

연구논문

만족도 측정 방법의 비교 가능성 연구:

행정서비스 만족도의 4, 5, 11점 리커트형 문항을 중심으로

An Empirical Study on Comparability of Satisfaction Measurement Tools: Focusing on 4, 5 and 11 Likert Type Items for Public Service Satisfaction Survey

한혁^{a)} · 금현섭^{b)}

Hyuk Han · Hyunsub Kum

본 연구의 목적은 서로 다른 개수의 선택지를 갖는 리커트형 척도 문항의 결과를 상호 비교할 수 있는지를 검토하는 데 있다. 현재 서베이를 이용한 설문조사에서 상이한 개수의 선택지를 갖는 문항으로 측정된 값을 비교하는 경우가 많다. 그러나 이는 동일한 조건하에서 각각 측정된 문항들의 측정 결과를 전환할 수 있는 적절한 환산방법이 있어야 한다는 전제하에서만 가능하다. 그렇지 않은 경우 상이한 개수의 선택지 문항으로 측정된 값들의 비교는 타당성을 확보할 수 없기 때문이다. 하지만 이러한 환산방법의 적절성에 대한 검토는 그동안 그리 많이 시도되지 않은 채 일종의 관행으로서 100점 환산방법이 활용되고 있다. 본 연구는 4, 5, 11점 척도 문항으로 측정된 자료를 이용하여 기존에 제시된 100점 환산방법하에서 동일한 결과를 나타내는지, 그렇지 않다면 어떠한 보정 방법을 제시할 수 있는지를 살펴보고자 하였으며, 이를 서베이 실험을 통해 확인하고자 하였다. 분석 결과, 기존의 다양한 100점 환산방법으로는 4, 5, 11점 척도 문항으로 측정된 결과가 상이하다는 점을 확인하였고, 각 척도 문항이 동일한 결과를 산출할 수 있는 환산방법을 검토한 결과 단순한 선형 환산식으로는 적절한 방법을 구성하기 힘들다는 점 역시 확인할 수 있었다.

a) 서울대학교 행정대학원 박사과정.

b) 교신저자(corresponding author): 서울대학교 행정대학원 교수 금현섭.

E-mail: hyunsk@snu.ac.kr.

결과적으로 상이한 선택지 개수의 리커트형 척도 문항으로 측정된 결과를 선형 환산식을 이용하여 비교하는 것은 적절치 않으며, 특별한 상황에서만 제한적으로 가능하다고 할 수 있다.

주제어 : 리커트형(Likert-type) 척도, 서베이, 행정서비스 만족도, 조사 설계

The aim of this study is to examine whether results of various Likert type scale items can be compared. Currently, survey results are often compared with questions measured by different scale items, but one of the two assumptions is needed for comparing those results. First, various Likert type scale items should have same results by previous conversion methods under same circumstances or second, there is a new way to convert different results from different scale items to same results. If these assumptions are not met, it is inappropriate for this comparison. So, we try to examine whether results of 4, 5 and 11 Likert type scale items are same under well-known 101(0~100) scale conversion methods. If not, we should consider which conversion methods can be suggested. Additionally, all this process was conducted based on a survey experiment. As a result, we confirm that 4, 5 and 11 Likert type scale items produce different results under well-known conversion methods and it can be difficult to suggest simple linear conversion methods. Consequently, we conclude that it is inappropriate to compare results from different Likert type scale items by linear conversion and it can be possible only under particular conditions.

Key words : Likert scale, survey, public service satisfaction, survey design

I. 서론

최근 들어 중앙정부 및 지방자치단체를 포함한 공공영역 전반에서 수요자 만족도 조사가 증가하고 있다. 실제 국민들이 체감하는 바를 파악함으로써 정책수요의 진단이나 수행된 정책을 평가하기 위해 이루어지고 있으며, 정부업무평가 국민만족도, 공공기관 고객만족도(PCSI), 각 시·도 및 시·군·구에서 이루어지는 행정서비스·민원 만족도 조사 등이 그 예가 될 수 있다. 여기에 각종 언론사와 연구기관의 조사까지 포함하면 관련된 재원 측면에서 상당한 공적 노력이 투입되고 있다고 할 수 있다.

일반적으로 조사를 통해 도출된 결과들은 비교를 통해 의미가 부각된다. 준거기준이 명확하다면 모르겠지만 만족도와 같이 절대적 기준이 분명하지 않는 경우, 그 의미는 횡단면적이거나 시계열적 비교를 통해서만 가능하기 때문이다. 이때 횡단면적이란 함은 동일 설문에 대한 집단별 비교와 동일 집단에 대한 설문별 비교를 의미하고, 시계열적이란 함은 동일 설문에 대한 월별 또는 연도별 비교를 의미한다. 즉, 설문에 대한 다양한 집단의 체감 수준과 그 변동을 통해 상대적으로 평가하고자 하는 것이다. 물론, 이러한 경향은 만족도 조사에만 국한되지 않는다. 추상적인 개념을 측정하는 모든 과정에서 척도¹⁾가 도입되고, 이로부터 측정된 결과들 역시 비교를 통해 의미가 부각되기 때문이다. 인간 심리에 대한 측정, 인지에 대한 측정, 태도와 관련한 측정 등 굳이 만족도 조사가 아니더라도 무수히 많은 측정에서 비교가 진행되고 있다.

본 연구는 각종 조사 과정에서 많이 활용되면서도 다소 간과되고 있는 상이한 개수의 선택지로 구성된 척도 문항 간 비교 가능성을 살펴보고자 하였다. 예를 들어, 가장 보편적으로 사용되는 리커트형 척도 문항의 경우 상황에 따라 4점,

1) 척도(scale)는 단지 하나의 문항을 지칭하는 것이 아니라 일정한 숫자의 부여방식과 합산 방식을 거쳐 특정한 개념의 측정값을 나타낼 수 있도록 만들어진 일련의 문항을 토대로 한다.

5점, 7점, 11점(0~10점), 101점(0~100점) 등 다양한 선택지 개수의 척도가 사용되고 있지만, 이를 이용한 설문 결과들을 어느 수준까지 비교하여 사용할 수 있는지에 대한 검토는 그리 많지 않다. 가령, 동일 설문지 내에서 4점 척도 문항으로 측정된 설문 결과와 5점 척도 문항으로 측정된 설문 결과를 비교할 수 있는지, 혹은 동일한 설문을 한 해에는 4점 척도 문항으로 측정하고 다음 해에는 11점 척도 문항으로 측정하였다면 이를 비교할 수 있는지에 대해 뚜렷한 답이 부재한 상황이다.

