당성을 확보하기 위한 방법에 대한 관심은 상대적으로 적었던 것이 현실이다.

따라서 본 보고서에서는 신뢰성과 타당성을 기준으로 만족도 조사를 분석하여 신뢰성과 타당성의 저해요소들을 파악하고 이에 따른 정책적 제언을 하고자 한다.

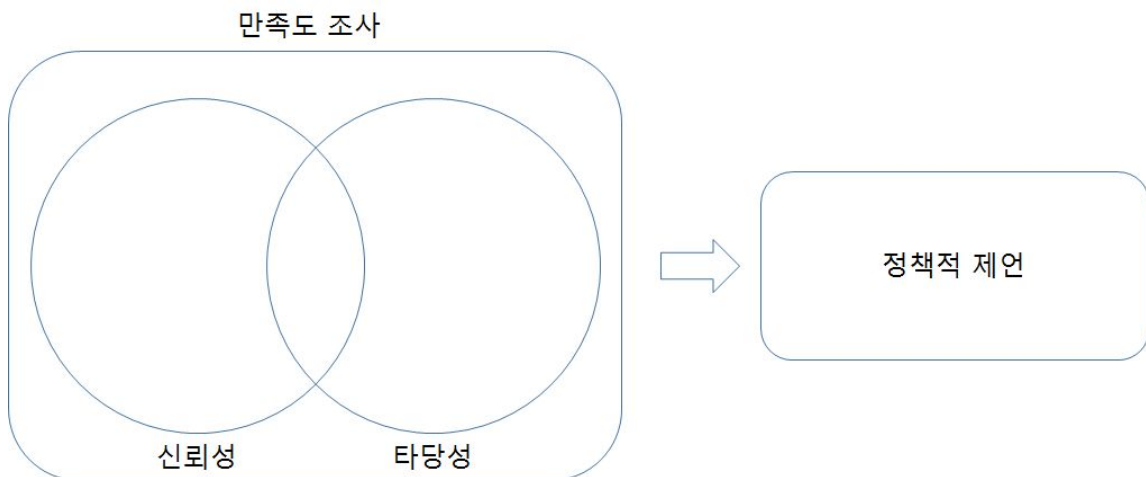


그림 1 분석틀

1.4. 자료 수집과 현황

연구는 국무조정실에 제출된 만족도 지표를 대상으로 하였다. 사용된 자료는 21개 조직에서 제출한 77개의 만족도 조사이다.

각 조직에서 제출된 자료는 만족도지표 세부현황, 조사 실시계획, 설문지로 구성되어 있다. 만족도지표 세부현황은 조직에서 수행중인 만족도 조사의 성과지표명, 조사 목적, 조사주체, 조사대상, 점수산출방식이 설명되어 있고, 조사 실시계획에는 일반적으로 조사기관, 조사기간, 조사방법, 예산, 표본에 대한 설명 등이 포함되어 있다. 설문지는 실제 조사에 사용된 설문지이다.

이중 만족도지표 세부현황은 77개의 만족도 조사와 관련하여 모두 제출되었다. 그러나 조사 실시계획은 46개의 만족도에 대해서만 제출되었고, 설문지는 31개의 만족도에 대해서 제출되었다.

표 1 분석자료 현황

자료명	세부현황	조사실시계획	설문지
개수	77	46	31

만족도 조사와 관련하여 조사기간은 43개의 만족도에서 확인할 수 있었고, 예산은 21개의 만족도에서만 정보가 제공되었다. 조사방법에 대한 설명도 43개의 만족도에서만 확인할 수 있었다. 표본추출방법은 60개의 만족도에서 나타나 있었고, 사회 경제적 변수는 제출된 31개의 설문지 중 29개 설문지에서 확인할 수 있었다.

표 2 만족도조사 세부항목 확인현황

항목	조사기간	예산	조사방법	표본추출방법	사회경제적 변수
개수	43	21	43	60	29

2. 조사방법론

- 2.1. 모집단 설정의 중요성
- 2.2. 표본추출 방법
- 2.3. 사회, 경제적 요인에 따른 만족도
- 2.4. 표본의 크기
- 2.5. 설문지 작성

2. 조사방법론

본격적으로 만족도 조사에 대한 분석을 수행 하기에 앞서 설문조사의 설계와 관련된 이론적 내용들을 먼저 살펴보고자 한다.

2.1. 모집단 설정의 중요성

통계조사를 실시할 때 설계과정에서 발생할 수 있는 오차에 집중한 나머지 실제로 무엇을 조사해야하는지의 문제에 대해서는 소홀한 경우가 많다. 모집단의 범주와 분석수준을 잘못 설정할 경우, 통계조사가 완벽하게 이루어졌다고 해도 조사결과는 실제와 왜곡되어 나타날 수 있다. 모집단의 범주와 관련된 문제는 월-로저스 현상으로 살펴보고 모집단의 분석수준과 관련된 문제는 심슨의 역설을 통해서 살펴보도록 한다.

월-로저스 현상은 전체적으로는 실질적인 점수의 상승이 발생하지 않아도 구성원의 범주를 변화시킴으로써 각 집단의 평균점수가 상승한 것처럼 나타나게 만든다.

표3 월로저스 현상

기간	1학기		2학기	
그룹	A반	B반	A반	B반
점수	5	5	5	5
	6	10	6	15
	7	15	7	20
		20	10	
평균점수	6	12.5	7	13.3

전체 총점을 보면 1학과 2학기 점수는 각각 68점으로 변동이 없다. 학생들이 1학기 때 받은 점수와 2학기 때 받은 점수는 똑같다고 설정하였기 때문이다. 그런데 2학기 성적을 보면 A반과 B반 모두 평균점수가 1학기 보다 상승하였다. 1학과 2학기의 다른점이 있다면, 1학기에는 B반에서 10점을 받은 학생을 2학기에는 A반으로 옮겼다는 사실이다. 이처럼 측정하고자 하는 대상의 범주를 잘못 설정하거나 임의로 바꾼다면 왜곡된 해석을 초래할 수 있다.

심슨의 역설은 분석수준을 설정하는 방식에 따라 비율에 대한 해석이 달라질 수 있음을 보여준다. UC버클리 사건은 심슨의 역설을 보여주는 대표적인 사례이다. 1973년, UC버클리 대학원이 여성 지원자들을 차별하고 있다는 문제가 제기되었다. 남성 지원자들은 지원자의 44%가 대학원에 합격했는데, 여성의 경우는 35%에게만 입학 허가가 주어졌다는 내용의 문제 제기였다.

표4 심슨의 역설

지원 및 합격 현황					학과별 합격현황(단위:%)	
학과구분	여성 지원자	합격자	남성지원자	합격자	여성	남성
1	10명	8명	80명	50명	80	63
2	5명	4명	60명	40명	80	67
3	80명	20명	40명	10명	25	25
4	30명	15명	40명	10명	50	25
합계	125명	47명	220명	110명		
전체 합격현황 (단위:%)	37.6		50			

※출처: 보르바흐, 코르프(2016)

전체 합격률을 보면 여성이 37.6%이고 남성은 50%이다. 차별이 존재하는 것처럼 보이는 수치이다. 그러나 학과별 합격률을 보면 상황은 달라진다. 4개의 학과 모두에서 여성지원자의 합격률이 남성의 합격률보다 낮은 곳은 없다. 오히려 학과별 합격률을 보면 남성에 대한 차별이 존재하는 것처럼 보인다. 전체 합격현황에서 여성의 합격률이 떨어진 이유는 3번학과에서 찾을 수 있다. 여성지원자는 총 125명인데 그중 80명이 3번 학과에 지원했다. 그런데 3번학과는 지원자의 75%를 불합격 시켰다. 즉, 여성들이 가장 많이 지원한 학과에서 많은 수의 여성들이 불합격하여 전체적으로 여성의 합격률이 급격히 떨어지는 현상이 발생하였다. 반면 남성 지원자들은 1번 학과와 2번 학과에 많이 지원하였는데, 이 학과는 전체적으로 합격률이 높아 남성의 전체 합격률은 올라가는 결과를 초래했다.

이러한 문제들을 종합해볼 때, 통계조사를 시행하는데 있어서 모집단을 명확히 설정하고 조사를 진행하는 것이 얼마나 중요한지 알 수 있다.

2.2. 표본추출 방법

표본을 추출하는 방법은 그 설계방식에 따라 크게 확률표본추출과 비확률표본추출로 나눌 수 있다.

2.2.1. 확률표본추출

모집단에 속한 모든 추출단위들에 대해서 표본으로 추출될 확률을 파악하고 있는 경우, 확률표본추출을 사용할 수 있다. 확률표본추출은 세부적인 방법에 따라서는 차이가 있지만, 표본추출설계과정에서 최소한 한 단계 이상 무작위표본추출과정을 포함한다. 이를 통해 조사자의 작위적 개입여지를 줄여준다.

단순무작위 표본추출(Simple Random Sampling)

단순무작위 표본추출법에서는 모집단을 구성하는 개별 구성요소들이 표본으로 선택될 가능성측면에서 동일하다. 예를들어 모집단의 구성원이 300,000명이고 표본의 크기가 300명이라면, 개별 구성요소들이 표본으로 선택될 가능성은 $1/1000$ 이 되는 것이다. 이처럼 개별 구성요소들의 선택확률을 동일하게 설정함으로써, 개별적인 구성요소들의 특성이 상쇄되어 표본은 집합적으로 보면 매우 성공적인 대표성을 지니게 된다. 즉, 제3의 설명가능성을 발생시킬 수 있는 체계적인 요인을 배제할 수 있다.

방법

- ①표본프레임을 작성한다.
- ②각 구성요소들에 고유번호를 할당한다.
- ③표본의 크기를 결정한다.
- ④난수표, 컴퓨터 프로그램을 활용하여 필요한 표본의 수만큼 개체를 추출한다.

표5 단순무작위 표본추출의 장점과 단점

장점	·모집단에 대한 자세한 지식이 불필요하다. ·외적 타당도를 통계적으로 추론할 수 있다. ·대표집단을 쉽게 얻을 수 있다. ·분류에 따른 오류의 가능성을 제거할 수 있다.
단점	·모집단에 대한 완전한 명부를 얻어야 한다. ·항상 대표성을 지닌 표본이 추출되는 것은 아니다. ·다른 표본추출절차에 비해 표본오차가 높아지는 경향이 있다.

체계적 표본추출(Systemic Sampling)

체계적 표본추출은 표본프레임에 배열된 구성요소들의 목록에서 일정한 간격으로 개체를 추출하는 방법이다. 표본추출간격(Sampling Interval)은 모집단 크기를 표본 크기로 나누어 결정한다. 예를들어 모집단의 구성원이 300,000명이고 표본의 크기가 300명 이라고 하면, 표본추출간격은 1000명이 되고 표본프레임 목록에서 매 1000번째 개체를 선택하여 표본을 구성하면 된다. 그런데 만약 모집단의 구성요소들이 일정한 주기나 규칙에 따라 목록화 되었다면, 표본 추출과정에서 편향이 발생할 수 있다. 따라서 체계적 표본추출을 할 때에는 목록의 특성을 주의 깊게 살펴보고, 만약 편향을 초래할 수 있는 방법으로 목록이 구성되어 있다면 조치를 취해야한다.

방법

- ①표본프레임 작성한다.
- ②표본프레임의 특성 검토한다.
- ③표본의 크기를 결정한다.
- ④표본추출간격을 도출한다.
- ⑤출발점을 무작위로 선택하여 표본추출간격에 따라 개체 추출한다.

표6 체계적 표본추출의 장점과 단점

장점	·표본선정이 쉽다. ·단순무작위 표본추출보다 표본선정이 더 정확할 수 있다. ·일반적으로 표본추출비용이 적게 든다.
단점	·모집단에 대한 완전한 명부를 얻어야 한다. ·주기성이 일어날 수 있다.

※출처: 김병섭(2008).

층화표본추출(Stratified Sampling)

층화표본추출은, 모집단 구성요소들의 특성 중에서 기준이 될 만한 특성을 선택하고, 그 기준에 따라 모집단을 몇 개의 집단으로 분류한 후, 각 집단에서 할당된 수만큼 표본을 추출하는 방법이다. 이 과정에서 각 하위 집단은 구성요소들이 모집단으로

함께 뭉쳐있을 때보다 좀 더 동질적인 집단이 된다. 층화표본추출은 각 하위집단에서 표본을 추출하는 비율에 따라 비례층화표출과 비비례층화표출로 구분할 수 있다. 비례층화표출은 모집에서 각 집단이 차지하는 실제비율에 따라 각 집단에서 무작위 추출을 하여 표본을 구성하는 것이고, 비비례층화표출은 모집단에 대한 사전지식을 바탕으로 조사자의 결정에 따라 실제 구성비율과 다르게 각 집단에서 무작위추출을 한다.

방법

- ①표본프레임을 작성한다.
- ②분류 기준을 설정하여 하위집단을 구성한다.
- ③표본의 크기를 결정한다.
- ④각 집단에서 할당된 만큼 개체를 무작위 추출한다.

표7 층화표본추출의 장점과 단점

장점	·관련된 변수의 대표성이 보장된다. ·다른 모집단들과의 비교가 가능하다. ·동질적인 집단에서 표본이 추출된다. ·표본오차가 감소된다.
단점	·표본추출 이전에 모집단에 대한 지식이 필요하다. ·표본추출과정에서 비용이나 시간이 많이 들 수 있다. ·집단 간 응답률이 차이가 날 수 있다.

※출처: 김병섭(2008). 편견과 오류 줄이기

다단계집락표본추출(Cluster Sampling)

다단계집락표본추출은 모집단의 전체 구성요소들에 대한 목록을 작성하는 것이 불가능하거나 실용적이지 않은 경우에 사용된다. 모집단을 하위집단으로 분류하고 그 하위 집단에서 무작위로 개체를 추출한다. 하위 집단을 선택할 때, 각 집락이 뽑힐 확률을 설정하는 방법에 따라 동일확률집단표본추출법과 비동일확률집단표본추출법으로 구분된다. 층화표본추출과 비슷해보이지만 하위집단을 분류하는 원칙이 다르다. 층화표본추출에서는 각 하위집단을 동질적으로 구성하여 각 집단간 차이를 부각시키는

설계를 하였다면, 다단계집락표본추출에서는 하위집단 간 차이를 줄이기 위해 각 집단의 구성원을 이질적인 요소로 구성하여야한다. 집단들 간의 동질성이 커질수록 대표성이 증가하고 표본오차가 감소하는 것이다.

방법

- ①표본의 크기를 결정한다.
- ②분류기준을 설정하여 하위집단을 구성한다.
- ③선택할 군집의 수와 각 군집별 표본 추출 개체 수를 결정한다.
- ④선택된 군집의 표본프레임을 작성한다.
- ⑤각 집단에서 할당된 만큼 개체를 무작위 추출한다.

표8 다단계집락표본추출의 장점과 단점

장점	·모집단의 일부분만 고려하면 된다. ·집락을 잘 규정하면 비용이 절감된다. ·집락의 특성을 평가하고 이를 모집단의 특성과 비교할 수 있다.
단점	·표본오차가 있다. ·집락이 모집단을 대표하지 못할 수 있다.

※출처: 김병섭(2008).

2.2.2. 비확률표본추출

비확률표본추출은 확률표본추출이 불가능하거나 시간적·금전적 제약이 많은 경우에 사용된다. 이러한 설계에서는 모집단 구성요소들의 추출확률을 모르기 때문에 표본으로부터 도출된 결론을 일반화할 수 있는 가능성이 현저하게 떨어진다.

표9 확률표본추출방법과 비확률표본추출방법의 비교

확률표본추출방법	비확률표본추출방법
·연구대상이 표본으로 추출될 확률이 알려져 있다.	·연구대상이 표본으로 추출될 확률이 알려져 있지 않다.
·무작위적 표본추출	·인위적 표본추출
·모수추정에 bias 없다.	·모수추정에 bias 존재 가능성 있다.
·표본분석결과의 일반화 가능	·표본분석결과의 일반화 제약
·표본오차의 추정 가능	·표본오차의 추정 불가능
·시간과 비용이 많이 소요	·시간과 비용이 적게 소요

※출처: 김병섭(2008).

판단/유의표본추출(Judgemental/Purposive Sampling)

판단/유의표본추출은 조사자가 연구목적상 자의적인 판단에 따라 표본을 추출하는 것이다. 조사자가 모집단과 그 구성요소들의 특성을 잘 파악하고 있다면, 개별구성요소들을 동등한 비중으로 간주하고 표본을 추출하는 것보다 특정지역이나 집단에 가중치를 두어 표본을 추출하는 것이 연구목적에 부합하는 대표성을 확보하는 방법이 될 수 있다.

할당표본추출(Quota Sampling)

할당표본추출은 표본이 모집단의 특성을 잘 대표할 수 있도록 모집단의 특성을 나타내는 하위집단별로 표본수를 미리 배정한 다음 표본을 추출하는 방법이다. 할당표본추출을 하기 위해서는 연구대상인 모집단의 특성을 미리 알고 있어야 한다. 그래야만 연구자가 연구문제와 관련이 있는 기준에 따라 하위집단을 구성하고 모집단을 대표할 수 있는 표본을 획득할 수 있다. 하위집단은 한 가지 기준에 의해서만 범주화되는 것이 아니라 여러 가지 기준에 따라 다양한 범주(category)가 정해질 수 있는데, 각 범주에 대한 할당표(quota matrix)를 작성하고 각 범주에 배당된 표본은 편의표본추출방법이나 판단표본추출방법이 사용된다.

편의/임의표본추출(Convenience/Accidental Sampling)

편의/임의표본추출은 조사자가 쉽게 이용 가능한 대상들을 표본으로 선택하는 방법이다. 따라서 조사결과를 일반화하는 것은 매우 위험하기 때문에 본격적인 연구에 앞서 수행하는 사전조사나 탐색적 연구에 적합한 방법이다.

눈덩이표본추출(Snowball Sampling)

눈덩이 표본추출은 모집단 구성원들이 서로를 잘 알고 있는 경우에 연구자가 임의로 선정한 제한된 표본에 해당하는 사람들로부터 추천을 받아 다른 표본을 선정하는 과정을 반복하면서 표본을 누적해가는 방법이다.