많은 경우 100점 환산 정보를 제시하고 있지만 이 역시 혼란스럽기는 마찬가지이다. 각 조사마다 각기 다른 100점 환산방법을 택하고 있으며, 일부 환산 정보가 제시되어 있지 않은 경우도 있다. 가령, 어떤 조사에서는 척도 문항의 최저점을 0으로, 최고점을 100으로 환산하는 방법을 택하기도 하고, 다른 조사에서는 4점 척도 문항의 1점을 25점, 5점 척도의 1점을 20점으로 환산하기도 한다. 또 다른 조사에서는 측정값의 최댓값과 최솟값을 활용하여 100점으로 환산하기도 하는 등 100점 환산이라고 하더라도 서로 다른 환산방법을 사용하는 상황이다.²⁾ 당연히 동일하지 않는 100점 환산방법을 사용하여 비교하는 것은 금물이지만, 그렇다고 동일한 환산방법을 적용하는 경우에도 이들 문항 간 비교가 적절한지에 대해서도 되묻지 않을 수 없다.

각기 다른 척도 문항의 비교 가능성 즉, 자료분석을 위해서는 다음의 조건에 대한 고려가 필요하다. 첫째, 각기 다른 선택지 개수를 이용한 척도로 측정된 결과물이더라도 동일한 환산방법을 적용하는 경우 동일한 결과(평균)를 유지해

2) 다양한 환산방법에 대한 예는 인터넷 검색을 통해 쉽게 찾을 수 있다. 이를 통해 검색한 몇 가지 예를 살펴보면 2006년도 국립춘천병원 고객만족도 조사, 2015년도 전문연구정보 활용사업 이용자 만족도 조사의 경우에는 100점 환산 과정에서 최저점을 20점 혹은 25점으로 환산하였으며, 2008년 별정우체국연합회 기타공공기관 고객만족도(PCSI), 2012년도 고객만족도 설문조사(통일공원, 임해자연휴양림, 국민체육센터), 2012년도 기상업무 국민 만족도 조사, 고위공직자 청렴도 평가 계획, 민원수렴 및 환류시스템 구축을 위한 여론조사(輿論調査)실시 결과(관세청)에서는 최저점을 0점으로 환산하였다. 또한, 2014년도 서울대학교병원 자체감사 인식도조사에서는 최저점을 40점으로, 최고점을 100점으로 환산하여 사용하였다. 각 자료에 대한 출처는 참고문헌 URL 주소를 참조하면 된다.

야 한다. 이 경우 응답자들이 인식하는 정도를 동일하게 측정했다고 볼 수 있을 것이다. 둘째, 각 척도가 측정하는 값이 상이하다고 해도 응답강도 혹은 응답빈도를 조정할 수 있다면 비교가 가능할 것이다. 예를 들어 4점 척도를 고려한다면, 집단의 대표 값인 평균은 선택지인 1, 2, 3, 4라는 응답강도에 대해 각각 몇 명이 응답하였는지(응답빈도)를 종합한 값이다.³⁾ 5점 척도의 경우에도 마찬가지이다. 따라서 4점 또는 5점 척도에서의 응답강도 혹은 응답빈도를 조정하여 참값에 근사해질 수 있다면 이들 문항 간 비교가 가능해진다. 다만, 응답빈도는 이미 주어진 응답강도에 반응하여 결정된 값이기 때문에 실질적으로는 응답빈도가 아닌 응답강도만 조정이 가능할 것이다. 가령, 응답자들이 1.3의 응답강도를 느꼈지만 선택지가 1, 2, 3, 4로 주어졌다면 가장 근사한 값인 1을 응답할 수밖에 없을 것이고 이 경우 0.3만큼의 오차가 발생하게 된다. 따라서 0.3만큼의 응답강도 차이를 조정할 수 있다면 참값에 근사해질 수 있다고 보는 것이다.

본 연구는 위의 두 가지 전제조건을 검증함으로써 리커트형 척도 문항의 비교 가능성을 살펴보고자 하였다. 첫 번째 조건에 대해서는 기존에 제시되고 있는 다양한 100점 환산방법들에 따라 선택지 개수가 상이한 리커트형 척도 문항들이 동일한 결과(평균)를 산출하는지 살펴보았다. 만약 동일하다면 어떠한 이유로 동일한지, 동일하지 않다면 어떠한 이유로 동일하지 않는지를 확인하고자 하였다. 두 번째 조건에 대해서는 상이한 개수의 선택지에 따라 참값에 근사해지는 응답강도를 추정하였다. 문제는 실제 실험을 통하지 않고서는 참값 파악이 어렵다는 점인데, 본 연구에서는 하나의 척도 문항을 기준으로 다른 척도 문항들의 상대적 응답강도를 확인하는 방법을 활용하였다. 실제 참값은 알 수 없지만 비교 가능한 조건들을 탐색하고자 하였기 때문이다.

추가적으로 가상의 참값⁴⁾에 대해 상이한 선택지 개수의 리커트형 척도 문항들의 응답강도가 어떻게 분포되는지 확인해 보았다. 이를 통해 실제로는 서열척도이지만 대체로 등간척도로 간주되어 사용되고 있는 리커트형 척도가 실제 등간성을 갖는지, 또는 척도 활용 과정에서 기존연구들에서 제시되지 않았던 또 다른

3) 평균은 $\frac{1}{n} \sum \text{응답강도} \times \text{응답빈도}$ 이기 때문이다.

4) 가상의 참값에 대한 내용은 이후 선행연구와 연구방법에서 다루어질 예정이다.

특징들이 있는지 파악하고자 하였다.

본 연구가 활용하는 자료는 응답자들에 대해 동일 설문을 다양한 개수의 선택지로 구성된 즉, 4점, 5점, 11점 척도를 이용한 전화 서베이를 진행하여 채집하였다. 이를 위해서는 응답자의 동질성이 무엇보다 중요하지만 실제 서베이에서 이를 확보할 수 없기 때문에 동일 지역의 인구사회학적 특성이 유사한 응답자들을 층화추출을 통해 선발하여 진행하였다. 즉, 서베이를 이용한 유사실험(pseudo-experiment) 형태를 유지하려고 한 것이다.

II. 선행연구 검토

1. 행정서비스 만족도 조사의 대두 배경

서비스 만족도 조사가 널리 활용되고 있는 배경에는 사회 변화와 깊은 연관이 있다. 이상민 외(2000)는 이러한 사회 변화를 두 가지 요인으로 설명하고 있는데 첫째, 1900년대 후반의 시장의 경쟁 심화로 인해 기업들이 더 나은 제품과 서비스를 위해 노력하면서 실질적인 품질에서의 차별화가 힘든 상황까지 도달하게 되었고, 따라서 그 다음 단계로 고객 만족이 강조되었다는 것이다. 둘째, 정보 기술이 발달함에 따라 기업들은 체계화된 정보 구축뿐만 아니라 이를 통한 고객 관리 역시 가능하게 되었고, 또한 인터넷 환경의 구축을 통해 고객과의 쌍방향 의사소통이 쉽게 이루어지게 된 환경 변화에 있다. 이러한 변화는 단순히 고객의 만족 수준을 파악하는 것에 그치지 않고 고객을 찾고, 고객과의 소통을 통해 무엇이 부족한지를 파악하여 대처할 수 있는 환경이 마련되었다고 할 수 있다.