표10 비확률 표본추출방법 요약

표본추출방법	특징	장점	단점
판단/유의표본추출	연구자의 사전지식을 토대로 특정의 기준에 적합한 사례를 표본으로 선택	연구자의 사전지식 활용할 수 있음	사전지식의 타당성에 의존함
할당표본추출	모집단을 사전에 설정된 범주별로 할당한 다음, 편의표본추출방법을 사용	범주에 대한 정의와 구성비율에 관한 자료가 정확하면 대표성이 있음	범주와 구성비율에 대한 최신자료가 필요함
편의/임의표본추출	연구자가 쉽게 이용가능한 사례를 표본으로 선택	비용과 시간절약	표본의 대표성이 없음
눈덩이표본추출	소수의 사례를 표본으로 선정한 다음, 그 사람으로부터 추천받아 다른 표본을 선정하는 과정을 되풀이	표본추출프레임이 없을 경우에도 사용할 수 있음	추천의 편의 때문에 표본의 대표성을 보장하기 어려움

※출처: 남궁근(2006).

2.3. 사회·경제적 요인에 따른 만족도

확률표본 추출의 경우에는 표본의 대표성이 확보될 수 있기 때문에 문제가 되지 않지만, 비확률표본 추출의 경우에는 표본의 대표성을 장담하기 쉽지 않으며, 통제되지 않은 요소들로 인하여 오류가 발생할 우려가 있다. 따라서 비확률 표본 추출의 경우에는 사회·경제적 요인도 설문에서 함께 조사하여 조사 결과의 판단에 적용하여야 한다.

실제 경험적 연구에서도 사회·경제적인 요인이 정책이나 사업의 만족도에 영향을 미치는 것으로 나타나고 있다.

김미숙과 정경란(2008)은 영재교육 담당교사의 영재교육 만족도를 분석하였다. 다단계집락표본추출방법을 사용하여 영재교육 담당교사 447명을 표본으로 추출하여 설문조사를 실시하였다. 만족도는 67개의 문항을 바탕으로 9개의 하위개념으로 측정하여 도출하였다. 응답범주는 5점 척도로 구성되었다. 분석결과 교사의 만족도는 기관 유형에 따라, 기관의 소재지에 따라, 학교급별에 따라, 성별에 따라, 학력에 따라, 경력에 따라 차이가 발생하는 것으로 나타났다.

김인, 우양호, 조정현(2007)은 행정서비스 현장과 관련된 고객만족도의 영향 요인을 측정하였다. 비례층화표본추출로 민원인 1220명을 표본으로 선정하여 조사를 실시하였다. 만족도는 10개의 문항으로 2개의 하위개념으로 측정하였다. 분석결과, 사회경제적 변수로 설정된 생활수준, 연령, 성별은 모든 모형에서 통계적으로 유의하지 않았다. 연구에서는 응답자의 사회경제적 특징이 유의하게 나오지 않은 이유가 행정이 사회경제적 특징에 따른 차별 없이 형평성있게 이루어졌기 때문으로 분석하고 있다.

박기관(2010)은 지방정부 상수도 서비스의 고객만족도 영향요인에 관한 연구를 진행하였다. 만족도는 16개의 문항을 바탕으로 5개의 하위개념으로 측정하여 도출하였다. 응답범주는 5점 척도로 구성되었다. 분석결과 서비스 만족도는 연령에 영향을 받는 것으로 분석되었다.

박홍식과 나현(2010)은 민원행정서비스의 만족도에 대한 연구를 하였다. 구청 민원실을 방문한 민원인들을 대상으로 비비례층화표본추출의 방법으로 244명에 대하여 설문을 실시하였다. 만족도는 6개문항을 바탕으로 2개의 하위개념으로 측정하였고, 응답범주는 5점 척도로 구성되었다. 분석결과 서비스 만족도는 성별, 나이, 교육수준과 유의한 상관관계가 나타나지 않았다.

신두섭과 박승규(2012)은 공공문화기반시설 이용만족도에 대한 결정요인을 분석하였다. 할당표본추출로 1410명을 표본으로 선정하였고, 만족도는 한 개의 문항을 가지고 단일개념으로 측정하였다. 응답범주는 5점척도로 구성되었다. 분석결과 이용 만족

도는 성별, 지역, 나이, 소득에 따라서 영향을 받는 것으로 나타났다.

2.4. 표본의 크기

표본추출방법을 선택하기 위해서는 표본의 크기에 대한 고민이 필요하다. 너무 적은 수의 표본을 추출하게 되면 대표성의 문제도 있지만 과장된 결과에 결과가 사용될 수 있다. Doakes사의 치약을 사용하면 23%나 충치가 줄어든다는 광고가 있었다. 통계 조사는 독립된 연구소에서 수행되었고 광고의 공정성도 보증되었다고 한다. 그런데 실상을 알아보니 통계적으로 불충분한 표본을 채택하여 진행된 연구였다. 치약의 시험자는 단지 12명에 불과했던 것이다(더렐 허프, 1997). 그렇다고 대규모의 표본이 좋은 것만은 아니다. 대규모의 표본의 경우 극단값들은 규모가 작은 경우보다 훨씬 더 극단적이 된다. 모집단에서 표본을 추출할 경우 대개는 소규모 표본에서보다 대규모 표본에서 극단값들이 추출되는 경우가 많다(로버트 후크, 2015). 대부분의 연구는 평균값을 도출하여 통계조사를 진행하는데 이러한 극단값들이 추출되어 나온다면 모집단의 평균값은 왜곡될 수 있다.

실제로 표본의 크기는 편견에 의존하여 결정되는 경우가 많다. 첫 번째 편견은 모집단의 10%를 표본으로 얻어야 한다는 1/10법칙이다. 그런데 Lin(1976)이 제시한 모형에 의하면 모집단이 작을 경우에는 1/10법칙이 적절하지만 모집단이 커질수록 이 법칙은 모형과 차이를 보인다. 대체로 모집단의 크기가 10,000을 넘어서는 경우에는 요구되는 표본의 크기가 600개 정도로 나타난다. 두 번째 편견은 표출단위수가 일정 정도 이상이 되어야 한다는 견해이다. 그러나 표본크기가 30이상이 될 때 표본분포가 정규분포를 따른다는 점을 감안할 때 무조건 표출단위 수가 많아야한다고 볼 수 없다. 세 번째 편견은 표본의 크기를 증가시킴으로써 표본오차를 최대한 줄여야 한다는 견해이다. 그러나 표본의 크기를 증가시키면 표본오차는 줄어들지만 비표본 오차는 오히려 커지는 결과가 발생한다. 따라서 총오차를 극소화 시키는 선택이 필요하다(김병섭, 2008). 필요한 표본의 크기에 영향을 미치는 주요 요인들은 연구의 목적, 모집단의 크기, 정밀도와 신뢰도, 모집단의 분산, 표본추출방법, 예산이나 시간과 같은 자원의 정도 등이 있다(Lin, 1976).

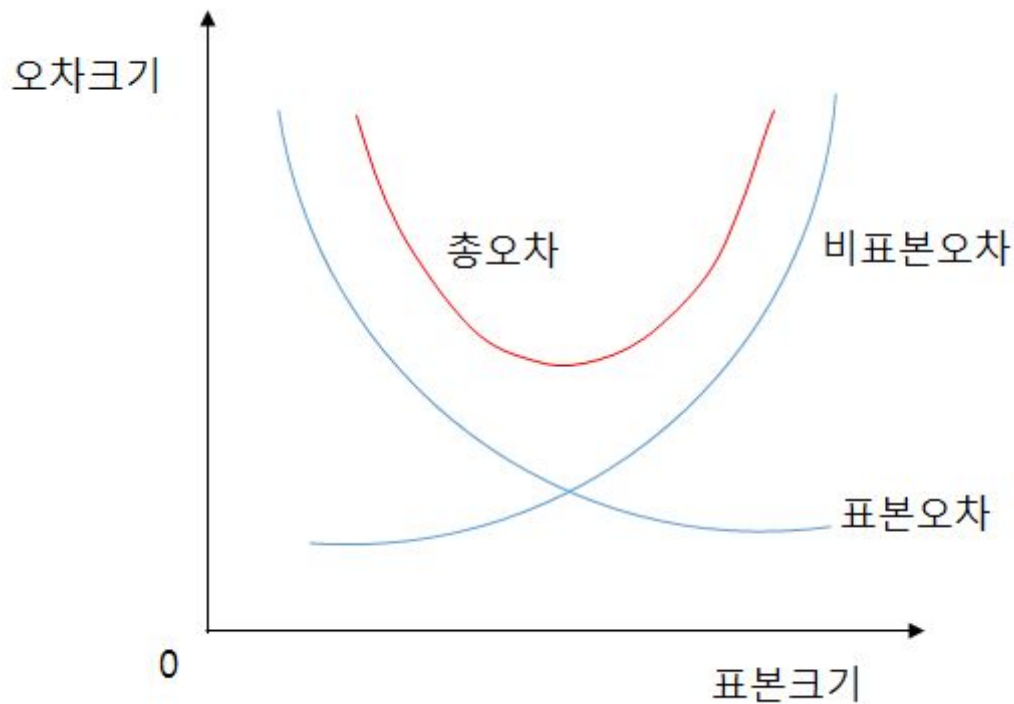


그림2 표본크기에 따른 표본오차와 비표본오차의 관계

2.5. 설문지 작성

2.5.1. 질문구성의 유의점

부정확한 질문(Imprecise Question)

설문문항의 모호성으로 인하여 응답자들이 설문문항을 서로 다르게 해석할 경우 오류가 발생한다.

- ① 일상용어를 사용하는 경우(예: 상당히, 더 나은 등)
- ② 복잡한 단어나 전문용어를 사용하는 경우
- ③ 시간이 명확하게 표시되지 않은 경우(예: 주기적, 한동안 등)

이중질문(Double-Barreled Questions)

한 가지 질문에서 두 가지 사항을 묻고 있을 때, 두 가지 사항에 대해 응답자가 다른 견해를 가지고 있다면, 응답의 신뢰도가 떨어진다. 이 경우 무응답이 많거나 사실과 왜곡된 응답이 나타날 수 있다.

유도질문(Loaded Questions)

연구자가 원하는 대답을 설문문항에 암시하게 되면, 응답자는 자신의 행동이나 의견에 따라 응답하는 것이 아니라 설문문항에 반응하여 대답을 하는 경우가 발생한다.

- ①문제상황에 대한 가치판단을 제시한 후, 해결책을 질문하는 경우
- ②호의적인 대상이나 친숙한 대상을 대답과 연계시키는 경우

부적절한 용어(Improper Terms)

응답자가 이해하기 힘들거나 다르게 해석할 수 있는 용어의 사용은 피해야한다.

- ①전문용어, 속어, 은어
- ②선동적인 용어
- ③약어, 줄임말
- ④포괄적인 용어(예: 절대로, 항상 등)

부적절한 질문(Irrelevant Questions)

응답자가 자신의 경험과 지식의 범위를 벗어난 질문에 응답할 경우, 왜곡이 발생할 수 있다. 응답할 수 없는 위치의 사람에게 질문을 하는 것은 바람직하지 않다.

긴 항목(Longthy Items)

질문을 명확히 하려다 보면, 질문이 길어지거나 여러 항목들을 나열하는 경우가 발생하는데, 이 경우 응답자는 제시된 항목을 기억하지 못하거나 질문을 오해할 우려가 있다. 따라서 설문문항은 간결하고 짧은 항목으로 구성하는 것이 바람직하다.

2.5.2. 설문의 순서

McClendon 과 O'Brien(1988)에 따르면 설문 순서의 배치는 설문 결과에 영향을 미칠 수 있다고 보았고, 고길곤(2014)은 이를 인지과정모형으로 설명한다. 응답에 대한 인지과정은 질문에 대한 이해(comprehension), 정보 검색(retrieval), 판단과 평가(judgment and estimation), 최종응답(reporting)으로 구분할 수 있다. 이해단계는 응답자가 제시된 질문에 의미를 부여하는 과정이다. 이해는 질문 자체의 의미(literal meaning) 뿐만 아니라 질문에 포함된 의도나 맥락을 추론하여 현실적인 의미(pragmatic meaning)를 형성하는 단계이다. 연구자, 조사자, 응답자들 사이에 질

문의 의미를 공유하지 못한 경우에 응답 오차가 발생하며 응답자들 간 이해의 차이가 현저할 경우에도 오차가 발생할 가능성이 크다. 정보검색단계는 응답자가 질문에 응답하기 위해 관련된 지식이나 기억들을 찾아내는 단계이다. 그러나 응답자는 자신에게 주어진 시간과 질문에 대한 관심정도에 따라 검색의 범위와 강도를 조절하게 된다. 만약 응답자가 설문과정내내 동일한 검색의 범위와 강도로 떠오른 지식이나 기억에 한정하여 질문에 응한다면 일관성 효과가, 응답의 차이를 주기 위해 다른 지식과 기억으로 질문에 응한다면 대비효과가 발생하게 된다. 판단과 평가의 단계는 검색된 지식과 기억을 결합하고 보완하는 과정을 뜻한다. 개인들에게 검색된 정보가 동일하다고 해서 그들의 판단이 일치하지는 않는다. 정보가 주어진다고 하더라도 쉽게 접근 가능한 정보만을 이용하여 판단하거나 특정 정보를 과장하여 일반화 시키는 행위를 할 수도 있다. 특히 윤리적 판단이나 가치의 문제가 결부된 경우 판단과 평가 단계에서 정보에 따른 응답결과의 차이가 커질 수 있다. 최종응답은 설문지에서 선택지를 최종적으로 선택하거나 면접자의 질문에 응답하는 단계이다. 이 단계에서 흔히 나타나는 순서오차는 리커트 척도로 된 설문지에 대한 응답에서 볼 수 있다. 응답자가 앞에서 너무 많이 3점을 선택했다고 생각되면 자신이 생각했던 점수를 선택하지 못하고 균형배분을 위해 3점이 아닌 다른 점수를 선택하는 행태가 나타난다. 이러한 결과는 응답자가 정보에 대한 판단과 평가에 의해서 응답을 결정하는 것이 아니라 최종응답 시점에서 점수의 배분을 위해 선택을 망설이거나 변경하는 과정에서 발생하는 오차이다.

설문 순서와 관련된 이슈들

①일관성 효과(Consistency Effects)

응답자는 질문에 대한 대답을 할 때, 그 질문을 평가할 비교 기준을 먼저 정한다. 일관성 효과란 각 질문들에 대해서 동일한 기준을 적용하는 것을 말한다. Shuman & Presser(1981)의 연구에 따르면 동일한 내용을 묻는 질문이 가깝게 배치되어 있을 수록 멀리 떨어져 배치되어 있을 때보다 응답이 가까운 값으로 수렴한다고 한다. 여기서 중요한 것은 단지 질문 문항의 물리적 거리가 아니라 문항 간에 기준을 변경시킬 수 있는 요소들이 존재하고 있는가이다. 따라서 조사자는 질문들을 배열할 때, 다른 질문의 응답에 영향을 미칠 수 있는 질문들이 있는 지를 이론적·논리적으로 파악해야 한다.

②대비효과(Contrast Effects)

응답자가 어떤 항목에 대하여 그 이전 항목과 차이를 주어야 한다고 판단했을 때, 실제 응답에 있어서 그 항목을 독립적으로 질문했을 때보다 과장되게 대답하는 것을 의미한다. Moore(2002)의 연구에 의하면, 전임 백안과 대변인인 Newt Gingrich와 전 상원 여당의 원내총무인 Bob Dole의 진실성에 대해서 각각 독립적으로 물었을 경우에는 41%와 60%로 조사되었는데, 둘에 대한 진실성을 연달아 물었을 경우에는 Newt Gingrich에 대한 응답은 33%로 떨어지고 Bob Dole에 대해서는 64%로 증가하였다. 이와 같이 서로 상반되는 특성을 지닌 대상에 대한 질문의 경우, 각각을 독립적으로 질문했을 때보다 연달아 질문할 경우 응답이 과장될 수 있다.

③피로효과(Fatigue Effect)

응답자가 설문에 피로함을 느끼게 되면 응답을 정확하게 하지 않을 우려가 있다. 피로감으로 인하여 질문을 제대로 이해하지 않고 중앙값으로 응답을 선택하거나 무응답을 할 수 있다. 이러한 피로함은 관련성 있는 여러 개의 질문들이 연달아서 이루어질 때 발생할 수 있다.

④중복효과(Redundancy Effect)

조사자가 했던 질문을 반복해서 질문하는 경우, 혹은 응답자가 그렇게 느끼는 경우, 응답자는 그 질문에 주의 깊게 대답하지 않게 된다. 보통 연구자들은 설문 항목을 만드는데 커다란 노력을 기울이기 때문에 자신에게 필요한 세세한 사항을 설문으로 작성하게 된다. 그러나 일반인인 응답자들은 그렇게 세세하게 만들어진 질문들의 차이점을 인식하지 못하는 경우가 발생할 수 있다. 이 경우 의도하지 않게 중복효과가 발생할 수 있다. 따라서 연구자들은 차이점이 있는 문항의 경우 그 구분이 명백하지 살펴보아야한다.

질문 배열 방식

위와 같은 오류의 가능성을 배제하며 편향되지 않도록 설문을 배열해야한다. 우선 일반적인 요령을 살펴보면, 첫째, 설문지의 첫 번째 항목은 응답자와 이해관계가 있는 것이거나 응답자가 접근하기 쉬운 것이어야한다. 그래야 응답자가 흥미를 가지고 설문 끝에 끝까지 임할 수 있다. 둘째는, 질문이 설문지 안내문에 나와 있는 내용과 관련되어 있어야한다. 질문이 안내문에서부터 논리적으로 전개되어 갈 때, 응답자는 설문

에 능동적으로 참여할 가능성이 높아진다. 셋째는 처음 질문은 이해하기 쉽고 위협적이지 않아야 응답자의 신뢰를 유도할 수 있다. 이러한 신뢰관계가 형성되면 친밀관계 효과(Rapport Effect)도 발생할 수 있다. 넷째는 대답하기 어려운 질문들은 응답자가 피로감을 느끼기 전에 먼저 배치해야한다. 대답하기 쉬운 사회통계학적 질문들은 보통 설문지의 끝에 배치하게 된다. 이러한 질문들은 피로감에 그리 크게 영향을 받지 않을 뿐만 아니라 사적인 질문이기 때문에 친밀한 관계가 형성되고 난 후에 가장 좋은 응답을 얻을 수 있기 때문이다. 다섯째는 질문은 응답자가 대답하기 쉽도록 주제별로 묶는 것이 좋다.

연구자가 질문의 배열을 결정할 때 항목들 간의 연속성을 유지하기 위해 노력하게 되는데, 이를 위해 깔때기 기법(funnel approach)이 사용된다. 이 방법에 의하면 처음에는 민감하지 않은 사항에 대해서 일반적인 질문들을 하고, 설문을 진행해가면서 응답자와 조사자 간의 의사소통을 통하여 서로간의 신뢰가 형성되면 점점 어렵거나 민감한 질문들을 진행시켜 나가는 방법이다. 처음에는 일반적이고 폭넓은 질문에서 시작해 좀 더 구체적인 질문으로 설문을 진행하는 것도 깔때기 방법의 하나이다.

이와 반대로 구체적인 질문에서 시작하여 일반적인 질문의 순서로 설문을 진행하는 것을 역깔때기 기법이라고 한다. 응답자가 조사 질문에 대하여 기초지식이나 의견을 가지고 있지 않다고 생각할 때, 응답자에게 먼저 정보를 주고 관심을 유도한 다음에 보다 크고 일반적인 질문을 하는 것이 대표적이다.

두 가지 방법 중 어떤 것이 더 바람직한지는 단정하기 어렵지만, 행정서비스와 관련하여 주민들의 만족도를 설문을 통하여 평가하는 경우 전반적인 만족도를 먼저 질문하고 구체적인 서비스를 질문한 경우가 그 반대의 경우 보다 만족도가 낮은 것으로 나타났다. 이러한 경향은 상대적으로 교육수준이 낮고 특정 주제에 관심이 적은 시민들에게서 더욱 많이 나타난다(Benton & Daly, 1993).

3. 측정

- 3.1. 정책/사업의 성과로서의 만족도
- 3.2. 측정의 개념
- 3.3. 척도문제
- 3.4. 신뢰성
- 3.5. 타당성

3. 측정

3.1. 정책/사업의 성과로서의 만족도

만족도 지표가 정책이나 사업의 성과를 정확하게 측정하기 위한 필요조건들에 대한 연구에 앞서 만족도 지표로 성과를 판단하는 것이 정당한지에 대한 논의가 선행되어야 한다. 이와 관련하여 견해가 대립하고 있다. 즉, 주관적 지표(subjective indicator)로 객관적 성과(objective performance)를 평가할 수 있는지에 대한 문제이다.

우선 만족도 조사를 성과평가와 연결시키는 것에 대하여 부정적인 입장이 있다. Stipak(1979)는 만족도 조사는 주관적 차원의 평가이기 때문에 이용자의 만족도가 공공서비스의 객관적인 산출과 통계적으로 유의하지 않게 나타날 수 있으며, 유의한 관계가 있다고 하더라도 개념적, 통계적인 문제들 때문에 조사의 타당성과 객관성을 확보하기 쉽지 않다고 말한다. Los Angeles 도심지역의 시민들을 대상으로 한 연구에서 경찰 서비스에 대한 시민들의 평가는 범죄해결비율(clearance rate), 도난품 회수율(property recovery rate), 지출(per capita expenditure), 인력(per capita employee), 범죄율(crime rate)과 유의한 관계가 잘 나타나지 않았다. Brown과 Coulter(1983)도 Tuscaloosa와 Alabama지역에 대한 연구에서 만족도와 공공서비스 수준은 완전히 독립적이라고 주장하고 있다.

둘째는 만족도 조사의 유용성을 긍정하는 견해이다. Parks(1984)는 주관적 지표와 객관적 성과의 관계가 유의하지 않게 나타나는 이유는 두 측정이 서로 다른 영역을 측정하고 있기 때문이며, 통계학적으로는 객관적인 성과를 측정할 때 평균으로 통합(aggregation)하여 측정하는 과정에서 정보의 손실이 발생하기 때문이라고 말한다. 따라서 객관적 성과를 투입(input)이 아니라 산출(output)에 초점을 맞추고, 산출에 대해서도 구할 수 있는(available) 부분적인 데이터만으로 측정할 것이라 아니라 정책의 모든 측면을 포괄하는 데이터로 측정을 해야한다고 설명한다. 통계적 통합의 문제와 관련해서는, 응답자는 자신이 생활하는 접촉영역을 기준으로 만족도를 평가하지만 객관적 성과는 주로 광범위한 지역을 단위로 지역의 수치를 통합하여 지표로 사용하기 때문에 발생하는 문제라고 말한다. 따라서 이러한 문제점을 보완하면 만족도 조사는 정책이나 사업을 평가하는 지표로 유용하게 사용할 수 있다고 말한다.

현대 행정은 공급자 중심에서 점차 수요자 중심으로 변화하고 있다. 이에 발맞추어 오늘날 정부 혁신의 방향도 고객만족주의, 주민중심주의로 나아가고 있다. 고객지향적 정부의 본질은 관료제가 아니라 고객요구의 충족이다(Osborne, 1993). 아무리

산출이 높은 정책이라고 해도 수요자가 인식하지 못하고 만족하지 못한다면 정책이나 사업에 대한 수정이 필요하다. 이제는 온정주의(paternalism)에 입각한 행정이 아니라 소비자에게 선택권을 부여하는 행정으로 전환하는 시점임을 생각할 때, 만족도 조사는 정책과 사업 평가의 유용한 지표라고 할 수 있다.

3.2. 측정(measurement)의 개념

사회현상을 수치로서 이해하고 평가하는 오늘날, 측정은 보편화되었으며 피할 수 없다. 그래서 이러한 측정과 관련된 문제들에 대해서 논의가 많이 이루어지고 있다. 그러나 아직까지도 어떠한 측정이 좋은 측정인지에 대한 체계적이고 공통적인 합의가 이루어지지 않은 실정이다. Stevens(1951)에 따르면 측정은 정해진 규칙에 의거하여 사건이나 대상에 숫자를 부여하는 것이다. 물론 이와 같은 정의가 완벽한 것은 아니다. 우리가 측정하는 사회현상은 추상적인 경우도 많기 때문에 사건이나 대상처럼 실체를 갖고 있지 않은 것도 많다.

Stevens의 측정은 사회과학적 의미의 측정이라기 보다는 자연과학과 좀 더 어울리는 개념정의라고 할 수 있다. 좀 더 사회과학적인 측정의 정의는 Riley(1963)의 의해서 제시되었다. Riley에 따르면 측정은 추상적인 개념을 경험적인 지시물(empirical indicants)로 연결시키는 과정이다. 즉 명백하게 구조화된 계획을 가지고 연구자의 마음속에 있는 일반적인 개념을 수치화 시키는 것이다.

이와 같은 측정의 정의는 측정이 경험적인 측면과 이론적인 측면을 함께 포함하고 있음을 보여준다. 경험적인 측면에서는 관찰가능한 응답에 초점을 맞춘다. 관찰가능한 응답은 인터뷰에 대한 대답, 관찰연구를 통한 대상의 행태, 질문지에 대한 응답자의 대답 등이 있다. 이론적 측면은 관찰가능한 응답을 통하여 드러내고자 하는 관찰불가능한(직접적으로 측정불가능한)개념에 관심이 있다. 따라서 측정의 주요 목적은 관찰불가능한 이론적 개념을 어떻게 분석가능한 지표로 만드느냐이다. 만약 지표가 제대로 설계되어 지표가 이론적 개념을 잘 대표한다면 지표를 통한 사회현상의 연구는 의미있는 사실을 밝혀낼 수 있다. 그러나 부정확한 지표는 오히려 사회현상에 대한 잘못된 추론과 이론에 대한 의심을 초래할 수 있다. 이를 방지하기 위해 그동안 사회과학에서는 이론을 검증하고 개발하기 위한 경험적 도구들을 발전시켜왔다. 측정은 결국 사회과학 연구의 시발점이자 발전의 중요한 필요조건이라고 할 수 있다.

3.3. 척도문제

사회현상은 대상을 직접 측정하는 것이 아니라 간접적으로 측정하기 때문에, 만족도 측정에 있어서도 척도의 구성은 중요하다. 만족도 측정의 경우 척도가 정책이나 사업의 성과를 정확히 측정할 수 있도록 설계되어야 측정 결과를 정책에 반영할 수 있다.

척도의 질은 신뢰성과 타당성에 의하여 평가된다. 신뢰성은 척도의 일관성을 의미하는데, 척도 간 일치정도, 척도 내 일관성으로 구분할 수 있다. 타당성은 보다 포괄적인 개념으로써 측정하고자 하는 개념을 얼마나 실제에 가깝게 측정하고 있는지에 대한 개념이다.

$$X=T+\epsilon_s+\epsilon_r$$

X=측정값

T=참값(true value)

ϵ_s =체계적 오차(systematic error)

ϵ_r =무작위 오차(random error)

측정은 X를 최대한 T와 가깝게 보여주는 것을 목표로 한다. 그러나 측정오차는 응답자의 개인적 상태, 측정 시 환경적 요인, 측정방법의 문제, 측정도구와 응답자의 상호작용, 연구자와 응답자의 상호작용 등에 의해서 발생할 수 있다(채서일, 2003). 이때 무작위 오차가 작고, 측정값의 오차가 대부분 체계적 오차로 구성되면, 그 척도는 신뢰성이 높다고 할 수 있다. 더 나아가서 측정오차($\epsilon_s+\epsilon_r$)가 작다면 그 척도는 타당성이 높다.

3.4. 신뢰성(reliability)

신뢰성은 타당성을 확보하기 위한 필요조건일 뿐만 아니라 타당도 보다 상대적으로 확보하기가 용이하므로 먼저 살펴보도록 한다.

신뢰성 높은 척도의 구성과 관련하여서는 다양한 이슈들이 있다.

①개념을 한가지 질문으로 측정할 것인가, 다양한 질문들로 구성하여 측정할 것인가. 일반적으로 척도를 구성하는 질문(item)이 많아질수록 그 척도의 신뢰성(reliability)이

높아진다고 알려져 있다. 왜냐하면 질문의 수가 증가할수록 전체 분산(variance)에서 차지하는 체계적인 분산(systematic variance)의 비율이 증가하기 때문이다.

②개념을 다차원적으로 측정할 것인가? 하나의 개념은 그 개념을 구성하는 다차원적인 개념(multi-dimensional concept)을 내포하는 경우가 있다. 따라서 개념을 측정하기 위해서는 그 개념에 내포된 각 차원의 개념을 측정하여 따로 측정하여 개념을 종합적으로 도출되어야 한다.

③질문이 어렵거나 길게 작성되어 이해하기 힘들면 신뢰성이 떨어진다. 질문을 명확하기 위해서는 질문이 길어질 수 있는데, 오히려 질문을 명확하게 하기 위해 불필요한 설명이 길어지면 신뢰성이 떨어질 수 있다.

④부정적 내용의 질문과 긍정적 질문의 내용이 혼합되어 있으면, 응답자의 혼란을 초래하여 역점수(reverse scoring)의 문제가 발생할 수 있고, 이는 신뢰성을 하락시킨다. 고길곤,탁현우,강세진(2014)은 역점수가 응답자의 주의를 환기시켜 집중력 감소와 인지능력 저하를 예방하여 합리적인 응답을 유도할 수 있다는 기존의 주장에 의문을 제기한다. 이들은 오히려 역점수를 사용하면 긍정적 질문과 부정적 질문에 대한 응답자의 비대칭적인 대담으로 인하여 응답오차가 발생할 수 있다는 가설을 제시한다. 분석결과 부정적 질문의 평균점수가 긍정적 질문에 비해서 더 높은 것으로 나타났다. 이러한 차이는 응답자들이 긍정적인 문항에 대해서는 극단적인 긍정보다는 중간적 긍정으로 대답하고, 부정적인 문항에 대해서는 중간적 부정보다는 극단적인 부정으로 응답하는 현상 때문에 발생한다.

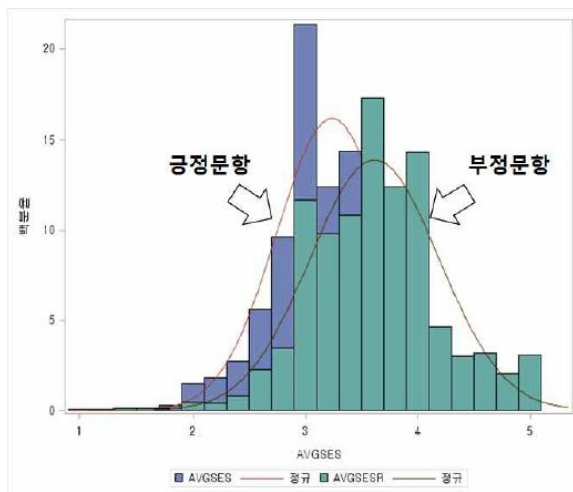


그림2 자기효능감 문항형태별 응답분포
※출처: 고길곤,탁현우,강세진(2014).

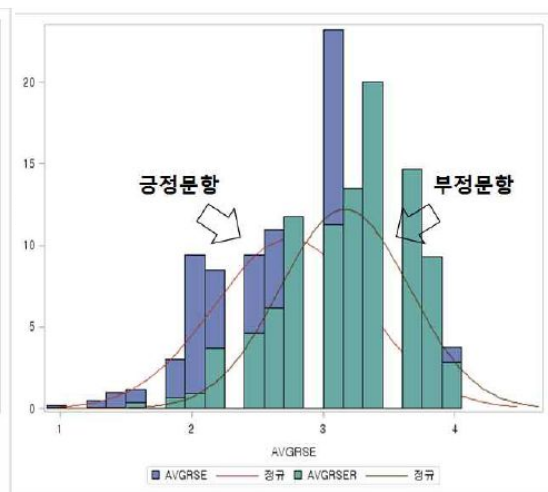


그림3 자아존중감의 문항형태별 응답분포
※출처: 고길곤,탁현우,강세진(2014).

⑤각 문항의 응답범주 수를 늘리면, 분산이 커지기 때문에 보다 높은 신뢰성을 확보할 수 있다.

김득성,장서일,고준희,김진모(2009)는 청감실험을 통하여 5점,7점,11점 척도의 비교 연구를 하였다. 실험결과 반응평균치에 있어서는 큰 문제가 발생하지 않았지만, 비율로 평가하는 경우에는 각 척도마다 기준이 다를 수 있고, 영역이 겹치는 경우가 발생하기 때문에 문제가 발생하는 것으로 나타났다.

Dawes(2012)는 5점,7점,10점 척도를 비교하면서 평균(mean), 왜도(skewness), 첨도(kurtosis)에서 어떠한 차이가 발생하는지를 분석하였다. 분석결과 평균에 있어서는 5점과 7점 척도는 평균이 같았지만, 10점 척도는 약간 낮은 수준으로 나타났다. 그리고 척도의 범주가 증가할수록 분산이 증가하였다. 그러나 왜도와 첨도에서는 유의한 차이가 나타나지 않았다.

표11 10점환산

5점척도 원점수	변환점수	7점척도 원점수	변환점수	10점척도 원점수	변환점수
1	1.0	1	1.0	1	1
2	3.25	2	2.0	2	2
3	5.5	3	4.0	3	3
4	7.75	4	5.5	4	4
5	10	5	7.0	5	5
		6	8.5	6	6
		7	10	7	7
				8	8
				9	9
				10	10

※출처: Dawes(2008)

표12 평균점수

질문 문항	평균점수: 5점 척도를 10점 척도로 변환	평균점수: 7점 척도를 10점 척도로 변환	평균점수: 10점 척도	평균점수: 5점 척도 rescaled minus 10 point	평균점수: 7점 척도 rescaled minus 10 point	평균점수: 5점 척도 rescaled minus 7 point
1	7.8	8.1	7.4	0.4	0.7	-0.3
2	7.4	7.3	6.9	0.5	0.4	0.1
3	5.1	4.6	4.8	0.3	-0.2	0.5
4	7.9	8.1	7.4	0.5	0.7	-0.2
5	6.8	6.9	6.6	0.2	0.3	-0.1
6	7.0	6.9	6.6	0.2	0.3	0.1
7	7.1	7.3	7.6	-0.5	-0.3	-0.2
8	5.9	6.0	5.3	0.6	0.7	-0.1
전체 평균	6.9	6.9	6.6	0.3	0.3	0.0

※유의수준 0.05에서 통계적으로 유의한 차이를 보임

※출처: Dawes(2008)

표13 표준편차

질문 문항	표준편차: 5점 척도를 10점 척도로 변환	표준편차: 7점 척도를 10점 척도로 변환	표준편차: 10점 척도	표준편차: 5점 척도 rescaled minus 10 point	표준편차: 7점 척도 rescaled minus 10 point	표준편차: 5점 척도 rescaled minus 7 point
1	2.7	2.4	2.5	0.2	-0.1	0.3
2	2.7	2.9	2.7	0.0	0.2	-0.2
3	3.2	3.1	3.1	0.1	0.0	0.1
4	2.4	2.3	2.6	-0.2	-0.3	0.1
5	3.0	2.9	2.8	0.2	0.1	0.1
6	2.7	2.7	2.7	0.0	0.0	0.0
7	2.7	2.6	2.4	0.3	0.2	0.1
8	2.9	2.9	2.8	0.1	0.1	0.0
전체평균	2.0	1.9	2.0	0.0	-0.1	0.1

※출처: Dawes(2008)

표14 왜도

질문 문항	왜도: 5점 척도를 10점 척도로 변환	왜도: 7점 척도를 10점 척도로 변환	왜도: 10점 척도	왜도: 5점 척도 rescaled minus 10 point	왜도: 7점 척도 rescaled minus 10 point	왜도: 5점 척도 rescaled minus 7 point
1	-1.2	-1.4	-0.8	-0.4	-0.6	0.2
2	-0.8	-0.8	-0.6	-0.2	-0.2	0.0
3	0.2	0.4	0.3	-0.1	0.1	-0.2
4	-1.1	-1.4	-0.9	-0.2	-0.5	0.3
5	-0.7	-0.7	-0.5	-0.2	-0.2	0.0
6	-0.7	-0.6	-0.5	-0.2	-0.1	-0.1
7	-0.6	-0.8	-1	0.4	0.2	0.2
8	-0.1	-0.2	0.0	-0.1	-0.2	0.1
전체평균	-0.5	-0.4	-0.4	-0.1	0.0	-0.1

※출처: Dawes(2008)

표15 첨도

질문 문항	첨도: 5점 척도를 10점 척도로 변환	첨도: 7점 척도를 10점 척도로 변환	첨도: 10점 척도	첨도: 5점 척도 rescaled minus 10 point	첨도: 7점 척도 rescaled minus 10 point	첨도: 5점 척도 rescaled minus 7 point
1	0.6	1.3	-0.1	0.7	1.4	-0.7
2	-0.4	-0.5	-0.7	0.3	0.2	0.1
3	-1.3	-1.1	-1.3	0.0	0.2	-0.2
4	0.3	1.3	-0.2	0.5	1.5	-0.1
5	-0.7	-0.6	-0.8	0.1	0.2	-0.1
6	-0.4	-0.5	-0.8	0.4	0.3	0.1
7	-0.8	-0.2	0.2	-1.0	-0.4	-0.6
8	-1.1	-1.0	-1.1	0.0	0.1	-0.1
전체평균	-0.4	-0.5	-0.4	0.0	-0.1	0.1

※출처: Dawes(2008)

⑥응답범주에 중립적 범주가 있는지의 여부는 신뢰성에 영향을 미친다. 응답자가 해당 질문에 대해서 중립적인 의견을 가지고 있다면 중립적인 대답을 할 수 있는 기회를 주어야한다. 4점척도, 6점척도와 같이 중립적인 범주가 제시되지 않는 척도는 응답자에게 선호의 여부를 강요하는 결과를 초래할 수 있다. 또한 중립적 범주(anchor point)는 응답자가 그 지점을 기준으로 정도의 판단을 보다 정밀하게 할 수 있게 해준다. 그런데 다른 한편으로, 중립적 범주가 제시되었을 때, 응답자의 반응이 중립적 범주로 쏠리기는 현상이 발생하기도 한다(Narayan &Krosnick, 1996). 이러한 현상은 질문 내용을 잘 이해하지 못하거나 응답자가 대답을 결정하지 못한 경우에 중립적 범주로 대답을 선택하기 때문으로 볼 수 있다(Raaijmakers et al., 2000). Borgers, Hox, Sikkel(2004)는 어린이와 성인의 응답을 비교하는 연구를 하면서 어린이들에게는 4점척도가 적합하다는 주장을 하기도 했다.