공공분야 역시 이러한 흐름에 예외가 아니다. 1990년 이전에는 중앙 정부의 일방적인 서비스 공급만 존재하여 경쟁의 여지가 없었지만, 이후 지방의회가 구성되고 주민들에 의해 지방자치단체장이 선출되면서 본격적인 행정서비스 경쟁 시대가 대두되었다고 할 수 있다(최병대 2003). 특히 지방자치단체장과 지방의회 의원은 4년마다 주민들에 의해 선출되기 때문에 재선을 위해서라면 유권자인 지역 주민의 요구에 민감할 수밖에 없고, 동시에 김대중 정부 이후 책임운영기관,

민영화, 민간위탁 등 신공공관리 요소가 강한 행정개혁이 진행되면서 공공영역 전반에 경쟁을 토대로 한 효율성의 가치가 더욱 강조되었기 때문이다(권기현 2007). 따라서 지방자치단체 수준에서도 주민의 만족 수준은 지방행정 전반에 있어서 중요한 가치가 되었고, 정보 기술의 발달로 주민들이 타 지방정부의 활동을 손쉽게 확인하고 비교하며 보다 용이한 행정서비스 요구와 불만의 빠른 전파·증폭이 가능하게 되었다. 이로 인해 지방자치단체의 적극적 대응이 더욱 유도되었다고 할 수 있다.

<표 1> 민간/공공분야에서의 고객/시민 관점의 변화

분야	분류	1970년대	1980년대	1990년대	1990년대 후반	2000년대
민간 분야	특징	One-way Selling	Customer Satisfaction	Database Marketing	Customer Relationship Management	
	대(對) 고객관점	수동적 구매자	선택적 구매자	개성화, 다양화된 구매자	쌍방향적 파트너	
	고객과의 관계	전체시장에서 일방적 공급	고객만족도 측정, 일방적 관계	그룹화된 고객과의 일방적 관계	개별 고객과 쌍방향 의사소통	
공공 분야	특징	Supply-Oriented Approach		Customer Satisfaction		Customer Relationship Management
	시민에 대한 관점	고객개념의 부재 (수동적 대상자)		고객개념의 등장 (능동적 대상자)		적극적 고객관리 (능동적 파트너)
	시민과의 관계	전체 시민(주민)에 대한 무차별적인 일방적 공급		고객(시민)만족도 측정, 일방적 관계		그룹화되고 계층화된 고객과의 쌍방향 관계

출처: 앤더슨컨설팅(2000, p.15~18), 이상민 외(2000), 최병대(2003)를 재구성

이처럼 공공영역에서도 최종 수요자에 대한 행정서비스 만족 수준이 무엇보다

다 중요한 요소로 떠오르면서, 서비스 공급자로서 공공조직에 대한 평가와 이를 통한 발전을 목적으로 만족도 조사가 빈번해지고 있다. 지방정부, 공기업은 말할 것도 없이 공공조직의 성과를 평가하고자 하는 중앙정부, 연구기관, 언론 역시 만족도 조사를 통해 공공분야를 평가하고 있는 상황이며, 향후에도 이에 대한 관심은 지속될 것이다.

결국 만족도 조사에 대한 관심 증대는 국민의 만족 수준이라는 추상적인 개념을 측정할 수 있도록 구체화하려는 시도가 발현된 것으로 볼 수 있으며, 이러한 현상은 만족도뿐만 아니라 인간의 추상적인 심리를 측정하고자 하는 모든 영역에 확대·적용되고 있다. 본 연구 역시 주된 연구대상은 행정 서비스 만족도 조사이지만, 궁극적으로는 추상적인 개념을 척도를 통해 구체화하고자 하는 모든 영역에 대한 논의라고 할 수 있다.

2. 상이한 선택지 개수의 척도 문항 환산 및 비교

실제 많은 조사연구들을 보면 상이한 개수의 선택지로 구성된 문항으로 측정된 결과를 100점으로 환산하여 보고하고 있다. 다양한 이유가 있겠지만, 가장 큰 이유는 100점 환산점수의 인식 용이성 때문일 것이다. 7점 척도 문항의 4점보다는 100점 척도 문항의 50점이 보다 직관적으로 인식되는 것과 같은 이치이다.

문제는 이러한 인식의 용이성 때문에 환산된 값들이 아무런 주의나 경고 없이 나열되어 비교되고 있다는 점이다. 가령, 4점 척도 문항으로 조사되어 환산된 75점과 7점 척도 문항으로 조사되어 환산된 70점을 동시에 접한다면 은연중에 75점이 더 크다고 생각하게 된다는 것이다. 물론, 서로 다른 선택지 개수의 척도로 측정된 결과를 비교할 수 있다는 근거가 뚜렷하다면 문제가 되지 않겠지만 현실은 제 각각의 100점 환산방법을 통해 결과를 제시하고 있으며, 이러한 결과들이 비교 가능한지에 대한 근거 역시 빈약하다는 데 있다.

이러한 비교가 빈번히 이루어지고 있는 이유는 기본적으로 척도를 등간척도로 가정하고 있기 때문이다. 즉, 척도 문항 선택지의 응답강도가 배수로 나열되어 있기 때문에 여기에 일정 숫자를 곱해 주거나 나누어 준다고 할지라도 그 간격이 일정할 것으로 보는 것이다. 하지만 리커트형 척도는 기본적으로 서열척도로서

그 응답강도가 비례적인 것을 가정하지 않는다. 다만 많은 실험을 통해 응답자가 등간척도에 근사한 것으로 인지하는 경우가 많음이 제시되었고, 이를 자료분석에서의 용이성 차원에서 원용하는 것이 실제 상황이다. 따라서 많은 연구들이 리커트형 척도에서의 응답강도가 등간이 아님을 지적하고 있으며 기계적인 환산 비교의 위험성을 우려하고 있으나 국내에서는 이에 대한 연구들이 그리 많지 않은 상황이다.⁵⁾

척도의 선택지 수와 관련된 해외 연구들은 주로 심리학과 마케팅 분야에서 이루어졌는데, 대부분 설문 성격에 따른 가장 적절한 선택지 수를 탐색하는 연구들이다.⁶⁾ 즉, 선택지 수에 대한 등간성 여부가 아니라 선택지 수에 따라 상이한 행태를 보이기 때문에 이로부터 도출된 결과들 역시 달라질 것에 주목하는 것이다. 척도의 선택지 수에 따라 결과의 타당성, 신뢰성 등이 달라진다는 가정 하에 많은 서베이 실험이 진행되었고 이 중 척도 문항의 비교 가능성과 관련된 연구는 주로 타당성과 관련하여 진행되었다. 이용된 척도 문항들이 타당하다면, 다시 말해 참값에 근사하다면 척도 문항 간 비교도 자연스럽게 가능해지기 때문이다.