Churchill과 Peter(1984)는 신뢰성과 관련된 위의 이슈들에 대해서 기존연구들에 대해서 메타 연구를 실시하였다. 1964년부터 1982년까지의 중요한 경영학 저널에 수록된 6484개의 연구들에서 추출한 154개의 척도에 대해서 메타분석을 실시한 결과, 개념을 측정하는 질문의 수와 응답범주의 수가 척도의 신뢰성에 강한 영향을 미치는 것으로 나타났다.

3.5. 타당성(validity)

3.5.1. 타당성의 개념

타당성은 연구자가 연구하고자 하는 대상을 정확하게 연구하였는가를 판단하는 기준이라고 할 수 있다. 사회과학에서는 100% 정확한 연구란 불가능하기 때문에 오류를 최대한 줄여 타당성이 높은 연구를 수행하는 것이 중요하다. 타당성은 구성 타당성, 내적 타당성, 외적 타당성, 통계적 타당성으로 분류할 수 있는데, 구성 타당성은 측정도구가 이론적 구성개념을 얼마나 잘 측정하고 있는가와 관련된 타당성으로 개념을 수치로 전환하는 과정에서 범할 수 있는 오류를 파악할 수 있게 해준다. 내적 타당성은 인과성의 조건과 관련된 타당성으로서 개념 간의 인과성 규명에서 발생할 수 있는 오류를 보여준다. 외적 타당성은 연구 결과의 일반화에 대한 타당성이다. 연구 결과가 다른 대상, 다른 시기, 다른 상황에서도 유효한지를 평가하는 것인데, 분석단위와 표본 추출과 관련되어 있다. 통계적 타당성은 개념사이의 공변관계가 참인지 아

닌지에 대한 통계적 의사결정의 타당성이다. 제1종 오류, 제2종 오류와 같은 문제를 다룬다(김병섭, 2008). 본 연구에서는 측정과 관련하여 구성적 타당성을 중심으로 타당성을 살펴보고자 한다.

타당성이란 개념은 사회현상을 기술하고 설명하는 지식과 연구를 체계화하기 위해서 고안된 개념이다. 구성(constructs)이라는 개념은 두가지 의미를 지닌다. 첫째는 체계적(systemetic)이다. 체계적이란 말은 개념이 지표로 해석되는 과정이 임의적 판단에 의해서가 아니라 이론적 근거와 기준에 의해서 진행된다는 점이다. 둘째는 관찰적(observational)이다. 관찰적이라는 말은 지표가 설명력을 갖기 위해서는 그 지표가 직접적이든 간접적이든 조작가능해야한다(operational)는 말이다(Kaplan, 1964). 구성의 속성에 관찰적 특징이 없다면 구성 또한 추상적인 영역에 머물러 기존의 개념과 차이가 없어지게 된다. 그리고 체계적 특징이 없다면 구성개념은 단순한 관찰에 지나지 않는다.

구성적 타당성은 추상적·관념적 수준의 관찰불가능한 개념이 조작가능한 수준의 지표와 얼마나 잘 일치하고 있는가를 평가하는 것이다.

3.5.2. 타당성과 관련된 이슈들

타당성 문제를 평가하기 위해 사용되는 가장 잘 알려진 측정도구는 Campbell과 Fiske(1959)의 multitrait-multimethod matrix(MTMM)이다. 최근 구성적 타당성을 평가하기 위한 연구들은 대부분 MTMM의 한계와 임의성을 극복하기 위한 것이라고 할 수 있을 정도이다. 여기서는 측정방법에 대한 논의보다는 구성적 타당성을 좀더 잘 이해할 수 있고, 실제 설계된 측정도구의 타당성과 관련된 이슈에 대해서 살펴보고 한다.

내적 일관성(internal Consistency)

구성적 타당성에서의 일관성은 측정하고자 하는 개념에 사용된 척도 간의 상관관계를 의미한다. 척도 간의 상관관계는 측정의 오차에서 체계적 오차가 차지하는 비중이 많아질수록 높아진다. 일반적으로 한가지 개념을 측정하기 위해 사용된 척도들도

100%완벽한 상관관계를 가지고 있지 않다. 다만 질문들 간의 상관관계가 높다는 것은 질문들이 같은 곳을 겨누고 있다는 의미로 해석할 수 있기 때문에 이를 내적 일관성(internal consistency)이 높다고 말한다.

각 질문들이 같은 곳을 겨누고 있다면(높은 상관관계를 가지고 있다면), 측정에서의 오차는 대부분 체계적 오차에서 비롯된 것으로 판단할 수 있다. 이렇게 볼 때, 내적 일관성이 높은 경우 체계적 오차만 수정하면 구성적 타당성을 확보할 수 있기 때문에 일관성은 구성적 타당성 확보의 중요한 필요조건이다.

실제 연구에 있어서 일관성과 관련된 이슈는 문항들 간의 상관관계의 정도가 각 문항의 개별적 신뢰성(reliability)에 의해서 영향을 받는다는 것이다.

04. 만족도 조사 분석과 정책적 제언

- 4.1. 평가주체의 문제
- 4.2. 측정의 문제
- 4.3. 모집단 설정과 구성적 특성에 따른 만족도
- 4.4. 표본 추출의 문제
- 4.5. 설문지 어구
- 4.6. 설문지 배열
- 4.7. 점수 변환의 문제
- 4.8. 무응답의 문제

4. 만족도 조사 분석

4.1. 평가주체의 문제

이론적 논의

평가는 기획(arrange), 수행(produce), 필요(needs)의 주체가 누구인지에 따라 내부평가와 외부평가로 구분할 수 있다. 기획은 평가 수행을 결정하고 그에 따른 비용을 지불하는 것을 의미하고, 수행은 평가를 설계하고 데이터를 모으고 분석하여 보고하는 과정을 말한다. 필요는 평가를 통하여 만들어진 정보를 누가 필요로 하고 사용하는가에 대한 개념이다(Vedung, 1997). 이와 같은 세 가지 개념을 기준으로 내부평가와 외부평가를 구분해보면 다음과 같다. 결국 평가의 수행이 외부에서 이루어졌는지 내부에서 이루어졌는지에 따라 평가의 유형이 결정된다고 볼 수 있다.

표 16 외부평가와 내부평가 분류기준

평가 유형	수행	기획	필요
외부평가	외부	외부	외부
	외부	내부	외부
	외부	외부	내부
내부평가	내부	내부	내부
	내부	외부	내부
	내부	내부	외부

※출처: Vedung(1997)

외부평가는 내부평가보다 신뢰성(credibility)이 높다고 알려져 있다. 내부평가의 경우에는 정책이나 사업이 낮은 평가로 인하여 제한되거나 축소될 것을 우려하여 내부 평가자들이 평가결과를 그대로 발표하는 것을 주저할 수 있다. 평가결과 상에 나타난 단점보다는 장점에 가중치를 높게 주어 발표를 한다거나 긍정적 효과만을 부각시키는 기만적 평가결과가 도출될 수 있다. 물론 외부평가의 경우에도 외부평가기관이 독립성을 유지하기란 쉽지 않다. 외부평가기관은 대부분 민간 사업체의 경우가 많은데, 계약을 체결하고 유지하기 위해서는 평가 작업 발주기관의 눈치를 보지 않을 수가 없기 때문이다.

정책적 제언

만족도 조사를 정책 평가의 일부분이라고 볼 때, 평가주체를 누구로 할 것인지와 관련하여 논쟁이 있다. 여기서는 평가주체의 선정과 관련하여 평가의 목적과 표본추출의 측면에서 살펴보도록 하겠다.

첫째, 평가의 목적과 관련해서 알아본다. 평가의 목적은 크게 책임성 확보(accountability), 정책개선(improvement), 정보와 지식의 획득(basic knowledge)으로 나눌 수 있다.

일반적으로 책임성 확보를 위한 정책 평가는 평가의 객관성이 중요하기 때문에 독립적인 외부기관에서 수행되어야 한다고 알려져 있다. 통계법 제18조에 의한 국가승인 통계나 동법 제17조에 의한 지정통계의 경우, 그 결과가 공표되기 때문에 해당 법의 적용을 받는 통계조사는 행정관료의 책임성 확보의 수단이 될 수 있다. 정보공개는 행정에 대한 시민들의 이해와 관심을 증진시켜 행정과정에 대한 참여를 확대시킬 뿐만 아니라, 행정관료들의 자주성을 증진시키는 효과를 촉발하여 행정의 책임성을 확보하는 작용을 하기 때문이다(이승중, 1990). 따라서 국가승인 통계나 지정통계의 경우에는 독립적이고 전문적인 외부기관이 적절한 평가 방법이나 이론을 통하여 평가를 진행하는 것이 더 적절해 보인다. 다만 이때에도 외부평가가 독립적으로 이루어질 수 있도록 해주어야 한다. 손성규(2004)에 따르면, 외부감사인은 시간적·공간적 제약을 갖고 있기 때문에 조직 내부인에게 의존하게 된다. 뿐만 아니라 평가를 진행하는 과정에서 내부인과 지속적인 접촉을 하면서 내부인의 의도와 방향에 동조하거나 압박에 순응하는 현상이 나타날 수 있다. 그러므로 평가역량과 평가업무수행에 필요한 자원이 충분한 외부기관을 선택하는 것이 중요하며, 평가가 독립적으로 이루어질 수 있도록 최대한 접촉을 자제해야한다. 또한 평가시간이 충분히 주어져야 외부기관이 평가업무를 충분히 숙달할 수 있고 이에 따라 평가의 진행이 원활해진다.

정책개선과 관련된 평가는 내부적으로 수행하는 것이 적절해 보인다. 만족도 조사의 경우, 통계법 제3조 1항에 따라 기관 내부적으로 사용할 목적으로 수행된 만족도 조사들이 여기에 해당된다. 내부적으로 정책의 개선을 위하여 수행되는 평가는 그 평가 결과를 공개할 의무는 없다. 다시 말하면, 그 결과에 따라 조직 외부로부터 정책의 지속여부에 대한 압력을 받을 가능성이 낮다. 그래서 내부적으로 평가가 이루어지더라도 평가결과를 왜곡하거나 과장할 유인이 상대적으로 적다. 또한 정책개선을 위한 평가는 평가의 객관성과 방법론적 정밀성 보다는 신속하게 평가결과를 활용하고 그 평가결과로부터 개선점을 찾아내는 것이 중요하다. 결국 정책이 제대로 효과를 발휘하도록 하는 것이 주된 목적이기 때문이다. 따라서 평가를 기획하고 필요로 하는

기관이 필요한 부분에 대하여 신속하게 평가를 수행하는 것이 평가의 목적상 적합하다. 외부기관을 통해서 평가가 이루어질 경우, 평가의 객관성과 엄밀성을 확보할 수 있지만, 외부기관은 평가가 이루어지는 목적이나 행정적 맥락에 대한 이해가 부족한 경우가 많으므로, 정작 조직에서 필요로 하는 부분에 대한 평가가 빠지거나 제대로 이루어지지 않을 우려가 있다. 특히 시범사업에 대한 평가의 경우에는 더욱더 내부적 평가가 요구된다고 하겠다.

지식의 획득과 관련된 평가는 보편적인 지식을 도출하여 정부운영에 적용할 목적으로 이루어진다. 이러한 평가는 대학, 연구소 등의 외부기관에 의해서 이루어지는 것이 바람직하다. 하지만 실제 이러한 목적으로 실시되는 평가는 드물다.

둘째, 표본추출의 문제와 관련하여 살펴본다. 표본추출은 크게 확률표본추출과 비확률표본추출로 나눌 수 있다. 지루한 작업을 견딜 수 있는 인내심과 모집단을 확정하고 표본프레임을 구성할 수 있는 비용만 있다면 확률표본추출을 사용하는 것이 대표성을 확보하고 연구결과의 일반화에 있어서 유리하다. 그렇지만 표본프레임을 구성하는 것이 불가능하거나 연구의 편의를 추구할 때, 목적론적 방법을 사용하여 비확률적인 표본추출을 사용할 수 있다. 비확률표본 추출의 목적은 개체들이 선택될 확률이 알려지지 않은 상태에서 얼마나 확률표본추출에 근접하게 대표성을 확보할 수 있는가에 있다.

확률표본 추출의 경우에는 표본프레임을 구성하고 각 개체들이 표본으로 선택될 확률을 확인하고 무작위추출이 포함된 표본추출방법을 사용하면 된다. 물론 비용과 시간이 많이 소요되지만 정해진 절차와 방법에 따라 대표성 있는 표본을 추출할 수 있고 표본 오차까지 확인할 수 있다. 따라서 확률표본 추출의 경우에는 전문기관이 정밀하게 확률표본추출 절차에 따라 표본을 확보하는 것이 바람직해보인다.

그러나 비확률표본 추출의 경우에는 대표성을 확보하기 위해서 고려해야 할 점이 복잡하고 유동적이며 특수하다. Feild와 동료들(2006)은 한 연구에서 확률적 표본추출 방법과 비확률적 표본추출을 사용하여 추출된 표본을 비교하였다. 비확률 표본을 추출하기 위하여 다양한 방법이 사용되었는데, 먼저 연구 대상 관련기관과 협약을 맺었다. 대상자들에 대한 정보를 얻거나 지원자들을 확보하기 위하여 관련기관의 협조를 얻는 것이 필요했다. 그리고 신문지상에 광고를 통해서 연구에 참여할 지원자들을 모집하였다. 뿐만 아니라 하위집단의 수를 목적론적으로 구성하여 우편물을 발송하였다(targeted mailings). 또한 지역 사회 도서관, 약국, 슈퍼마켓에 연구관련 공고문을 부착하였다. 마지막으로 연구에 1차적으로 선택된 참여자들에게 연구기준에 적합한 타인을 추천받았다. 분석결과 연령, 성별, 결혼기간, 임금, 직업상태, 교육수준에서는

확률표본추출과 유사한 구성을 보였고, 종교, 인종 구성에서는 차이를 나타내었다. 이러한 연구결과를 바탕으로 살펴볼 때, 비확률표본 추출을 통해 대표성 있는 표본을 구성하기 위해서는 모집단의 성격에 대한 지식이 필요하고 표본을 수집하기 위한 네트워크나 권한이 중요하다는 것을 알 수 있다.

따라서 확률표본추출의 경우에는 외부기관에서 만족도 조사를 수행하는 것이 바람직하지만 비확률표본추출에서는 정책과 관련된 지식과 자원을 확보하고 있는 내부기관이 만족도 조사를 수행하는 것이 더 효율적이라고 판단된다.

평가의 목적과 표본추출의 방법을 기준으로 적절한 평가기관을 설정하면 아래와 같다.

표 17 평가목적과 표본추출방법에 따른 평가주체

표본추출방법 \ 평가 목적	평가 목적	
	책임성 확보	정책 개선
확률표본추출	외부기관	연구자 판단
비확률표본추출	연구자 판단	내부기관

위에서 살펴본 논의대로 분석할 경우 평가의 목적과 표본추출이 요구하는 기관이 상충하는 경우가 발생한다. 이 경우 평가의 환경과 맥락을 고려하여 연구자가 평가기관을 선정해야겠지만, 평가의 목적에 조금 더 가중치를 두는 것이 바람직하다고 판단된다.

4.2. 측정의 문제

이론적 논의

정부업무평가 기본법의 시행으로 인하여 중앙행정기관의 장은 소속기관의 정책 등에 대하여 자체평가를 실시하고 있다. 다양한 평가 방법이 있지만 그 중에서 정책 만족도 조사는 중요한 평가 도구 중의 하나이다. 그런데 문제는 만족이라는 개념을 측정하기가 쉽지 않다는 것이다. 정책 수요자의 만족은 정책의 궁극적 목표로서 공공정책의 산출이 수요자에게 전달되어 발생하는 심리적 상태라고 할 수 있다. 만족은 주관적 상태이므로 그 범위와 정의를 구체화 시키기가 어렵다. 일반적으로 정확한 측정과 관련하여 신뢰성과 타당성을 필요조건으로 제시하고 있다. 그런데 신뢰성과 타당성을 사후적으로 평가하는 연구는 많지만 사전에 신뢰성과 타당성을 확보하기 위한 연구는 제대로 이루어지지 않았다. 여기서는 신뢰성과 타당성을 사전에 확보하기 위한 방법을 탐색해보려고 한다.

만족의 개념과 관련하여 이시경과 이중화(2006)는 여섯 가지 특징을 제시한다. 첫째, 만족이란 주관적인 심리 현상이나 서비스 이용 경험에 의해서 형성되는 마음의 상태이다. 둘째, 만족은 행정서비스에 대한 규범적 기대와의 비교를 통하여 발생하는 평가반응이다. 셋째, 만족은 반드시 경험이나 심리적 기대에 의해 형성된다. 넷째, 만족은 집합적인 경험의 반영으로서 다양한 측면들이 포함된다. 다섯째, 만족은 고정되고 정적인 현상이 아니라 경험이나 기대의 변화에 지속적으로 반응하는 동적인 현상이다. 여섯째, 만족은 수요와 욕구의 충족에 대한 반응이다. 이와 같은 관점에서의 만족이란 정책에 대한 전반적이고 주관적인 평가 반응의 결과로서 느낌이나 감정 뿐만 아니라 판단적 가치를 모두 포함하는 것이된다. 만족도와 관련된 선행연구들을 살펴 보더라도 만족도를 이처럼 포괄적인 정서적 반응으로 규정하고 있다는 느낌을 받게 된다.

원래 만족도에 대한 연구는 정책과 관련된 한정된 인식으로 시작되었다. Parasuraman, Zeithaml, Berry(1988)은 SERVQUAL에 기반한 만족도 연구에서 만족도를 유형성(tangibles), 신뢰성(reliability), 대응성(responsiveness), 보증성(assurance), 공감성(understanding/knowning custoers)의 5가지 차원으로 분석하였다. 유형성은 시설, 장비, 인적자원 등과 같이 서비스 수행에 필요한 자원이 준비된 정도를 의미하고, 신뢰성은 약속된 서비스를 정확하게 수행할 수 있는 능력을 의미한다. 대응성은 고객들을 도와주고 신속하게 서비스를 제공하는 태도이고, 보증성 고객들에게 신뢰와 확신을 줄 수 있는 지식과 친절함을 의미하고, 공감성은 고객들에게

개인적인 관심과 배려를 주는 자세이다. 이와 같은 맥락에서 이향수와 안형기(2011)도 행정서비스에 대한 주민만족도 평가 연구에서 만족도를 접근성, 신속정확성, 유형성, 대응성, 업무처리능력으로 측정하였다. 이에 반해 만족도를 포괄적으로 측정하는 관점에서는 정진경(2009)의 연구와 같이 기관환경(사무실의 청결, 쾌적)을 만족도 측정에 포함시키기도 하고, 이시경과 이중화(2006)의 연구에서 나타나듯이 정리정돈 상태나 주차의 편리성 등을 세부측정지표로 사용한다. 이러한 현상은 조성한과 강정석(1998)이 중앙행정기관의 민원행정서비스 만족도를 측정하기 위한 연구에서 만족도를 접근용이성(accessibility and availability of service), 편리성(convenience), 신속·정확성(speed and correctness), 쾌적성(pleasantness), 대응성(responsiveness), 공정성(fairness), 환류성(feedback)의 7가지 차원으로 포괄적인 측정을 한 것이 지속적으로 영향을 미친 것으로 보인다. 물론 만족도 측정은 다양한 차원에서 접근하는 것이 좋다. 또한 만족도와 관련된 다양한 요인들을 측정하면 추세연구와 회귀분석을 하는데 있어서 통제변수로 사용할 수 있기 때문에 유용하다. 그러나 정책평가로써의 만족도 측정에 있어서 정책과 직접적인 관련성이 떨어지는 차원을 무리하게 측정에 포함시키면 그 측정지표는 불필요한 측정오차를 포함하게 된다. 따라서 만족도와 관련하여 다양한 요인을 측정하는 것은 추가적인 연구를 위해서 유용할 수 있지만, 만족도의 측정지표는 신중하게 선택하여 정책과 직접적인 관련이 있는 요인에 관하여 제한적으로 구성하는 것이 바람직해 보인다.

만족도 조사 분석

병무청의 징병검사 만족도는 요소만족도와 전반적 만족도를 가중평균하여 도출하는데, 요소만족도와 관련된 문항에서 쾌적성과 용모단정성을 묻고 있다. 그런데 ‘깨끗하다’, ‘단정하다’는 용어는 판단의 기준이 모호할 뿐만 아니라 정책과 직접적인 관련성이 떨어져 보인다.

징병검사 만족도(병무청)

징병검사장 환경	만 족 / 공 감 중 -----> 불만족/공감 <----- 안함										
1) 쾌적성 - 징병검사장의 시설들은 깨끗하게 관리되고 있었습니까?	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①	①
2) 용모단정성 - 징병검사장에서 근무하는 담당 직원들의 용모는 단정하였습니까?	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①	①

Oliver(1981)는 인식된 서비스의 질(perceived quality)과 만족(satisfaction)을 구분하였다. 인식된 서비스의 질은 대상에 대한 지속적인 태도를 의미하지만 만족은 즉흥적인 인상이나 기대불일치로 인하여 발생하는 일시적인 감정상태라고 말한다. 비록 만족도 조사가 만족이라는 용어를 사용하고 있지만, 만족도 조사가 측정하려고 하는 것은 만족이 아니라 인식된 서비스의 질이라고 할 수 있다. 따라서 특정한 인상으로 인하여 일시적인 감정에 영향을 미칠 수 있는 지표는 만족도를 도출할 때 제외하는 것이 필요하다.

실제 진행되는 있는 만족도 조사는 다양한 양태를 보이고 있고 체계적으로 분류하여 현황을 분석하기는 어렵지만, 몇몇 만족도 조사에서는 만족도를 단일차원으로 측정하고 있다.

어린이집 이용자 만족도(보건복지부)

자녀가 다니는 어린이집의 전반적인 만족도는 어느 정도입니까? ①매우 불만족 ②불만족 ③보통 ④대체로 만족 ⑤매우 만족
--

보육·양육지원 만족도(보건복지부)

정부는 2013년부터 0~5세 어린이집·유치원 미이용 영유아에게는 가정양육수당, 0~2세는 보육료, 3~5세는 누리과정을 지원하고 있습니다. 귀하는 이렇게 0~5세 모든 영유아에게 보육료·교육비 또는 가정 양육수당을 지원하는 정책에 어느 정도 만족하십니까? ①매우 불만족 ②불만족 ③보통 ④대체로 만족 ⑤매우 만족
--

유아학비 지원 학부모 만족도(교육부)

유아학비 지원 전반에 대한 학부모님의 만족도는 어느 정도입니까? ①매우 만족 ②다소 만족 ③보통 ④별로 만족하지 못함 ⑤전혀 만족하지 못함
--

대입정보포털 만족도 조사(교육부)

대입정보포털 서비스는 이용하기에 어땠나요?

	매우 만족	만족	보통	불만족	매우 불만족
진로정보	⑤	④	③	②	①
대학정보	⑤	④	③	②	①
학과정보	⑤	④	③	②	①
전형정보	⑤	④	③	②	①
학습진단	⑤	④	③	②	①
온라인상담	⑤	④	③	②	①

만족도 지표와 관련된 설문지를 만족도 측정의 차원(dimension)과 차원을 구성하고 있는 질문(item)으로 분류하면 다음과 같다

표 18 만족도 측정의 차원과 문항

	단일 차원 측정	다차원 측정(불완전)
단일문항 측정	유아학비 지원 학부모 만족도(교육부) 어린이집 이용자 만족도(보건복지부) 보육·양육지원 만족도(보건복지부)	생애설계서비스 만족도(고용노동부) 직업소개종사자 교육만족도(고용노동부) 일학습병행제 만족도(고용노동부) 개발된자료의수업개선 적합성에 대한 교원 만족도(교육부) 지역 평생교육 활성화 사업 주민참여 만족도(교육부) 지방대학 재정지원 정책 수요자 만족도(교육부) 고등교육 재정지원 사업 재편에 대한 정책 수요자 만족도-BK21플러스사업(교육부) 고등교육 재정지원 사업 재편에 대한 정책 수요자 만족도-지방대학특성화 사업(교육부) 일반농산어촌개발 사업관계자 고객 만족도(농림축산식품부) 징병검사 수검자 체감 만족도(병무청) 사회복무요원 소양교육 만족도(병무청) 산림치유프로그램 만족도(산림청) 숲태교 만족도(산림청) 소통경험 소비자 만족도(식품의약품안전처) 식품의 영양표시 소비자 활용도(식품의약품안전처) 한부모가족 지원사업 만족도(여성가족부) 청소년의 건강한 성장 지원 성과목표 달성을 위해 국립청소년수련원 서비스에 대한 만족도(여성가족부) 청소년가족연계서비스 참여 만족도(여성가족부) 찾아가는 가정폭력 예방교육 수혜자 만족도(여성가족부) 통합환경관리제도 인지도 및 만족도(환경부)

	단일 차원 측정	다차원 측정(불완전)
다문항 측정	중·고교에서 직업 체험 교육 참여자 만족도(교육부) 대입정보포털 이용 만족도(교육부) 회사채 투자자 만족도(금융위원회) 식품안전체감도(식품의약품안전처) 연안정비사업 고객만족도(해양수산부) 민간기업 기후변화 리스크평가 활용 만족도(환경부)	치안고객 만족도(경찰청) 청년창직활성화 만족도(고용노동부) 자유학기제 운영 학교 학생 학교생활 만족도(교육부) 학습자 참여 활동 중심 교과서 단원모형에 따른 자기 주도 학습능력 향상도(교육부) 학교스포츠클럽 활동 17시간 이상 활동 등록률 및 만족도(교육부) 학교폭력 예방교육 만족도(교육부) 고등교육 재정지원 사업 재편에 대한 정책 수요자 만족도-특성화전문대학육성사업(교육부) 농어촌주민의 정주 만족도(농림축산식품부) 숲길 이용자 만족도(산림청) 산림교육만족도(산림청) 청소년의 건강한 성장 지원 성과목표 달성을 위해 프로그램 공모사업 서비스에 대한 만족도(여성가족부) 청소년의 건강한 성장 지원 성과목표 달성을 위해 방과후아카데미 서비스에 대한 만족도(여성가족부) 아동·청소년 성폭력 예방교육 만족도(여성가족부) 센터심리지원 만족도(여성가족부) 조달물자품질만족도(조달청) 도시가치 및 비전에 대한 주민 이해도 제고(행정중심복합도시건설청)

※국무조정실에 제출된 설문지만을 대상으로 함.

정책적 제언

사회적 개념의 측정은 비유적으로 볼 때, 고고학에서 유물을 발굴하는 과정과 유사하다. 도자기를 우리가 측정하려는 만족도 개념이라고 비유할 때, 고고학자들이 도자기를 땅속에서 단번에 끄집어내지 못하듯이 사회과학자들도 만족도라는 개념을 하나의 질문으로 단번에 도출할 수 없다. 사회과학자들은 만족도라는 개념이 어떻게 생긴

지 모르기 때문에 이론적으로 드러난 윤곽을 따라 다양한 차원에서 조심스럽게 그 형태를 발굴해나가야한다. 그러므로 만족도의 측정은 기존 이론적 연구에서 이루어진 논의를 기반으로 하여 다차원적으로 접근하여 측정하는 것이 바람직하다. 현재 대부분의 만족도 조사들이 만족도를 다차원적으로 측정하고 있지만, 지표의 타당성을 확보하기 위해서는 기존의 연구들에서 이루어진 이론적 논의들을 설문지에 보다 정밀하게 반영할 필요가 있다.



그림 4 만족도 측정 접근방법

4.3. 모집단 설정과 구성적 특성에 따른 만족도

이론적 논의

모집단의 설정은 만족도의 조사대상을 결정하는데 있어서 중요한 기준이 된다. 이와 관련하여 앞에서 모집단의 범주와 분석수준을 잘못 설정할 경우 발생할 수 있는 문제에 대해서 살펴보았다. 만족도 조사에 있어서의 모집단은 정책대상집단(policy target group)이다. 정책대상집단은 직·간접적으로 정책의 영향이나 적용을 받는 사람이나 집단으로서 정책을 통해서 행동의 변화가 발생하는 대상을 의미한다(남궁근, 2012). 정책대상집단은 정책유형에 따라 개인, 특정 집단, 국민전체 등 다양한 분석수준으로 설정될 수 있다. 또한 설정된 정책대상집단은 반응하는 행태에 따라 다양한 구성적 특성을 지니게 된다.

정정길과 동료들(2016)은 정책대상집단을 성질에 따라 수혜집단과 정책비용부담자로 구분한다. 정책수혜집단은 정책에서 제공하는 서비스나 재화를 받는 자들로서 주로 정책을 통하여 혜택을 받는 집단이고 정책비용부담자들은 정책실행과정에서 피해를 입거나 희생을 당하는 사람들이다. 그러나 이처럼 이분법적으로 정책대상집단을 구분하는 것은 현재 복잡해진 정책환경을 단순화시킬 우려가 있으며 구분자체가 상호배타적(mutually exclusive)이지도 않다. 실제 서비스나 재화를 제공받는 자들 중에도 정책을 혜택이 아닌 부담으로 느끼는 부류가 있을 수 있고, 정책 수혜집단과 정책비용부담자들의 정책에 대한 태도가 구분된 집단에 귀속적이지 않을 수 있다.

Ingram, Schneider, deLeon(2007)은 사회적 형성이론(social constraction theory)를 기반으로 정치적 영향력(political power)과 사회적 인식(positive or negative image of deservedness)에 따라 정책대상집단을 구분하고 있다.

표 19 사회적 인식과 정치적 영향력에 따른 정책대상집단 구분

	사회적 인식	
	긍정적	부정적
정치적 영향력 ↑	수혜집단(advantaged)	이익집단(contenders)
	의존집단(dependents)	일탈집단(deviants)

※출처: Ingram, Schneider, deLeon(2007)

수혜집단은 정치적으로도 영향력을 가지고 있고 사회적 지위에 대한 자격에 대해서

도 긍정적인 평가가 이루어지는 집단이다. 이익집단은 정치적 영향력은 강력하지만 이기적이고 신뢰성이 떨어진다고 인식되는 집단이다. 의존집단은 정치적 영향력은 약하지만 사회적으로 연민과 동정의 대상이 되는 집단이다. 일탈집단은 정치적 영향력이 약하고 사회적으로 불균형적인 분배와 제재를 받는 집단이다. 이와 같은 구분은 정책과 관련하여 다양한 정책대상집단이 존재하고, 각 집단별로 다른 정책들이 적용될 수 있다는 내용을 보여준다. 그러나 특정정책에 대해서 그 정책과 관련된 정책대상집단이 어떻게 형성되고 구분되는지에 대해서는 설명하지 못한다.

따라서 정책대상집단의 태도는 정책수용성으로 살펴보는 것이 바람직해 보인다. 어떠한 정책이든 정책대상집단의 다양한 이해관계를 완벽하게 충족시키지는 못한다. 그 결과 정책대상집단 내에서 정책에 대한 의견 불일치가 발생할 수 있다. 정책수용성은 정책대상집단의 행태가 단순하게 행위규정에 부합하도록 행동하는 것이 아니라, 내적인 가치체계와 태도의 구체적인 변화까지를 포함하는 선호체계라고 할 수 있다 (Duncan, 1981). 이와 관련하여 정주용(2008)은 정책 수용성을 6가지 양상으로 나누어 구분하고 있다.

표 20 정책수용성의 구분

	적극적 찬성	피동적 중립	활동적 반대
이해적(효용적) 근거	I	II	III
이념적(미신적) 근거	IV	V	VI

※출처: 정주용(2008) 재인용

I의 영역은 정책수용성이 확보된 상태이다. 정책으로 인한 개인이나 집단의 효용 증가가 정책에 대한 적극적인 찬성을 유도하기 때문이다. II의 영역은 정책이 자신의 이해와 상관이 없거나 정책에 대해 무관심하여 중립적인 태도를 보이는 곳이다. III의 영역은 정책대상 집단이 정책으로 인하여 침해된 손실 때문에 적극적으로 정책에 반대하는 영역이다. IV의 영역은 이해관계와 상관없이 정책에 대한 이념적 근거를 통해 적극적으로 정책에 찬성하는 경우이다. V의 영역은 현실적으로 잘 나타나지는 않지만, 정책대상집단이 자신의 이념적 근거에 따라 정책을 판단하였을 때, 명확한 판단을 할 수 없는 경우에 발생하는 경우이다. VI의 영역은 정책 수용성이 가장 떨어지는 영역으로 이념적 근거에 의해 적극적으로 정책에 반대하는 입장이기 때문에 정책대상집단의 신념이 변화하기 전에는 정책의 수용성을 확보하기 어렵다.