그러나 실증 연구의 결과들은 뚜렷한 결론을 제시하지 못하고 있다. 대체로

5) 조사 과정에서 응답자의 행태에 영향을 미치는 요인은 다양하며, 이에 대한 위험성은 이미 많은 국내 학자들에 의해 제기되어 왔다. 설문 과정에서 질문 순서에 따라 응답자의 응답에 변화가 발생하는지(고길곤 2014; 이주현 2011), 측정 방법(전화/대면/온라인)에 따라 사회적 바람직성에 의한 편향(social desirability bias)이 어떻게 달라지는지(송인덕·조성겸 2013), 문항의 역코딩 여부가 어떻게 응답에 영향을 미치는지(고길곤 외 2014), 설문지의 길이가 응답의 질에 어떠한 영향을 미치는지(김권현 외 2015) 등 많지는 않지만 다양한 연구들이 진행되었다. 그러나 리커트형 척도 문항의 선택지 수와 관련된 연구는 극히 적는데 유흥준·김월화(2014)의 연구가 5점 척도 문항, 11점 척도(0~10점) 문항, 101점 척도(0~100점) 문항, 서열 부여 척도 문항을 비교하고 있지만, 각 리커트형 척도 문항에 대한 평균 및 표준편차 정보만 제공하고 있을 뿐 해당 척도 문항 간 측정값의 비교 가능성은 검토하고 있지 않다.

6) 이들 연구들은 대체로 선택지 수가 상이한 척도 문항의 특징들을 확인하고, 각 척도 문항의 타당성, 신뢰성을 통해 최적의 척도 문항을 찾아내려는 취지에서 진행되었으며(Green & Rao 1970; Preston & Colman 2000), 이를 통해 실제 활용 측면에서 응답자의 피로도, 선택지에 따른 응답 차별화가 확실한지에 대한 식별력(discriminating power) 등을 좋은 측정도구의 판단 기준으로 제시하고 있다(Preston & Colman 2000).

선택지가 많을수록 타당성이 증가한다는 연구(Lehmann & Hulbert 1972; Loken et al. 1987), 선택지 수가 타당성에 영향을 미치지 않는다는 연구(Matell & Jacoby 1971), 그리고 7개의 선택지까지는 타당성이 증가하지만 그 이후에는 별다른 효과가 발생하지 않는다는 조건부 결론 등이 병존하고 있다(Preston & Colman 2000).⁷⁾ 선택지 수와 타당성 간의 관계가 존재하지 않는다는 연구를 제외하면 선택지 수에 따라 다른 결과를 나타낸다는 점에는 동의하는 것으로, 따라서 척도 문항 간 상호 비교는 곤란하다는 결론에 다다르게 된다.

이런 결론에 대해서 몇 가지 원인을 추측해 볼 수 있는데 먼저, 응답분포의 상이성을 들 수 있다. 척도 응답 과정에서 응답자는 본인의 의사와 가장 “유사한” 선택지를 선택하게 된다. 하지만 본인의 생각과 가장 근접한 선택지를 선택한다는 의미는 이 과정에서 일정한 오차가 발생하게 된다는 것인데 가령, 4점 척도에서 1.3의 응답강도를 느끼고 있지만 주어진 선택지 1과 2 중에서는 1을 선택하는 것과 같다. 이 경우 0.3의 오차가 발생하게 되는데⁸⁾ 해당 오차는 1을 응답한 응답빈도와 결합하여 오차의 전체적 크기를 결정하게 된다. 즉, 1.3의 응답강도를 느낀 10 사람이 1을 응답하였을 때와 동일한 응답강도를 느낀 5 사람이 1을 응답하였을 때 오차의 크기는 달라진다. 결국, 응답빈도들을 종합한 응답분포는 오차에 적지 않은 영향을 미친다고 볼 수 있는데, 조사영역이나 설문문항마다 응답분포는 다르다는 점에서 오차의 전체적 크기 역시 달라질 수밖에 없다.⁹⁾

7) 참고로 본 연구에서 다루고 있는 타당성 영역이 아닌 신뢰성과 응답 시간에 대한 기존의 연구들을 살펴보면, 신뢰성의 경우에는 척도의 선택지 수가 일정 수준 이상이어야 내적 신뢰도가 확보될 수 있다는 연구(Green & Rao 1970), 내적 신뢰도의 경우에는 척도 문항에 따라 별 차이가 없지만 조사-재조사(test-retest) 신뢰도의 경우에는 오히려 선택지 수가 많을수록 감소한다는 연구(Preston & Colman 2000), 2, 3점 척도 문항에서만 두드러지게 감소한다는 연구(Bending 1954) 등이 혼재하고 있는 상황이다. 응답 시간의 경우에는 대다수 연구들이 척도의 선택지 수가 많을수록 응답 시간이 증가한다는 동일한 결론을 제시하고 있다(Maxell & Jacoby 1972; Preston & Colman 2000)

8) 척도의 선택지 개수가 증가할수록 타당성이 증가한다는 주장은 선택지가 증가할수록 오차를 감소시켜 주기 때문이다.

9) 척도에 대한 많은 연구들은 설문의 목적, 구성에 따라 상이한 결과가 도출될 수 있음을 가정하고 있다(Garland 1991; Jamieson 2004; Matell & Jacoby 1971; Maxell & Jacoby 1972; Preston & Colman 2000).

또 다른 원인으로 예상할 수 있는 것은 실제 참값을 알기 어렵다는 것이다. 타당성 연구를 진행하기 위해서는 먼저 참값을 확인하고 이로부터 얼마나 오차가 발생하는지 측정할 필요가 있다. 하지만 선행 실험이 없는 경우 참값을 알아내는 것은 불가능하다. 따라서 기존의 연구들을 통해 인정된 측정도구의 값과 비교하는 공인타당성(concurrent validity) 혹은 전후 측정을 통해 두 값을 비교하는 예측타당성(predictive validity)을 확인하고 있다. 하지만 이들은 각각 서로 다른 참값을 가정하고 있기 때문에 결과적으로는 비교 대상의 차이로 인한 다른 결론의 가능성이 존재한다.

이밖에 리커트형 척도 문항의 비교 가능성에 대해 실제 서베이 실험을 활용한 경우도 있다. Levine & Baden(1978)이 진행한 소비자 선호 조사에서 각 척도 문항의 비교 가능성 조건을 탐색하는 연구가 그 예이다. 분석결과는 3점과 5점 척도 문항의 경우 ‘측정값×100/선택지 수’ 방법으로 환산하였을 때 측정값들이 통계적 차이가 없었으며, 5점 척도 문항과 7점 척도 문항의 경우에는 척도의 중앙값을 활용한 방법¹⁰⁾으로 환산하였을 때 차이가 없다고 보고되고 있다. 하지만 차이가 없었다는 의미는 다수의 실험 중 차이가 나는 경우가 상대적으로 소수에 그쳤다는 것이지 항상 동일했다는 것은 아니며, 차이가 나거나 나지 않는 이유에 대해서는 명확히 밝히고 있지 않다.