이와 같은 범주를 기반으로 하여 정책대상집단이 정책에 대한 만족도를 평가하는

과정에서 정책 내용 외적으로 편견이 개입할 수 있는 집단을 구성해볼 수 있다. I, IV 집단은 정책에 적극적으로 찬성하기 때문에 만족도 조사에서 긍정적으로 답변할 가능성이 높고, III, VI 집단은 정책에 대해 부정적 인식이 강할 수 있다. II, V 집단은 정책에 대해 무관심하거나 이해도가 떨어질 것으로 예상할 수 있다. 따라서 모집단을 구성할 때, 이론적이나 선형적으로 I, IV 집단에 속할 것으로 판단할 수 있는 응답자가 많이 포함되도록 모집단을 설정하면 정책에 대한 평가가 다른 설정에 비하여 긍정적으로 나타날 수 있다. 더욱 심각하게 고려해야 할 사항은 IV, VI 집단이다. I, II, III 집단은 정책에 대한 이해관계에 따라 정책에 대한 태도가 변화하는 집단이기 때문에 정책으로 정책대상집단의 태도를 변화시킬 수 있다. 또한 V 집단은 정책에 대한 설명과 이해가 동반되면 정책에 대한 태도가 바뀔 수 있는 집단이다. 그러나 IV, VI 집단은 정책 집행자가 정책대상집단의 태도에 영향을 미칠 수 있는 여지가 적으며, 그 집단의 태도 또한 잘 변하지 않고 강력하다. 정책 평가의 목적으로 이루어지는 만족도 조사가 만족도 조사결과를 가지고 정책을 개선하거나 책임성을 확보하는 수단으로 사용된다고 했을 때, IV·VI 집단의 조사결과를 다른 집단과 통합하여 만족도를 도출하면 정책 그 자체에 대한 평가가 왜곡될 수 있다. 왜냐하면 IV·VI 집단은 집단 구성원들이 선형적으로 가지고 있는 이미지나 신념에 의해서 정책을 바라보기 때문에 정책의 개선이나 변화만으로는 그 이미지나 신념을 바꾸기가 쉽지 않으며, 그러한 목적에서 정책을 수정하는 것은 바람직하지 않다. 따라서 정책평가의 모집단을 설정할 때, IV·VI 집단이 I, II, III, V 집단과 함께 섞여있지 않은지에 대한 분석이 필요하다. 만약 분석을 통해서 IV, VI 집단이 만족도 조사에 섞여 있는 것이 확인되면 이 집단에 대한 결과는 따로 분리해서 보고하거나 통제하여 만족도 조사를 도출해야 한다.

만족도 조사 분석

현재 시행되고 있는 여러 만족도 조사에서 설문조사의 모집단에 여러 하위집단이 함께 포함된 경우가 있다. 이러한 설문조사를 대상으로 하위집단에 대한 통제가 이루어지고 있는지, 설문지를 분리하여 사용하고 있는지에 따라 만족도 조사를 분류하면 다음 표와 같다.

표 21 하위집단 통제여부와 집단 간 설문지 분리여부

	설문지 분리	설문지 통합
통제	고등교육 재정지원 사업 재편에 대한 정책 수요자 만족도-특성화전문대학육성사업(교육부) 학교스포츠클럽 활동 17시간 이상 활동 등록률 및 만족도(교육부) 아동·청소년 성폭력 예방교육 만족도(여성가족부)	직업소개종사자 교육만족도(고용노동부) 대입정보포털 이용 만족도(교육부) 고등교육 재정지원 사업 재편에 대한 정책 수요자 만족도-지방대학특성화 사업(교육부) 지방대학 재정지원 정책 수요자 만족도(교육부) 통합환경관리제도 인지도 및 만족도(환경부) 일반농산어촌개발 사업관계자 고객 만족도(농림축산식품부)
비통제		일학습병행제 만족도(고용노동부)

위에서 나타난 4가지 유형 중에서 가장 위험한 것은 하위집단에 대한 통제없이 설문지를 통합적으로 사용하고 있는 경우이다.

‘일학습병행제 만족도’에 대해서 살펴보면, 정책 수혜자인 모집단에 사업주, 기업현장교사, 학습근로자가 포함되어 있다. 그런데 이들은 이념적 근거나 이해적 근거에 의해서 정책에 대한 태도가 명확하게 구분되는 집단으로 볼 수 있다. 따라서 집단 간 구분을 통제할 필요성이 있다. 비록 설문지에서 만족도 문항 앞에 집단 구분이 되어 있다고는 하지만 응답자는 자신의 집단에 대한 판단없이 첫 번째로 나오는 집단의 항목에 설문을 할 우려가 있다. 세 집단에 대한 설문은 그 구성과 형식이 동일하기 때문에 설문지를 따로 분리하여 진행하던지 응답 항목에 대한 자세한 안내와 도움이 요구된다.

하위집단에 대한 통제는 있지만 설문지가 분리되어 있는 경우도 유의해야한다.

‘고등교육 재정지원 사업 재편에 대한 정책 수요자 만족도-특성화전문대학육성사업’은 교직원, 산업체, 재학생을 포함하여 모집단을 설정하였다. 그런데 만족도를 도출하는 과정에서 세 집단을 분리하지 않고 통합하여 점수를 도출하면 문제가 될 수 있다. 이 설문조사에서는 재학생, 교직원, 산업체 각각의 집단에 서로 다른 설문지를 사용하고 있다. 재학생들은 교육내용, 교육 운영 및 지원, 유형성으로 만족도를 측정하였고, 산업체를 대상으로는 직무능력성취도, 졸업생의 역량, 산학협력으로 만족도를 측정하였다. 교직원은 사업비 편성 및 집행, 사업 운영, 사업성과를 기준으로 만족도를 평가한다. 물론 이러한 영역들이 전체적으로 특성화전문대학육성사업을 구성하지

만, 각 분야의 점수를 통합하여 하나의 점수로 도출하는 것은 문제가 있어 보인다. 따라서 개별적으로 집단 별로 점수를 도출하여 각 영역에 대한 부분적 평가를 진행하는 것이 바람직하다.

‘아동·청소년 성폭력 예방교육 만족도’는 아동·청소년, 학부모를 대상으로 한다. 각 집단에 대한 설문지는 분리되어 있고, 문항의 구성과 설계가 동일하기 때문에 큰 문제가 없어 보일 수 있으나, 아동·청소년과 학부모는 이념적 근거에 따라 정책에 대한 태도가 다를 수 있는 집단이다. 따라서 점수를 분리하여 도출할 필요가 있다.

‘학교스포츠클럽 활동 17시간 이상 활동 등록률 및 만족도’는 교사와 학생을 대상으로 각 집단을 대상으로 작성된 설문지를 사용한다. 그런데 만족도를 구성하는 차원과 질문문항이 서로 다르다. 따라서 만족도를 통합적으로 도출하는 것은 바람직하지 않다.

하위집단은 통제되어 있지만 설문지가 통합되어 있는 경우에는 만족도 점수를 통합적으로 산출하는 것이 바람직한지 고려해보아야한다.

‘지방대학 재정지원 정책 수요자 만족도’에는 학생, 교직원, 교수가 함께 모집단에 포함되어 있다. 교직원과 교수는 정책에 대한 이해관계가 다를 수 있으며, 학생은 이념적으로 정책에 중립적일 가능성이 높다. 따라서 이들 집단의 만족도를 통합적으로 산출하지 말고 개별적으로 산출해야 결과의 정책적 활용성이 더 높아진다.

‘직업소개종사자교육만족도’는 정책수혜자를 모집단을 설정하고 있는데, 사원부터 임원까지 너무나 성격을 달리하는 조직 계층을 포함하고 있다.

‘대입정보포털 이용 만족도’는 학생, 학부모, 교사, 기타를 포함한 모집단을 설정하고 있다.

‘지방대학특성화 사업’은 학생, 직원, 교수를 통틀어 모집단으로 설정하였다.

‘지방대학 재정지원 정책 수요자 만족도’도 학생, 직원, 교수를 통틀어 모집단으로 설정하고 있다.

‘일반농산어촌개발 사업관계자 고객 만족도’에는 공무원, 공사직원, 추진위원장, 추진위원, 컨설팅 회사, 전문가를 모두 포함하여 만족도를 도출하고 있다.

‘통합환경관리 제도 인지도 및 만족도’는 사업장, 공공기관, 전문가, 환경기술인을 모두 포함하여 만족도를 측정하고 있다.

정책적 제언

정부의 정책은 대부분 광범위한 영향력을 지니기 때문에 이해관계자가 많으며 정책적 특성도 다양하다. 이와 같은 이유로 만족도 측정에 다양한 정책 대상자들을 포함시키는 것은 불가피한 현상으로 판단된다. 그렇지만 정책에 대한 통합적인 만족도 지표를 도출하기 위해 이들 각 집단의 만족도를 무분별하게 변환하여 합산하는 방식은 바람직하지 않다. 기본적으로 각 집단별·영역별 분리된 만족도 산출이 요구된다.

4.4. 표본 추출 방법의 문제

이론적 논의

국무 조정실의 만족도 지표 준수사항에서는 만족도 조사대상 선정과 관련하여 무작위로 대상자를 구성하거나 인구·통계학적 특성을 반영하여 한 집단에 편향되지 않게 표본을 구성하도록 요구하고 있다. 그래서 다수의 만족도 조사에서 무작위 표본 추출을 표본추출방법으로 선정하였다고 밝히고 있는데, 이에 대한 검토가 필요하다.

무작위표본추출은 확률표본추출방법으로 추출방법의 기초가 되는 가장 기본적인 방법이다. 무작위라는 뜻은 설계자가 임의대로 표본을 추출한다는 뜻이 아니라 설계자의 의도가 추출될 표본에 전혀 영향을 주지 않아 모집단 구성원이 표본으로 추출될 가능성이 완벽하게 공평한 추출방법을 의미한다. 일반적으로 단순확률추출법, 단순임의추출 등(simple random sampling)의 용어로 사용된다. 따라서 무작위표본추출을 진행하기 위해서는 모집단이 명확해야하고 표본프레임이 설계되어 있어야 진행이 가능하다. 그러나 국무조정실에 제출된 조사계획들을 살펴보면 일반적으로 알려진 무작위표본추출과 다른 방식으로 진행되는 조사들이 있다.

만족도 조사 분석

‘숲길 이용자 만족도(산림청)’는 숲길 현장출구에서 6월~7월, 9월~10월에 걸쳐 표본추출을 진행하고 있다. 출구에서 퇴장하는 이용자들을 무작위로 선정한다고 해서 선정된 표본을 무작위표본추출로 만들어졌다고 말할 수는 없다. 모집단을 숲길 전체 이용자로 보아야하는데, 표본 조사가 6월~7월, 9월~10월에만 이루어졌기 때문에 다른 기간에 숲길을 이용한 이용자들이 배제된다. 정확한 무작위표본추출을 사용하기 위해서는 전체 이용자의 리스트를 표본프레임으로 만들고 그 표본 프레임에서 무작위 추출을 진행하여야한다. 따라서 이러한 표본추출 방법은 무작위표본추출 방법이 아니라 비확률표본추출방법으로 보아야한다. ‘숲길 이용자 만족도’의 경우 이용자 리스트를 구성한다는 것이 사실상 불가능하기 때문에 이 경우에는 비확률표본추출을 정교하게 설계하여 진행할 수밖에 없다.

‘산림복지서비스 프로그램 참여자 만족도(산림청)’도 9월~10월에 걸쳐 현장출구에서 표본추출을 진행하고 있는데, ‘숲길 이용자 만족도’도 같이 비확률표본추출방법으로 판단된다. ‘산림복지서비스 프로그램 참여자 만족도’는 ‘숲길 이용자 만족도’와는 달리

교육 전에 참가자의 정보를 확보할 수 있을 것으로 판단된다. 이를 통해 참여자들의 리스트가 구성되면 그것을 표본프레임으로 사용하여 무작위표본추출을 진행할 수 있다. ‘치안고객 만족도(경찰청)’의 경우, 분야별 업무처리 과정에서 서비스 이용자들의 리스트를 만들고 그 리스트에서 무작위 표본추출을 진행하고 있다.

정책적 제언

무작위 표본추출을 진행하기 위해서는 표본프레임의 설계가 선행되어야 한다.

4.5. 설문지 어구

이론적 논의

정치학에서는 대중들의 정치적 선호나 선택이 프레이밍(framing)에 의해서 결정된다고 말한다. 프레임(frame)이란 포괄적 의미에서는 특정한 현상이나 사안에 대해서 질서와 의미를 부여하는 마음의 구조이며, 프레이밍은 심리적 관점에서 그 구조를 형성하기 위하여 대화나 소통을 사용하여 해석적 구조 체계(interpretive structure)를 형성하는 것이다(박원호·안도경·한규섭, 2013). 따라서 프레이밍을 설정하는 방식에 따라 시민들의 선호가 변화할 수 있다.

이와 같은 과정을 적용하여 설문지의 어구설정이 어떻게 프레이밍으로 작용할 수 있으며, 그에 따라 시민들의 만족도가 어떤 영향을 받는지를 살펴보기 위해 프레이밍에 대해 좀 더 자세히 살펴볼 필요가 있다. 프레이밍 이론에 따르면 특정한 이슈는 다양한 관점에서 바라볼 수 있으며, 여러 가지 가치와 제약조건들 속에서 개인에게 의미를 형성한다.

정책에 대한 선호 = $\sum V_i \times W_i$ V_i : 가치·믿음, W_i : 가중치($\sum W_i=1$)

Chong과 Druckman(2007)은 개인의 선호가 위와 같이 가치와 믿음들의 합으로 결정된다고 본다. 한 가지 정책은 여러 측면을 가치와 이슈를 가지고 있다. 예를 들어 무상급식의 경우 보편적 복지 측면에서는 무상급식을 긍정적으로 바라볼 수 있다(V_1). 그러나 재정적 측면에서 바라보았을 경우에는 세금부담과 재정지출의 증가라는 부정적 측면이 있다(V_2). 만일 개인이 부정적 측면에 더 큰 가중치를 두고 있다면, 무상급식 정책에 대하여 부정적인 태도를 보이게 된다.

만일 이와 같은 선호 결정의 매커니즘이 정확한 정보와 지식에 의해서 형성된 시민들의 가치와 믿음을 바탕으로 개인의 확신과 소신에서 비롯된 가중치로 인하여 작동한다면 큰 문제가 없겠지만, 시민들은 현실적 조건들로 인하여 정책에 대한 정보가 부족하기 때문에 자신들의 정치적 선호에 기반하여 일관된 정치적 선택을 제대로 할 수 없다. 따라서 대의민주주의 체제 속에서 정치적 선택지는 엘리트나 정당에 의해서 제공되고 시민들은 제공된 선택지 중에서 자신들의 정치적 신념에 부합하거나 사회·경제적 효용을 높일 수 있는 선택을 하게 된다. Zaller(1992)에 따르면 개인들의 정책 선호와 의견은 즉흥적인 고려들에 의해 형성되며 이는 시민들이 정책 사안에 대하여 확고하고 체계적인 태도를 가지고 있는 것이 아니라 정책에 대한 접근방식이나 강조

점이 바뀔에 따라 쉽게 태도가 달라질 수 있다는 사실을 보여준다.

이러한 프레이밍 이론은 설문조사에도 그대로 적용될 수 있다. 설문에서 정책에 대한 개인의 선호는 설문지 어귀 구성을 통하여 특정 가치나 측면을 부각 시킬 경우 개인의 선호를 교란시킬 수 있다. 정치적 태도와 관련된 프레이밍은 강한 프레이밍일 때가 약한 프레이밍일 때보다 정책선호의 왜곡이 심하고, 경쟁적인 프레이밍 환경에서는 사실상 정책선호의 왜곡이 사라진다(Chong&Druckman, 2007). 그러나 설문에서는 경쟁적으로 프레임을 제시할 수 없다. 따라서 최대한 프레이밍을 배제한 설문지를 구성할 필요가 있다.

만족도 조사 분석

여성가족부의 ‘한부모가족 아동 양육비 등 지원사업’의 만족도 조사의 설문지에서는 정책의 전반적 만족도에 대하여 다음과 같이 질문하고 있다.

한부모가족 아동양육비 등 지원사업」은 한부모가족의 안정적 자녀 양육을 위하여 아동양육비, 중고등학생 학용품비, 생활보조금 등을 지원하고 있습니다.
위 사업에 대해 전반적으로 만족하십니까?
① 매우 만족 ② 만족 ③ 보통 ④ 다소 불만족 ⑤ 매우 불만족

질문에서 한부모가족의 안정적 자녀 양육을 위하여 아동양육비 등을 지원하고 있다는 정책의 긍정적인 측면을 제시하고 그 사업에 대한 만족도를 물어보고 있다. 이 경우 응답자는 질문자에 의해 정책에 대한 태도를 긍정적인 방향으로 프레이밍 당하기 쉽기 때문에 응답결과는 왜곡될 우려가 있다.

보건복지부의 유아학비 지원에 대한 만족도 조사에서는 정책의 전반적인 만족도를 다음과 같이 질문하고 있다.

유아학비 지원 전반에 대한 학부모님의 만족도는 어느 정도입니까?()
① 매우 만족 ② 다소 만족 ③ 보통 ④ 별로 만족하지 못함 ⑤ 전혀 만족하지 못함

위 질문에서는 정책 만족도에 대한 특별한 프레이밍이 나타나지 않고 있다.

정책적 제언

프레이밍 이론에 따르면 개인의 정책에 대한 선호는 설문당시에 접할 수 있는 정보에 의해서 크게 영향을 받는 것으로 나타난다. 따라서 설문지 상에서 개인의 태도를 유도할 수 있는 정보를 질문과 가까이 배치하게 되면 이는 유도질문의 문제를 발생시킨다.

4.6. 설문지 배열

이론적 논의

고길곤(2014)은 설문도구 개발 단계에서 발생할 수 있는 응답오차에 대한 연구가 이루어지지 않고 있는 문제점을 제기하며 설문순서에 따라 발생할 수 있는 오차의 문제를 지적한다. 설문의 배열은 응답자의 심리적 잔향(psychological resonance)효과를 통해서 응답에 영향을 미칠 수 있는데, 예를 들어 선행 설문문항의 응답에 극단적인 선택을 했을 경우 다음 문항에 대해서는 상대적으로 온건한 선택을 하는 현상이 나타날 수 있다. 이 경우, 질문자는 설문순서의 배열을 통하여 결과를 의도적으로 왜곡시킬 수 있게 된다. 따라서 선행연구의 분석을 통하여 설문지 배열로 인하여 발생할 수 있는 오차와 편의들을 살펴보고 설문지 구성을 객관적이고 체계적으로 구성할 필요가 있다.