그렇다면 현재 주로 사용되고 있는 100점 환산방식에 대해 살펴볼 필요가 있는데, 대체로 다음과 같다.¹¹⁾ 첫째, ‘측정값×100/선택지 수’로서 예를 들어, 4점 척도의 1점을 25점, 2점을 50점, 3점을 75점, 4점을 100점으로 환산하는 방법이다. 이 방법은 리커트형 척도 문항의 간격 비율을 환산 후에도 그대로 유지함으로써 응답강도의 비율에 초점을 맞춘 것이라 할 수 있다. 가령, 4점 척도에서의 1점은 5점 척도의 경우 1.25와 같은 응답강도를 지닌다는 것이다. 두 번째 방법은 ‘(측정

10) 환산방법과 관련해서는 이후 과정에서 자세히 언급할 예정이다.

11) Levine & Baden(1978)은 기존 연구들에서 환산에 대한 표준적이 방법이 제시되지 않고 있기 때문에 통상적으로 사용되고 있는 방법들을 선정하여 비교하였다고 언급하고 있다. 예를 들어, 본고에서 제시한 첫 번째 방법, 중앙값을 통한 환산방법, Z-score 환산 등이 그것들이다. 본 연구에서는 그 이외에도 기존 연구들에서 제시된 환산방법들을 종합하여 검토하였다(Preston & Colman 2000).

값-1)×100/(선택지 수-1)'로 4점 척도 문항의 1점을 0점, 2점을 33.3점, 3점을 66.7점, 4점을 100점으로 환산하는 방법이다. 첫 번째 제시한 방법과는 다르게 척도들의 최솟값과 최댓값의 응답강도가 모두 동일하다는 전제하에 환산하는 방식으로 4점 척도 문항의 1점, 5점 척도 문항의 1점, 7점 척도 문항의 1점이 모두 동일한 응답강도를 갖게 된다.

<표 2> 각 환산방법에 대한 예시(100점 기준)

환산방식	측정값×100/ 선택지수	(측정값-1)×100/ (선택지수-1)	척도의 중앙값 활용
4 점	25, 50, 75, 100	0, 25, 75, 100	20, 40, 60, 80
5 점	20, 40, 60, 80, 100	0, 25, 50, 75, 100	17, 33, 50, 67, 83
11 점(0~10)	0, 10, 20... 90, 100	-10, 0, 10... 90	0, 10, 20... 90, 100

주: 11점 척도 문항의 경우, (측정값-1)×100/(선택지수-1)로 환산하였을 때 음수가 도출되기 때문에 이후 연구에서는 모두 0, 10, 20... 90, 100으로 환산하여 비교하였음.

이밖에도 Levine & Baden(1978)이 사용했던 척도의 중앙값을 활용한 환산방식과 100점 환산점수의 최솟값과 최댓값을 임의로 설정하여 환산하는 방식 등이 있다. 전자는 '측정값/환산할 척도의 중앙값×환산될 척도의 중앙값'으로 환산하는데, 예를 들어 4점 척도 문항을 5점 척도 문항으로 환산한다면, 4점 척도 문항의 1점은 $1 \times '1/2.5 (=4점\ 척도의\ 중앙값)' \times '3 (=5점\ 척도의\ 중앙값)'$ 으로 계산되어 5점 척도 문항에서는 1.2의 응답강도로 환산된다. 한편, 후자는 100점 환산점수의 최솟값과 최댓값을 임의로 설정하여 환산하는 방식인데 실질적으로는 위에서 제시한 두 번째 방법과 동일하다. 다만 최솟값을 40점, 최댓값을 100 등으로 임의로 고정시킨 상태에서 40~100 구간만 나누어 사용하는 것이다. 가령 4점 척도의 1점은 40점, 2점은 60점, 3점은 80점, 4점은 100점에 위치하게끔 한다.

쉽게 추론할 수 있는 것은 이들 환산방법들이 모두 서로 다른 가정하에 변환된 것이어서 각기 상이한 방식으로 환산한 결과들을 비교하는 것은 결코 긍정적이지 않다는 점이다. 하지만 동일한 환산방법을 이용하였다고 하더라도 사용된 척도가 등간척도가 아닐 가능성이 높기에 그 결과들을 비교할 수 있는지에 대한

의문은 여전히 남는다.

본 연구는 앞서 언급한 기존 연구들의 한계를 반영하여 일반적으로 정규분포를 띠고 있는 조사 결과에 대한 연구를 진행하고자 하며, 이를 다른 분포를 띠고 있는 조사로 확대 해석하지 않고자 한다. 동시에 각 척도 문항들의 환산값들이 동일하거나 동일하지 않는 이유에 대해 오차의 발생 원인을 토대로 설명하고자 한다. 그러나 본 연구 역시 전화 서베이 실험을 활용하고 있어 실제 참값을 알지 못한다는 한계가 있고 그 결과들이 얼마나 참값에 근사한지 확인할 수는 없다. 따라서 본 연구에서는 서로 다른 척도 문항들로 측정된 변수들을 통합할 때 빈번히 사용되는 방법¹²⁾을 이용하여 가상의 참값을 구성하고, 이에 각 척도에서의 응답들이 어떻게 대응하는지 살펴보았다.

Ⅲ. 연구 설계

1. 연구대상

실험이 비실험 방법과 다른 가장 큰 이유는 집단 간 동질성 확보에 있다. 실험은 자극이 주어지기 전에 실험집단과 비교집단 간 동질성을 최대한 확보하지만, 준실험 혹은 통계적 방법을 통해 집단 간 차이를 극복하려는 비실험은 그렇지 않다. 집단 간 동질성을 확보하고자 하는 배경에는 자극을 통해 발생한 효과가 집단의 특성에 기인한 것이 아니라 자극에 기인한 효과이길 기대하기 때문이다. 이러한 이유로 본 연구와 같이 각기 다른 척도에 대한 반응을 살펴보는 연구 역시 동일한 집단에서 자극이 주어지지 않는다면, 그 결과에서의 차이가 척도 차이에 의한 것인지 응답집단 간 차이에 의한 것인지 구별할 수 없게 된다. 따라

12) 본 연구에서는 일반적인 연구 과정에서 서로 다른 척도로 측정된 변수들을 통합할 때 빈번히 사용되는 방법을 사용하였다. 가령, 4점 척도로 측정된 A변수와 5점 척도로 측정된 B변수의 결과를 통합하고자 할 때, 일반적으로 $(A\text{변수의 평균}/4+B\text{변수의 평균}/5)/2$ 를 사용한다. 따라서 본 연구에서도 4, 5, 11점 척도의 평균을 동일하게 평균내어 가상의 참값을 구성하였으며 구체적인 식은 이후에 제시될 식(4)와 같다. 다만, 식(4)는 위와 동일한 방법이지만 11점을 중심으로 통합한 식이다.

서 집단 동질성 유지 또는 최소한의 유사성 확보는 매우 중요하다.