이주현(2011)은 설문 순서가 응답 결과에 미치는 영향을 분석하기 위하여 김포공항 주변 거주민의 소음에 대한 설문조사 자료를 분석하였다. 소음에 대한 성가심에 대한 반응이 설문순서에 따라 어떻게 달라지는지에 대하여 분석하였는데, 설문초기에 성가심에 대한 질문을 접한 응답자들이 설문 중간 단계에서 성가심에 대한 질문을 접한 응답자들 보다 비행기 소리에 더 성가심을 느끼는 것으로 나타났다. 특히 연립주택 거주자의 비율이 높은 지역에 사는 주민들은 이 차이가 통계적으로 유의하게 나타났다. 연구자는 연립주택은 방음이 잘 안되는 노후주택이 많기 때문에 항공기 소음에 더 많은 피해를 보게 되어, 위와 같은 결과가 도출되었다고 분석하였다. 따라서 만족도와 관련된 설문지를 구성할 때, 대상집단 중에 정책을 통하여 큰 혜택이나 피해를 보는 집단이 있을 경우 만족도와 관련된 설문문항을 앞으로 배치하면 그 집단의 응답 결과가 과장되어 나타날 우려가 있다.

만족도 조사 분석

정책 만족도 설문조사는 대부분 설문지의 문항이 적은 편이다. 그래서 만족도 관련 설문문항이 시작된 설문조사를 정리해보았다.

표 22 만족도 설문문항으로 시작된 설문조사

생애설계서비스 만족도(고용노동부)
 일학습병행제 만족도(고용노동부)
 고등교육 재정지원 사업 재편에 대한 정책수요자 만족도-BK21플러스 사업(교육부)
 소통경험 소비자 만족도(식품의약품안전처)
 한부모가족 지원사업 만족도(여성가족부)

안명식, 한인수, 오홍석(2016)은 직무만족도가 전반적인 만족도 질문과 세부적 만족도 질문의 배치에 따라 응답결과가 어떻게 변화하는지를 분석하였다. 분석 결과 전반적인 만족도 질문을 먼저하고 세부적 만족도 질문을 나중에 배치한 질문이 그 반대의 경우보다 응답 간의 상관성이 높게 나타났다. 따라서 연구자들은 전반적인 만족도에 중점을 둘 것인지 아니면 세부 영역별 만족도에 더 중점을 둘 것인지에 따라 설문지의 질문 순서를 다르게 구성할 필요가 있다.

표 23 설문지 배열 유형구분

전반적 ---> 세부적
생애설계서비스 만족도(고용노동부)
청년창직활성화 만족도(고용노동부)
일학습병행제 만족도(고용노동부)
지역 평생교육 활성화 사업 주민참여 만족도(교육부)
농어촌주민의 정주 만족도(농림축산식품부)
일반농산어촌개발 사업관계자 고객 만족도(농림축산식품부)
산림치유프로그램 만족도(산림청)
숲태교 만족도(산림청)
소통경험 소비자 만족도(식품의약품안전처)
한부모가족 지원사업 만족도(여성가족부)
청소년의 건강한 성장 지원 성과목표 달성을 위해 프로그램 공모사업 서비스에 대한 만족

도(여성가족부) 청소년의 건강한 성장 지원 성과목표 달성을 위해 국립청소년수련원 서비스에 대한 만족도(여성가족부) 아동·청소년 성폭력 예방교육 만족도(여성가족부) 도시가치 및 비전에 대한 주민 이해도 제고(행정중심복합도시건설청)
세부적--->전반적
치안고객 만족도 조사(경찰청) 자유학기제 운영 학교 학생 학교생활 만족도(교육부) 고등교육 재정지원 사업 재편에 대한 정책 수요자 만족도-지방대학특성화 사업(교육부) 지방대학 재정지원 정책 수요자 만족도(교육부) 고등교육 재정지원 사업 재편에 대한 정책 수요자 만족도-특성화전문대학육성사업(교육부) 징병검사 수검자 체감 만족도(병무청) 사회복무요원 소양교육 만족도(병무청) 식품안전체감도(식품의약품안전처) 숲길 이용자 만족도(산림청) 산림교육만족도(산림청) 청소년가족연계서비스 참여 만족도(여성가족부) 찾아가는 가정폭력 예방교육 수혜자 만족도(여성가족부) 조달물자품질만족도(조달청) 민간기업 기후변화 리스크평가 활용 만족도(환경부) 통합환경관리제도 인지도 및 만족도(환경부)
세부적만
직업소개중사자 교육만족도(고용노동부) 개발된자료의수업개선 적합성에 대한 교원 만족도(교육부) 학교스포츠클럽 활동 17시간 이상 활동 등록률 및 만족도(교육부) 학습자 참여 활동 중심 교과서 단원모형에 따른 자기 주도 학습능력 향상도(교육부) 고등교육 재정지원 사업 재편에 대한 정책 수요자 만족도-BK21플러스사업(교육부) 회사채 투자자 만족도(금융위원회) 청소년의 건강한 성장 지원 성과목표 달성을 위해 방과후아카데미 서비스에 대한 만족도(여성가족부) 연안정비사업 고객만족도(해양수산부)
전반적만
중·고교에서 직업 체험 교육 참여자 만족도(교육부) 대입정보포털 이용 만족도(교육부) 유아학비 지원 학부모 만족도(교육부) 어린이집 이용자 만족도(보건복지부) 보육·양육지원 만족도(보건복지부)

정책적 제언

대부분의 설문조사들이 정책 인식여부, 정책 경험 내용, 응답자의 사회경제적 조건 등과 같은 질문으로 설문을 시작하고 있었다. 그런데 정책 만족도와 관련된 정책들이 대부분 정책들로 인하여 혜택을 받고 있는 사람들을 대상으로 하고 있다는 점을 고려할 때, 설문을 정책에 대한 만족도로 바로 시작하는 것은 만족도가 왜곡되어 도출될 우려가 있다.

설문지 배열은 정책 평가의 목적과 상황에 따라 달라질 수 있지만, 설문배열이 전반적인 만족도로 시작하여 세부적인 만족도로 배치된 설문조사는 위의 내용(만족도 질문이 앞에 배치될 경우 결과가 과장될 수 있다.)과 결합하여 문제를 발생시킬 수 있다. ‘생애설계서비스 만족도(고용노동부)’, ‘일학습병행제 만족도(고용노동부)’, ‘소통 경험 소비자 만족도(식품의약품안전처)’, ‘한부모가족 지원사업 만족도(여성가족부)’가 여기에 해당한다. 정책혜택을 입는 응답자들이 앞에 배치된 만족도 질문에 과장되게 반응하고 그것이 세부만족도로 연결될 경우, 조사결과가 왜곡될 우려가 있기 때문이다.

4.7. 점수 변환의 문제

이론적 논의

점수를 변환하였을 때, 왜곡의 문제가 발생하는 것을 앞에서 살펴보았다. 일반적으로 점수는 변환하지 않고 그대로 사용하는 것이 좋다. 그러나 편의상 점수를 변환하거나 통계처리의 목적에서 점수를 다른 기준으로 환산할 경우가 발생한다.

만족도 조사 분석

점수를 변환하거나 환산하여 사용하는 만족도 지표는 아래와 같다.

표 24 각 지표의 점수 산출 방식

지표명	점수 산출 방식
치안고객만족도(경찰청)	각 문항 100점 환산하여 산술평균
청년창직활성화 만족도(고용노동부)	100점 환산 점수 산출
디지털교과서 사용자 만족도(교육부)	100점 만점 환산 (학생만족도+교사만족도)/2
중·고교에서 직업체험 교육 참여자 만족도(교육부)	만족도 평균을 100점 기준으로 환산
대입정보포털 이용 만족도(교육부)	100점 환산
고등교육 재정지원 사업 재편에 대한 정책수요자 만족도-BK21플러스 사업(교육부)	평균점수 산출 후 100점 환산
고등교육 재정지원 사업 재편에 대한 정책수요자 만족도-특성화전문대학육성사업(교육부)	평균점수 산출 후 100점 환산
지역 평생교육 활성화 사업 주민참여 만족도(교육부)	만족도 점수 평균을 100점 만점 기준으로 환산
wee센터 학생 상담 만족도 및 상담 증가율(교육부)	100점 환산
찾아가는 어린이 안전체험교	전체문항 합산 방식

지표명	점수 산출 방식
실 참여자 평가(국민안전처)	100점 환산
U-City 통합플랫폼 기반구축 사업 주민만족도(국민안전처)	전체문항 합산 방식 100점 환산
세부담수준 전문가 만족도(기획재정부)	전체문항 합산 방식 최종 평균 점수 산출방식(100점 환산)
공공부문 부채 산출 전문가 인지도 조사(기획재정부)	전체문항 합산 방식 최종 평균 점수 산출방식: 100점 환산
세원투명성 관련 전문가 평 가(기획재정부)	전체문항 합산 방식 최종 평균 점수 산출방식(100점 환산 등)
복지서비스 만족도 조사(농 림축산식품부)	전체 문항 합산 최종 평균 점수 산출방식(100점 환산)
일반농산어촌개발 사업관계자 고객만족도 조사 (농림축산식품부)	100점 환산
장애인 문화예술교육 수혜자 만족도(문화체육관광부)	전체문항 합산 방식 최종 평균 점수 산출방식(100점 환산)
무지개다리사업 참가자 만족 도(문화체육관광부)	전체 문항 일괄 합산 $\Sigma(7\text{점 척도}-1) \times 16.7 \} / \text{전체 응답자수}$
문화동반자 사업 참가자 만 족도(문화체육관광부)	100점 기준으로 환산 참가자 설문조사(50%) 및 전문가 면담평가(50%)로 구성
장애인 창작 아트페어 참가 자 만족도(문화체육관광부)	최종 평균점수로 산출(100점 만점)
문화유산관리체계 개선 전문 가 만족도(문화체육관광부)	100점 환산
국립중앙박물관 고객만족도 (문화체육관광부)	$\Sigma\{(7\text{점 척도}-1) \times 16.7\} / \text{전체응답자수}$
국립한글박물관 고객만족도 (문화체육관광부)	최종 평균 점수 산출방식(100점 환산)
아리랑 대축제 참여자 만족 도(문화체육관광부)	측정 후 100점 기준 환산
한민족정보마당 이용자 만족 도(문화체육관광부)	최종 평균 점수 산출방식(100점 환산)
국외여행 통합 정보 홈페이 지 이용자 만족도(문화체육 관광부)	5개 문항(만족도 관련)의 5단계 만족도를 100점 만점으로 환산(매우만족 100점, 매우 불만족 20점)하여 합산후 평균 점수
한국방문우대카드 이용 만족 도(문화체육관광부)	38개 문항(만족도 관련)의 5단계 만족도를 100점 만점으로 환산(매우만족 100점, 매우 불만족 20점)하여 합산후 평균 점수

지표명	점수 산출 방식
정병검사 수검자 체감 만족도(병무청)	요소 만족도(가중치 적용, 60%)+전반적 만족도(40%) 최종 평균 점수 산출방식 : 100점 환산
사회복무요원 소양교육 만족도(병무청)	요소 만족도(가중치 적용, 60%)+전반적 만족도(40%) 최종 평균 점수 산출방식 : 100점 환산
산림교육만족도(산림청)	체감만족도와 요소만족도로 구분하여 50점을 배점하고 각 만족도 문항 수에 계량 모델을 적용하여 점수 산출 최종 평균 점수 산출방식(100점 환산)
숲태교만족도(산림청)	반적 만족도와 요소만족도로 구분하여 각 5점씩 총30점을 배점하고, 각 만족도를 100점으로 환산하여 합산 적용 최종 평균 점수 산출방식(100점 환산)
산림치유프로그램 만족도(산림청)	반적 만족도와 요소만족도로 구분하여 각 5점씩 총30점을 배점하고, 각 만족도를 100점으로 환산하여 합산 적용 최종 평균 점수 산출방식(100점 환산)
숲길 이용자 만족도 조사(산림청)	체감만족도와 요소만족도로 구분하여 50점을 배점하고 각 만족도 문항 수에 계량 모델을 적용하여 점수 산출 최종 평균 점수 산출방식(100점 환산)
통합식품 안전정보망 사용자 만족도(식품의약품안전처)	100점 환산 3개 시스템별 만족도 점수의 산술 평균
소통경험 소비자 만족도(식품의약품안전처)	전체문항 합산 방식 최종 평균 점수 산출방식(100점 환산)
식품의 영양표시 소비자 활용도(식품의약품안전처)	3개 문항 응답자 비율 {[‘영양표시를 인지 하고 있다’ 응답자 비율] × [‘영양표시 내용이 식품을 고르는데 영향을 미친다(보통 이상)’ 응답 자 비율]× [‘영양표시를 읽는다(가끔 읽는다 이상)’ 응답자 비율]} × 100
한부모가족 지원사업 만족도 (여성가족부)	각 문항 가중평균 산출(100점 환산)
청소년의 건강한 성장 지원 성과목표 달성을 위해 프로 그램 공모사업 서비스에 대 한 만족도 조사(여성가족부)	최종 평균 점수 산출방식(100점 환산)
청소년의 건강한 성장 지원 성과목표 달성을 위해 국립 청소년수련원 서비스에 대한 만족도 조사(여성가족부)	전체문항 합산 방식 최종 평균 점수 산출방식(100점 환산)

지표명	점수 산출 방식
청소년의 건강한 성장 지원 성과목표 달성을 위해 방과 후아카데미 서비스에 대한 만족도 조사(여성가족부)	전체문항 합산 방식 최종 평균 점수 산출방식(100점 환산)
청소년가족연계서비스 참여 만족도(여성가족부)	전체문항 합산 방식 각 문항 가중평균 산출(100점 환산)
아동·청소년 성폭력 예방교 육 만족(여성가족부)	각 문항 가중평균 산출(100점 환산)
센터심리지원 만족도(여성가 족부)	각 문항 가중평균 산출(100점 환산)
찾아가는 가정폭력 예방교육 수혜자 만족도(여성가족부)	각 문항 가중평균 산출(100점 환산)
원자력안전 정책만족도(원자력안전위원 회)	각 문항별 5점 평균과 100점 평균 제시 5점 척도의 100점 환산 점수 산출방식 : $(\text{점수}-1)/(\text{척도}-1) \times 100$
조달물자 품질만족도(조달청)	전체문항 합산 방식: 각 응답문항별 점수를 100점 만점으로 환산 (환산점수=응답척도*20) 최종 평균 점수 산출방식: 점수 합산 결과를 유효표본수로 산술평균하여 산출 $\text{평균점수} = \Sigma(\text{환산점수}) / \text{유효표본 수}$
연안정비사업 고객만족도(해 양수산부)	점수는 9, 7, 5, 3, 1의 5단계로 구분 최종 점수는 문항별 점수를 합산하는 방식으로 점수를 산출 하나, 문항별 최고점이 9점이므로, 90점 만점으로 측정
도시가치 및 비전에 대한 주 민 이해도 제고(행복중심복 합도시건설청)	주민인지도 관련 항목(2개)을 합산하여 100점(%)으로 환산
민간기업 기후변화 리스크평 가 활용 만족도 조사	매우불만족 1점, 불만족 2점, 보통 3점, 만족 4점, 매우만족 5점으로 100% 환산
환경분쟁조정 인지도 및 만 족도	최종 평균 점수 산출방식(백분율 환산)
통합환경관리제도 인지도 및 만족도	항목별 합산 및 최종 평균 점수 산출(100점 환산)

※만족도 지표 현황에서 나타난 점수 산출 방식을 최대한 그대로 사용함.

정책적 제언

다수의 지표에서 점수변환과 환산을 사용하고 있다. 그러한 변환과 환산이 정당한지, 필요한지를 판단하기 위하여 점수변환이나 환산이 사용된 지표에 관하여 그 목적과 근거를 밝히는 것이 필요하다.

4.8. 무응답의 문제

이론적 논의

대부분의 연구설계는 조사과정에서 완전한 자료가 얻어졌다는 가정 하에서 진행된다. 그러나 실제 설문조사를 시행하다보면 계획하고 기대했던 만큼의 자료를 확보하지 못하는 경우가 많이 발생한다. 설문의 성격이나 상황, 수행방법에 따라 다를 수 있겠지만, 일반적으로 설문의 경우 응답률이 40%정도만 되어도 운이 좋다고 이야기한다(Gujarati&Porter, 2009). 이렇게 볼 때, 전수조사는 물론이고 표본조사의 경우 과연 40% 응답률이 모집단을 대표하는지에 대한 고민을 하지 않을 수 없다. 만약 무응답자들이 응답자들과 체계적인 차이를 보인다면 편향(bias)이 발생하게 된다. 설문을 통하여 정책의 성과를 제대로 분석하기 위해서는 그 분석의 기반이 되는 데이터의 질이 좋아야한다. 데이터가 무응답의 오류를 포함하고 있을 경우 그 분석결과는 타당성과 신뢰성이 떨어지게 된다. 그럼에도 불구하고 각부처의 만족도 조사 평가에서 무응답에 대한 보고와 평가가 이루어지지 않고 있다.

무응답의 오류가 위험한 이유는 그 오류의 크기가 어느정도인지 파악할 수 없다는 사실과 무응답의 오류를 발생시키는 무응답률과 거부율이 지속적으로 증가하고 있다는 점이다.

우선 무응답 오류의 크기는 다음과 같이 살펴볼 수 있다.

무응답 오류를 도출하기 위하여 모집단 평균을 확률론적으로 정의하면 아래와 같다.

$$\bar{Y} = \frac{Y}{N} = \lambda_0 \bar{Y}_0 + \lambda_1 \bar{Y}_1$$

$\bar{Y}$: 모집단평균, $\bar{Y}_0$: 무응답집단평균, $\bar{Y}_1$: 응답집단평균,
 λ_0 : 무응답집단비율, λ_1 : 응답집단비율

이 경우 무응답 편향은 다음과 같이 분석할 수 있다.

$$\text{무응답 오류} = \lambda_0 (\bar{Y}_1 - \bar{Y}_0)$$

따라서 무응답오류 무응답 비율이 작아질수록, 응답집단과 무응답집단의 평균차이가 작을수록 오류도 작아지게 된다. 그러나 현실적으로 무응답집단의 평균을 알 수

없기 때문에 무응답오류의 크기가 어느정도인지 확인하기 어렵다.