본 연구는 경상남도 창원시에 거주하는 240명의 지역 주민들을 대상으로 전문 조사 기관의 전화 조사를 통해 일련의 행정서비스에 대한 설문에 응답을 받았으며, 240명 중 80명에게는 4점 척도로, 또 다른 80명에게는 5점 척도, 나머지 80명에게는 11점(0~10점) 척도를 이용하였다. 즉, 동일 설문에 대한 3가지 다른 척도를 이용한 응답을 채집하였다. 집단 유사성 확보를 위해 층화추출 방식을 활용하였으며, <표 3>에서 볼 수 있듯이 응답자의 인구사회학적 특성 즉, 성비, 평균 나이 및 미혼 비율, 정치 성향, 소득, 학력 및 직업이 매우 유사하게 구성되었음을 알 수 있다.

<표 3> 응답자의 인구사회학적 특성

인구사회학적 특성	4점 척도	5점 척도	11점 척도
표본 수	80	80	80
남성비율	50%	50%	50%
평균 나이(만)	46.9	46.7	46.5
미혼 비율	26.25%	23.75%	28.75%
평균 정치성향 (5 점 척도: 1. 매우 진보,.. 5. 매우 보수)	3.2	3.2	3.0
평균 소득 (1. 월 100 만원 미만, 2. 월 100~199 만원 ... 7. 월 600 만원 이상)	4.3	4.1	4.4
직업			
농업/수산업/축산업/자영업	18.75%	15%	15%
판매/서비스직	11.25%	10%	15%
기능/숙련공/일반작업직/사무/기술직	27.50%	28.75%	30%
경영/관리직/전문/자유직	1.25%	1.25%	1.25%
공무원/공공기관 종사자	2.50%	2.50%	1.25%
전업주부/학생/무직/기타	38.75%	42.50%	37.50%
학력			
고졸 이상	47.50%	51.25%	46.25%
대졸 이상	52.50%	48.75%	53.75%

하지만 응답자의 인구사회학적 특성이 유사하게 구성되었다고 하더라도 집단 간 동질성 유지에 필요한 표본 크기가 부족하다는 비판이 제시될 수 있다. 그러나 본 연구와 유사한 유홍준·김월화(2014)의 연구는 137명을 대상으로, Preston & Colman(2000)은 149명을 대상으로 실험을 진행하여 결론을 도출하고 있다. 특히 본 연구의 대상이 되는 세 집단의 2개의 동일 척도(5점 척도)의 동일 문항에 대한 응답 결과를 비교하였을 때, 집단 간 평균값이 매우 유사함을 확인할 수 있었다 (문항1: 3.65, 3.49, 3.58 / 문항2: 3.41, 3.28, 3.31). 이를 바탕으로 본 연구는 집단 간 동질성은 어느 정도 확보되었다고 판단하였다.

<표 4> 설문조사 설계

지역	경상남도 창원시
조사 시기	2015 년 10 월
대상	총 240 명 (80 명×3)
사용 척도	4 점 척도 80 명, 5 점 척도 80 명, 11 점 척도 80 명
설문 항목	지방정부 행정서비스 체감 만족도 (14 개 영역×1 문항) 전반적인 서비스 만족도 (1 문항) 세금 대비 서비스 만족도 (1 문항)
지방정부 서비스조사 영역	재난방지, 보건의료, 보육, 농촌, 도로관리, 대중교통, 환경미화, 경제, 관광, 문화, 주거, 교육, 노인복지, 치안, 전반적 만족도, 세금 대비 만족도

조사 분야는 앞서 언급하였듯이 행정서비스 만족도 분야로 한정하였다. 이에 해당하는 설문 구성은 공공분야에서 빈번히 조사되고 있는 지방정부 행정서비스 체감 만족도를 총 14개의 행정서비스 영역(재난방지, 보건의료, 보육, 농촌, 도로 관리, 대중교통, 환경미화, 경제, 관광, 문화, 주거, 교육, 노인복지, 치안 영역)에 대해 묻고, 여기에 전반적인 서비스 만족 수준과 세금 대비 서비스 만족 수준을 질문하여 총 16개 문항으로 진행되었다. 14개의 행정서비스 영역 만족도에 대한

설문은 “귀하가 거주하는 창원시에서 제공하는 각각의 공공서비스에 대해 어느 정도 만족하시는지 말씀해주시요.”로, 전반적인 서비스 만족 수준은 “지금까지 응답해주신 영역들을 모두 포함하여 창원시의 공공서비스에 대해 종합적으로 얼마나 만족하십니까?”로, 세금 대비 서비스 만족 수준은 “귀하께서 지출한 세금에 비해 얼마나 만족스러운 공공서비스를 제공받고 있다고 생각하십니까?”로 이루어졌다.

2. 리커트형 척도 문항의 비교 가능성

앞서 언급하였듯이, 리커트형 척도 문항 간 비교가 가능하기 위해서는 첫째, 상이한 척도 문항으로 측정된 결과들이더라도 동일한 환산방법을 적용한 경우 동일한 결과(평균)를 유지해야 하거나 둘째, 각 척도에 따라 측정된 값이 상이하다고 해도 응답강도를 이용한 조정이 가능한 경우이어야 한다. 즉, 모든 측정값들이 참값 또는 유사한 값에 수렴(근사)함을 확인할 수 있거나 근사값으로 보정하는 방법이 있는 경우이다.

먼저, 동일한 환산방법 내에서 동일한 결과(평균)를 유지하는지 확인하기 위해 4점, 5점, 11점 척도 문항으로 측정된 결과에 대해 앞서 언급한 환산방법들을 적용하였다. 기존 연구에서 사용한 환산방법들은 ‘측정값×100/선택지 수’(이하 환산방법1), ‘(측정값-1)×100/(선택지 수-1)’(이하 환산방법2), ‘측정값/환산할 척도의 중앙값×환산될 척도의 중앙값’(이하 환산방법3)으로 각 리커트형 척도 문항으로 측정된 값들을 해당 방식으로 환산한 후, 척도 간 *t*검정을 진행하였다. 즉, 4점 척도 문항과 5점 척도 문항 간 측정값 비교, 4점 척도 문항과 11점 척도 문항 간 측정값 비교, 5점 척도 문항과 11점 척도 문항 간 평균 비교를 진행한 것이다. 이 과정에서 11점 척도 문항의 경우에는 환산방법2로 환산하였을 경우에는 음수가 도출되기 때문에 일괄적으로 0, 10, 20... 90, 100으로 환산하였다.

만약 응답집단이 동일하다는 전제하에 4, 5, 11점 척도 문항의 응답 결과가 100점 환산 후에도 일치한다면 세 척도 문항은 환산을 통해 비교 가능하다고 볼 수 있다. 물론 세 척도의 환산된 측정값이 모두 참값이라는 확신은 없지만,

동일한 결과를 나타내기 때문에 세 척도 문항 내에서는 비교가 가능해진다는 의미이다. 그러나 다른 환산 결과를 나타낸다면 척도 문항들이 서로 달리 측정한다는 것으로 비교가 불가능해진다. 이 경우, 두 번째 전제 조건인 응답강도의 보정을 통해서만 비교가 가능하다.

이때 4, 5, 11점 척도 문항의 응답강도 보정은 다음과 같이 설명할 수 있다. 가령, 4점 척도 문항과 11점 척도 문항이 동일한 결과를 가지기 위해서는 4점 척도 문항의 1점을 1.2, 2점을 4, 3점을 6.7, 4점을 9.7로 치환함으로써 11점 척도 문항의 평균과 같아진다고 하자. 11점 척도 문항에 대응하는 4점 척도 문항의 평균은 '1×빈도+2×빈도+3×빈도+4×빈도'가 아니라 '1.2×빈도+4×빈도+6.7×빈도+9.7×빈도'가 되는 것이다. 이 경우, 1.2, 4, 6.7, 9.7이라는 치환될 응답강도를 추정할 수 있다면 4점 척도 문항과 11점 척도 문항 간 평균을 동일하게 만들 수 있다. 이를 위해서는 세 척도 문항의 결과가 모두 동일해지는 응답강도를 추정해야 하는데, 상호 비교 가능한 응답강도의 추정을 위해서는 아래와 같은 식을 전개할 수 있다.

먼저 평균에 대한 공식은 식(1)과 같으며, 관측치 수를 고려한다면 아래와 같이 변형시킬 수 있다.

$$(4점\ 척도의\ 총점) = 1 \times (1의\ 응답빈도) + 2 \times (2의\ 응답빈도) + 3 \times (3의\ 응답빈도) + 4 \times (4의\ 응답빈도)$$

(1)

$$(4점\ 척도의\ 평균) = \frac{1}{n} (4점\ 척도의\ 총점)$$

만약, 4점 척도 문항의 총점을 종속변수 y 로 놓고 1의 응답빈도를 x_1 , 2의 응답빈도를 x_2 , 3의 응답빈도를 x_3 , 4의 응답빈도를 x_4 로, x_1 에 대응하는 응답빈도를 응답빈도 a , x_2 에 대응하는 응답빈도를 응답빈도 b , x_3 에 대응하는 응답빈도를 응답빈도 c , x_4 에 대응하는 응답빈도를 응답빈도 d 로 치환하면 식(2)와 같은 응답강도 추정 방정식을 제시할 수 있고, 총 관측치가 80명이므로 $x_1 + x_2 + x_3 + x_4 = 80$ 이 되어 x_1 이 제거가능하다.

$$y = \text{응답강도}a \times x_1 + \text{응답강도}b \times x_2 + \text{응답강도}c \times x_3 + \text{응답강도}d \times x_4 + \epsilon$$

$$y = \text{응답강도}a \times (80 - x_2 - x_3 - x_4) + \text{응답강도}b \times x_2 + \text{응답강도}c \times x_3 + \text{응답강도}d \times x_4 + \epsilon \quad (2)$$

$$y = \text{응답강도}a \times 80 + (\text{응답강도}b - \text{응답강도}a) \times x_2 + (\text{응답강도}c - \text{응답강도}a) \times x_3 + (\text{응답강도}d - \text{응답강도}a) \times x_4 + \epsilon$$

당연히 y 가 4점 척도 문항의 총점이라고 한다면 응답강도 a 는 1, 응답강도 b 는 2, 응답강도 c 는 3, 응답강도 d 는 4로 추정될 것이다. 그러나 y 가 5점 척도 문항의 총점 혹은 11점 척도 문항의 총점이라면 응답강도 $a \sim d$ 는 각 척도 문항에 대응하는 응답강도가 추정될 것이다. 본 연구에서 조사하였던 만족도 설문은 총 16 설문이었기 때문에 총 16개의 y 값을, 4점 척도 문항을 기준으로 총 16개의 x_1, x_2, x_3, x_4 (5점 척도 문항시에는 x_5 추가)를 갖는다. 이를 일반화하면 식(3)과 같다.

$$\begin{pmatrix} y_1 \text{ (재난방지)} \\ y_2 \text{ (보건의료)} \\ y_3 \text{ (보육)} \\ \vdots \end{pmatrix} = \begin{pmatrix} 80 & x_{21} & x_{31} & \cdots \\ 80 & x_{22} & x_{32} & \cdots \\ 80 & x_{23} & x_{33} & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} \text{응답강도}a \\ \text{응답강도}b - a \\ \text{응답강도}c - a \\ \vdots \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \vdots \end{pmatrix} \quad (3)$$

여기에서, 예를 들어 x_{21} 은 재난방지 문항에 2를 응답한 사람들의 빈도를 나타낸다.

이때 11점 척도 문항에 대응하는 4점, 5점 척도 문항의 응답강도를 추정하기 위해서 재난방지, 보건의료 등 16개 설문의 11점 척도 문항의 총점(각 문항의 평균 \times 80명)을 y 로, 16개 설문의 각 선택지 응답빈도를 4점 척도 문항의 경우에는 $x_2 \sim x_4$, 5점 척도 문항의 경우에는 $x_2 \sim x_5$ 로 놓고 회귀분석을 진행하면, 각 선택지에 대응하는 응답강도를 추정할 수 있다. 여기서 추정된 응답강도는 16개 행정 서비스 만족도 개념에 대한 세부 차원들의 평균적인 응답강도라고 할 수 있다.

마지막으로 가정된 참값의 경우 4점, 5점, 11점 결과를 환산 없이 종합할 수 없기 때문에 11점을 중심으로 환산방법1과 같이 척도의 길이를 조정한 후, 이를 식(4)를 이용하여 평균하였다. 가정된 참값 역시 16개 설문별로 4, 5, 11점 평균이

존재하기 때문에 총 16개의 y 값이 존재하게 되며, 이후 추정과정은 식(3)과 동일하다.

$$\text{가정된 참값} = n \times \frac{1}{3} \times \left(\frac{11}{4} \times 4\text{점 평균} + \frac{11}{5} \times 5\text{점 평균} + 11\text{점 평균} \right), \quad n \text{은 } 80\text{명} \quad (4)$$

IV. 연구 결과

1. 환산방법에 따른 척도 문항별 평균 비교

먼저 각 척도 문항으로 측정한 값을 환산방법1, 환산방법2, 환산방법3으로 환산한 후의 평균값을 비교한 결과는 <표 5>와 같다. 4, 5, 11점 척도 문항을 환산방법1을 통해 환산하여 비교하였을 경우, 4, 5점 척도 문항 비교에서는 재난방지 분야에서, 5, 11점 척도 문항 비교에서는 관광, 교육 분야에서, 4, 11점 척도 문항 비교에서는 경제 분야에서 통계적 차이가 존재하였다. 환산방법2를 통해 환산한 경우에는 4, 5점 척도 문항 비교에서는 재난방지, 농·어촌, 관광분야에서, 5, 11점 척도 문항 비교에서는 보건의료, 농·어촌, 치안분야에서, 4, 11점 척도 문항 비교에서는 거의 모든 분야에서 통계적 차이가 존재함을 확인할 수 있다. 마지막으로 환산방법3을 통해 환산하였을 경우, 5, 11점 척도 문항 비교에서 재난방지, 대중교통, 환경미화, 교육, 전반적 만족도 분야에서, 4, 11점 척도 문항 비교에서는 대중교통, 환경미화, 경제, 교육, 전반적 만족도 분야에서 통계적 차이가 존재하였고, 유일하게 환산방법3의 4, 5점 척도 문항 비교에서만 통계적 차이가 발생하지 않았다.