정책적 제언

일반적으로 무응답이 있을 경우 무응답을 확률변수로 대체하여 자료수집을 하는 경우가 있다.

$$\hat{Y}_i = R_i Y_i + (1 - R_i) Z_i$$

$\hat{Y}_i$: 추정값, R_i : i 개체의 응답확률, Z_i : 대체값

이 경우 대체값을 무응답자들의 실제 값과 비슷하게 설정할수록 추정값은 무응답자들의 실제 값과 유사해진다.

그러므로 무응답 오차를 줄이는 방법으로 응답율을 높이는 것과 대체값이 평균적으로 무응답자들의 평균과 유사하도록 만드는 방안이 제시된다. 대체값은 정확하게 설정하는 것이 가장 좋은 방법이지만 무응답자들의 평균값을 확인하는 것이 불가능하므로 현실적인 방법은 응답율을 높이는 것이다.

응답률이 낮은 원인은 크게 3가지로 나누어 볼 수 있다.

첫째, 응답자를 접촉하는 것이 어려워지고 있다. 전화조사의 경우 수신차단장치의 발전으로 인하여 원하지 않는 전화를 차단할 수 있게 되었다. 따라서 이러한 고려없이 표본수의 확보에만 치중하면 편향된 조사결과가 나타날 위험이 크다. 직접 설문 경우에는 경비시스템의 강화와 범죄의 위험 때문에 외부인들이 거주자들을 만나기가 쉽지 않다. 또한 여성의 사회참여 확대에 인하여 평일 낮시간에는 부재중인 가구가 많아 집에서 응답자를 연결해줄 사람을 찾기가 어려워졌다.

둘째, 대상자들이 응답에 참여하는 것을 거부하고 있다. 누구나 한번쯤은 전화나 길거리에서 설문조사요구를 거절한 적이 있을 것이다. 대부분의 경우 바빠서 조사에 응할 시간이 없거나 유사한 설문조사의 경험, 개인정보를 침해받지 않으려는 경향 때문에 조사를 거부하고 있다.

셋째, 조사가 불가능한 경우이다. 응답자가 육체적, 정신적으로 조사에 응할 수 있는 준비가 안되어 있을 때 발생한다. 또한 언어소통의 문제로 인하여 조사가 불가능

한 경우도 있다.

이에 대해 무응답을 방지하기 위한 여러 가지 전략들이 개발되어 있다.

첫째, 여러 번 반복해서 접촉을 시도하는 것이다. 많은 경우 응답자를 만날 수 없어 설문이 이루어지지 않는데, 접촉시도 횟수를 늘리면 그만큼 응답률을 높일 수 있다. 우편의 경우 설문지를 여러 번 우송하거나 미리 통고하여 접촉 기회를 늘릴 수 있고, 전화 조사의 경우, 응답하기 편리한 시간을 선택하여 여러 차례 전화를 걸어 설문을 요구할 수도 있다. 직접 설문을 할 경우, 처음에는 경계하며 응답을 꺼리더라도 추가적인 방문에서는 설문에 응할 수도 있다.

둘째, 유인장치의 제공이다. 설문은 응답하는 사람에게 시간과 노력에 따른 기회비용을 발생시킨다. 따라서 일종의 보상적인 차원에서 인센티브를 제공할 수 있다. 보상의 경우 금전적인 보상이 비금전적 보상보다 더 효과적인 것으로 나타났으며, 보상액이 클수록 응답률을 높이는 것으로 알려져 있다. 유인장치를 제공할 때 주의할 점은 보상으로 인해 응답자가 낙관적인 응답을 할 우려가 있다는 것이다. 따라서 보상이 너무 지나치게 설정되어서는 안된다. 교육부에서 실시하는 ‘대학입학정보 포털 관련 온라인 만족도 조사’에서는 설문에 참여한 대상자들에게 1,300원의 모바일 쿠폰을 지급하고 있다.

셋째, 면접자의 역량 강화. 직접 설문의 경우 면접자의 사회적 기술이 중요하다. 그래서 무응답 상황에 노련한 경험자를 우선적으로 면접자로 선발해야한다. 이들은 응답을 얻기 위한 과정에서 발생하는 문제점들을 인식하고 있고 해결 방법의 탐색에 있어서 능숙하기 때문이다. 부족한 면접자는 훈련을 통해서 확보한다. 훈련에는 응답자와의 접촉을 유지하는 방법, 응답자에게 설문을 소개하는 방법, 응답을 유도하는 방법 등에 대한 내용이 포함되어야 한다. 그리고 면접자에 대한 감독을 강화하여 면접자들이 성실하게 설문을 수행하는지 지속적으로 확인할 필요가 있다.

05. 결론

5. 결론

정책 수혜자를 중심으로 정책을 평가한다는 것은 민주행정의 밑거름이 된다는 의미에서 각 조직에서 이루어지는 만족도 조사는 큰 의미를 지닌다. 그러나 만족도 지표가 정책이나 사업의 성과를 정확히 측정하기 위해서는 척도의 신뢰성과 타당성이 필요하다.

신뢰성을 확보하기 위한 다양한 이슈들을 살펴보았는데, 경험적 연구를 통해서 모두 고려해야 할 필요성이 있다는 사실을 확인할 수 있었다. 그런데 실제 현실에 있어서는 예산과 시간의 제약으로 인해서 신뢰성과 관련된 모든 조건들을 충족시키기는 쉽지 않다. 따라서 Churchill과 Peter(1984)의 메타 연구에서 나타난 결과를 참고하여, 개념을 측정하는 질문의 수와 질문에서 선택할 수 있는 응답범주의 수를 늘리려는 노력을 해야 할 것으로 보인다.

타당성은 만족도 지표가 만족도를 얼마나 잘 수치화 시켰는지와 관련된 문제이다. 타당성을 확보하기 위해 연구자는 만족도를 여러 질문으로 측정하는 것이 좋다. 왜냐하면 단일 항목의 질문으로 추상적 개념을 완벽하게 측정하는 것이란 사실상 쉽지 않기 때문이다. 여러 질문을 통하여 각 질문들이 복합적으로 만족도라는 개념의 외형을 잘 표현할 때, 타당성은 높아진다. 이를 위하여 질문들은 서로 내적 일관성을 가져야 한다. 질문들은 연구자가 임의로 구성하는 것이 아니라 기존의 연구들을 참고하여 정밀화 시킬 때 내적 일관성을 높일 수 있을 것이다.

측정의 문제가 해결되고 나면 과연 누구를 대상으로 측정해야 하는가의 문제에 직면한다. 즉 모집단 설정의 문제이다. 아무리 측정을 정확하게 해냈다고 하더라도 대상이 잘 못 설정되면 결과가 왜곡될 수 있다는 것을 윌-로저스 현상과 심슨의 역설을 통해서 살펴보았다. 그럼에도 불구하고 만족도 조사 연구들에서는 모집단, 즉 측정대상에 대한 고민이 부족한 것으로 보인다. 일부 만족도 조사에서는 모집단과 표본 프레임이 불일치하는 경우가 발생하기도 하였다. 만족도 조사의 대상은 결국 정책관련자들이 되어야 한다. 그런데 실제 연구에 있어서는 표본 프레임을 설정할 수 없는 경우가 많이 발생한다. 이때에는 정책에 관련되었는지를 묻는 설문문항을 설문지에 포함시켜 정책관련자들로 표본을 구성하는 방법을 사용할 수도 있을 것이다.

설문지는 실제 만족도 조사의 설계와 준비가 마무리 되어 나타난 결과물로서 조사 결과가 왜곡되지 않도록 하기 위해서 작성시에 유념해야 할 사항들이 많다. 조사 문항이 정책을 포괄적으로 평가하기 위해서는 우선 설문문항들의 타당성과 신뢰성이 확보되어야 한다. 그리고 설문문항들의 배열에도 신경써야 하는데, 설문 문항 배열에 문제가 있을 때에는, 일관성 효과·대비효과·피로효과·중복효과 등의 문제가 발생할 수

있기 때문에 이를 방지하기 위한 설문배치가 필요하다. 간혹 정책을 잘 알지 못하는 사람에게도 설문조사가 이루어지는 경우가 있는데, 잘모르거나 어려운 질문에 대해서 응답자가 중앙값이나 무응답으로 대답하는 경향을 보일 수 있다. 조사대상의 문제에서도 살펴보았지만, 만족도 조사는 기본적으로 정책의 내용을 알고 있는 관련자들을 대상으로 이루어져야한다. 하지만 표본프레임의 작성이 어려워 부득이하게 정책을 잘 알지 못하는 사람들이 표본으로 추출될 가능성이 있다면, 설문문항에서 정책인식여부나 관련여부를 묻는 문항을 추가하여 설문결과의 해석에서 사용할 수 있어야 한다.

마지막으로 만족도 조사는 자체적으로 하는 것이 바람직한가의 문제가 있다. 만족도 조사는 기본적으로 공개의 목적이 아니라 정책을 자체적으로 평가하고 개선하기 위한 내부자료의 성격을 가지므로 의도적인 왜곡의 유인이 없다고 보인다. 따라서 자체적으로 만족도 조사를 실시하는 것이 큰 문제가 될 것으로 판단되지는 않는다. 그러나 문제는 정부조직 담당자들이 만족도 조사에 대한 전문지식이 부족할 경우 의도하지 않은 왜곡이 초래될 수 있다는 점이다. 또 하나 중요하게 지적하고 싶은 것은 표본을 추출할 때, 확률표본추출을 할 경우에는 외부기관에 조사를 위탁하는 것이 큰 문제가 안되지만, 비확률표본추출의 경우에는 오히려 자체조사를 하는 것이 바람직할 수도 있다는 것이다. 비확률표본추출은 모집단의 특성에 대해서 잘 알고 있어야하고 정책의 성격도 파악하고 있어야 대표성있는 표본을 추출할 가능성이 높아지기 때문이다.

참고문헌

- 고길곤 (2014). 설문순서에 따른 응답오차의 유무와 원인에 대한 연구, 「한국행정학회 학술발표 논문집」, 2014(12): 168-189.
- 고길곤·탁현우·강세진 (2015). 설문조사에서 문항의 역코딩 여부가 응답결과에 미치는 영향-자아존중감과 자기효능감 측정 사례를 중심으로. 「한국행정학보」, 49(3): 515-539.
- 김득성·장서일·고준희·김진모 (2009). 청감실험을 통한 5점, 7점 11점 척도 비교 연구. 「한국소음진동공학회 학술대회」, 2009(4): 184-185.
- 김미숙·정경란 (2008). 영재교육 담당교사의 영재교육 만족도 분석, 「The Journal of Korean Education」, 35(2): 153-172.
- 김현구 (2003). 정부업무 기관평가의 이론적 논고, 「한국행정학보」, 37(4): 57-78.
- 게르트 보스바흐·옌스 위르겐 코르프 (2016). 「통계의 거짓말」, 옮긴이:강희진, 작은책방.
- 김인·우양호·조정현 (2007). 행정서비스 현장 고객만족도 영향요인-부산광역시 민원행정서비스를 중심으로, 「한국행정논집」, 19(4): 925-954.
- 남궁근 (2012). 「정책학」, 법문사.
- 더렐 허프 (1997). 「통계의 마술-재미있는 통계 이야기」, 옮긴이: 김정흠, 청아출판사.
- 로버트 후크 (1995). 「통계학자와 거짓말쟁이」, 옮긴이:김동훈, 새날.
- 박기관 (2010). 지방정부 상수도 서비스의 고객만족도 평가 및 영향요인에 관한 연구, 「국가정책연구」, 24(1): 157-178.
- 박원호·안도경·한규섭 (2013). 제18대 대통령 선거의 경제민주화 쟁점에 대한 실험설문: 프레임 효과의 중심으로, 「한국정치연구」, 22(1): 1-28.
- 박흥식·나현 (2010). 민원행정서비스 만족도에 대한 투명성 효과: 직접 및 조절인자, 공정성을 통한 매개인가?, 「행정논총」, 48(4): 385-408.
- 손성규 (2004). 외부감사인의 내부감사의존도에 관한 연구, 「회계저널」, 13(1): 1-42.
- 신두섭·박승규 (2012). 공공문화기반시설 이용만족도 결정요인 분석, 「문화경제연구」, 15(3): 139-159.
- 안명식·한인수·오홍석 (2016). 직무만족 조사에서의 질문 순서 및 질문 의도 효과에 관한 연구, 「한국콘텐츠학회논문지」, 16(7): 423-430.
- 이승중 (1990). 정보공개의 행정책임성에 관한 효과 및 공개방향, 「지방행정연구」, 5(1): 27-43.
- 이시경·이중화 (2006). 기초자치단체 행정서비스의 주민만족도 측정방법의 재검토: 측정지표 구성과 중요도의 측정을 중심으로, 「한국공공관리학보」, 20(1): 99-117.
- 이주현 (2011). 「설문조사시 질문순서가 응답결과에 미치는 영향」, 서울시립대학교.
- 이향수·안형기 (2011). 고객지향적 행정서비스 품질과 주민 만족도에 관한 연구, 「행정논총」, 49(4): 217-245.
- 정정길·최종원·이시원·정준금·정광호 (2016). 「정책학원론」, 대명출판사.
- 정주용 (2008). 「정책수용성 급변현상에 관한 연구-방사성폐기물처리장 입지정책을 중심으로-」, 고려대학교.
- 정진경 (2009). 공적사회복지서비스 행정기관 고객만족도에 관한 연구: 동 주민자치센터 이용자 중심, 「한국공공관리학보」, 23(3): 1-22.

- 조성한·강정석 (1998). 중앙행정기관에 대한 국민고객만족도 조사. 한국행정연구원 연구보고 98-15.
- 채서일 (2003), 사회과학조사방법론, 학현사.
- Benton, J.E. & Daly, J.L. (1993). Measuring Citizen Evaluations: The Question of Question Order Effects. *Public Administration Quarterly*, 16(4): 492-508.
- Churchill Jr, Gilbert A., and J. Paul Peter. "Research design effects on the reliability of rating scales: A meta-analysis." *Journal of marketing research* (1984): 360-375.
- Dawes, J. G. (2012). Do data characteristics change according to the number of scale points used? An experiment using 5 point, 7 point and 10 point scales. *International Journal of Market Research*. 50(1): 61-77.
- Stipak, B. (1979). Citizen Satisfaction with Urban Services: Potential Misuse as a Performance Indicator. *Public Administration Review*, 39(1): 46-52.
- Parks, R. (1984). Linking Objective and Subjective Measures of Performance. *Public Administration Review*, 44(2): 118-127.
- Borgers, N., Sikkels, D., & Hox, J. (2004). Response effects in surveys on children and adolescents: The effect of number of response options, negative wording, and neutral mid-point. *Quality & Quantity*, 38(1): 17-33.
- Brown, K., & Coulter, P. (1983). Subjective and Objective Measures of Police Service Delivery. *Public Administration Review*, 43(1): 50-58
- Campbell, D.T., & Fiske, D.W. (1959). Convergent and Discriminant Validity by the Multitrait-Multimethod Matrix, *Psychological Bulletin*, 56: 81-105.
- Chong, D., & Druckman, J. N. (2007). Framing theory. *Annu. Rev. Polit. Sci.*, 10: 103-126.
- Churchill Jr, G. A., & Peter, J. P. (1984). Research design effects on the reliability of rating scales: A meta-analysis. *Journal of marketing research*: 360-375.
- Dawes, J. (2008). Do data characteristics change according to the number of scale points used ? An experiment using 5 point, 7 point and 10 point scales, *International Journal of Market Research*, 50(1).
- Duncan, W. J.(1981). *Organizational Behavior*, Boston: Houghton Mifflin Company.
- Feild, L., Pruchno, R. A., Bewley, J., Lemay Jr, E. P., & Levinsky, N. G. (2006). Using probability vs. nonprobability sampling to identify hard-to-access participants for health-related research: costs and contrasts. *Journal of Aging and Health*, 18(4): 565-583.
- Gujarati, D.N.&Porter, D.C. (2009). *Basic Econometrics*, McGraw Hill.
- Ingram, H., Schneider, A. L., & DeLeon, P. (2007). Social construction and policy design. *Theories of the policy process*, 2: 93-126.
- Kaplan, A.(1964). *The Conduct of Inquiry*, Scranton, Pennsylvania: Chandler Publishing.
- Leung, S. O. (2011). A comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-point Likert scales. *Journal of Social Service Research*, 37(4):

412–421.

- Lin, N. (1976). *Foundations of Social Research*, NY: McGraw–Hill.
- Moore, D.W. (2002) Measuring New Types of Question–Order Effects: Additive and Subtractive. *Public Opinion Quarterly*, 66: 80–91.
- Narayan, S., & Krosnick, J. A. (1996). Education moderates some response effects in attitude measurement. *Public Opinion Quarterly*, 60(1): 58–88.
- Oliver, R. L. (1981). Measurement and evaluation of satisfaction processes in retail settings. *Journal of retailing*, 57(fall): 25–49.
- Osborne, D. (1993). Reinventing Government. *Public Productivity & Management Review*, 16(4): 349–356.
- Parasuraman, V. A. Z., Berry, L. L., & Zeithaml, V. A. (1988). A Multiple Item Scale for Measuring Consumer Perception of Service Quality. *Journal of Retailing*, 64(1): 12–40.
- Raaijmakers, Q. A., Van Hoof, A., Hart, H., Verbogt, T.F.M,A, Vollebergh, W.A.M. (2000). ADOLESCENTS' MIDPOINT RESPONSES ON LIKERT–TYPE SCALE ITEMS: NEUTRAL OR MISSING VALUES?. *International Journal of Public Opinion Research*, 12(2): 209–217.
- Riley, M. W. (1963). *Sociological research* (Vol. 1). R. K. Merton (Ed.). New York: Harcourt, Brace & World.
- Shuman, H. & Presser, S. (1981). *Questions and Answers in Attitude Surveys*, NY: Academic Press.
- Stipak, B. (1979). Citizen satisfaction with urban services: Potential misuse as a performance indicator. *Public Administration Review*, 46–52.
- Vedung, E.(1997). *Public Policy and Program Evaluation*, Transaction Publishers.
- Yeager, D. S., Krosnick, J. A., Chang, L., Javitz, H. S., Levendusky, M. S., Simpser, A., & Wang, R. (2011). Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non–probability samples. *Public opinion quarterly*, 75(4): 209–747.
- Zaller, J. (1992). *The nature and origins of mass opinion*. Cambridge university press.