결과적으로 16개 설문에 대해 4, 5, 11점 척도 문항으로 측정한 경우 본 연구에서 사용된 환산방법으로는 거의 모든 분야에서 5% 유의수준에서 동일한 평균값을 보이지 않고 있다. 오직 환산방법3을 이용하는 경우에만 4점 척도 문항과 5점 척도 문항의 평균 점수가 16개 영역 모두에서 유사하게 나타날 뿐 그 이외에

<표 5> 환산방법에 따른 척도 문항별 평균 비교(*t*검정)¹³⁾

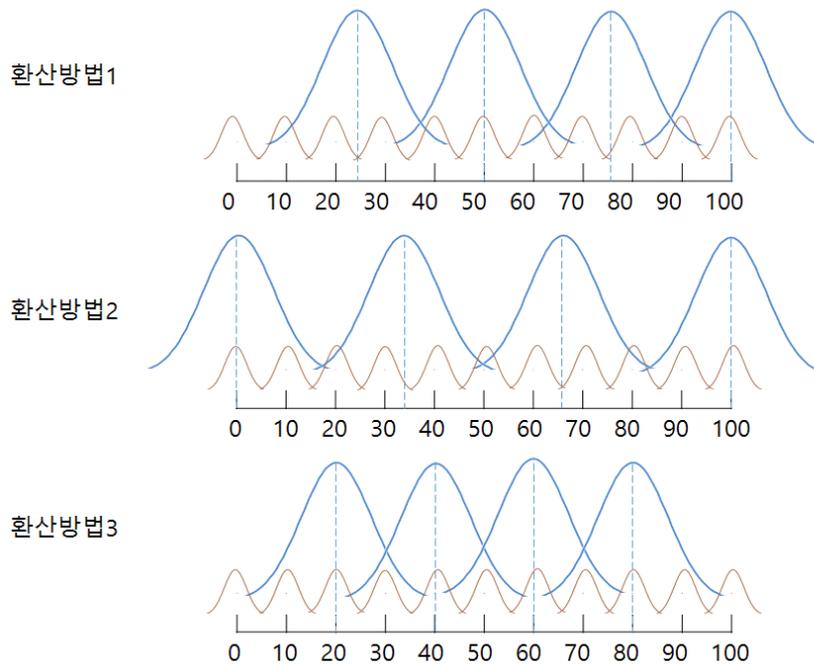
척도	원 점수			환산방법1에 대한 유의확률			환산방법2에 대한 의확률			환산방법3에 대한 유의확률		
	4점	5점	11점 (0~ 10점)	4, 5점 척도 비교	5, 11점 척도 비교	4, 11점 척도 비교	4, 5점 척도 비교	5, 11점 척도 비교	4, 11점 척도 비교	4, 5점 척도 비교	5, 11점 척도 비교	4, 11점 척도 비교
재난 방지	2.76	3.04	5.79	0.0445	0.0538	0.8072	0.0061	0.3665	0.0000	0.0604	0.0147	0.3556
보건 의료	2.88	3.43	6.13	0.603	0.8513	0.7075	0.2282	0.0153	0.0000	0.8553	0.1297	0.1562
보육	2.55	3.00	5.46	0.6516	0.2106	0.4177	0.1914	0.1078	0.0050	0.6695	0.1393	0.2288
농·어촌	2.56	2.94	5.24	0.2859	0.1964	0.9318	0.0445	0.0216	0.0000	0.2850	0.1871	0.6736
도로 관리	2.81	3.36	6.28	0.785	0.3935	0.6035	0.4258	0.2379	0.0480	0.9471	0.0563	0.0598
대중 교통	2.93	3.49	6.70	0.6502	0.194	0.4739	0.3176	0.3985	0.0670	0.8916	0.0034	0.0049
환경 미화	2.96	3.63	6.95	0.961	0.3128	0.2838	0.6367	0.3744	0.1610	0.6664	0.0038	0.0007
경제	2.26	2.85	5.31	0.2816	0.0618	0.002	0.8841	0.2337	0.2620	0.3607	0.0620	0.0058
관광	2.31	2.55	4.83	0.2339	0.0087	0.2637	0.0362	0.3913	0.0060	0.1554	0.0550	0.5159
문화	2.71	3.26	6.05	0.9055	0.3385	0.405	0.4519	0.1955	0.0410	0.9642	0.0724	0.0570
주거	2.60	3.10	5.68	0.8339	0.2525	0.3719	0.3300	0.1157	0.0140	0.8945	0.1022	0.1203
교육	2.63	3.16	6.11	0.9779	0.0389	0.0568	0.4139	0.4861	0.1510	0.9299	0.0035	0.0029
노인 복지	2.50	3.00	5.56	1.0000	0.1355	0.1623	0.4683	0.1860	0.0420	1.0000	0.0642	0.0638
치안	2.65	3.28	5.78	0.6117	0.798	0.4321	0.7931	0.0120	0.0050	0.4983	0.2653	0.0883
전반적 만족도	2.71	3.18	5.94	0.4032	0.0941	0.4475	0.0869	0.1203	0.0010	0.5157	0.0091	0.0324
세금 대비 만족도	2.24	2.71	4.93	0.6733	0.0556	0.0206	0.5566	0.0936	0.0240	0.8450	0.1420	0.0942

13) 진한 글씨로 표시된 부분은 유의수준 0.05에서 통계적으로 유의한 항목이다.

는 적게는 1개에서 많게는 13개 설문결과에서 환산된 평균값에 통계적으로 유의미한 차이가 있었다. 그러나 환산방법1의 경우에도 척도 문항에 따라 1~2개 영역에서 평균값 차이가 발생하고 있는데 이러한 차이가 5% 유의수준에서 우연에 의해 발생한 차이일 수 있어 실질적으로는 모든 영역에서 동일한 평균값을 나타내는 것일 수도 있다.

따라서 보다 정확한 판단을 위해서는 16개 설문 모두에서 유사한 평균값을 보인 경우와 일부 설문에서만 평균 차이가 발생한 경우에 대해, 앞서 제시한 응답강도를 보정하는 방법을 병용할 필요가 있다. 만약 응답강도를 보정하는 방법을 통해 추정된 결과가 환산결과와 유사하다면 해당 환산방법에 따른 척도 문항 간 비교가 가능하다고 볼 수 있을 것이다. 가령, 환산방법3의 검증은 식(3)을 통해 y 에 5점 척도의 총점을, x 에 4점 척도의 응답빈도를 입력하여, 응답강도 a 가 1.2, 응답강도 b 가 2.4, 응답강도 c 가 3.6, 응답강도 d 가 4.8에 대응하는지 확인해 보는 것이다. 해당 검증은 이하의 각 척도에 대한 응답강도를 추정하는 과정에서 같이 살펴보고자 한다.

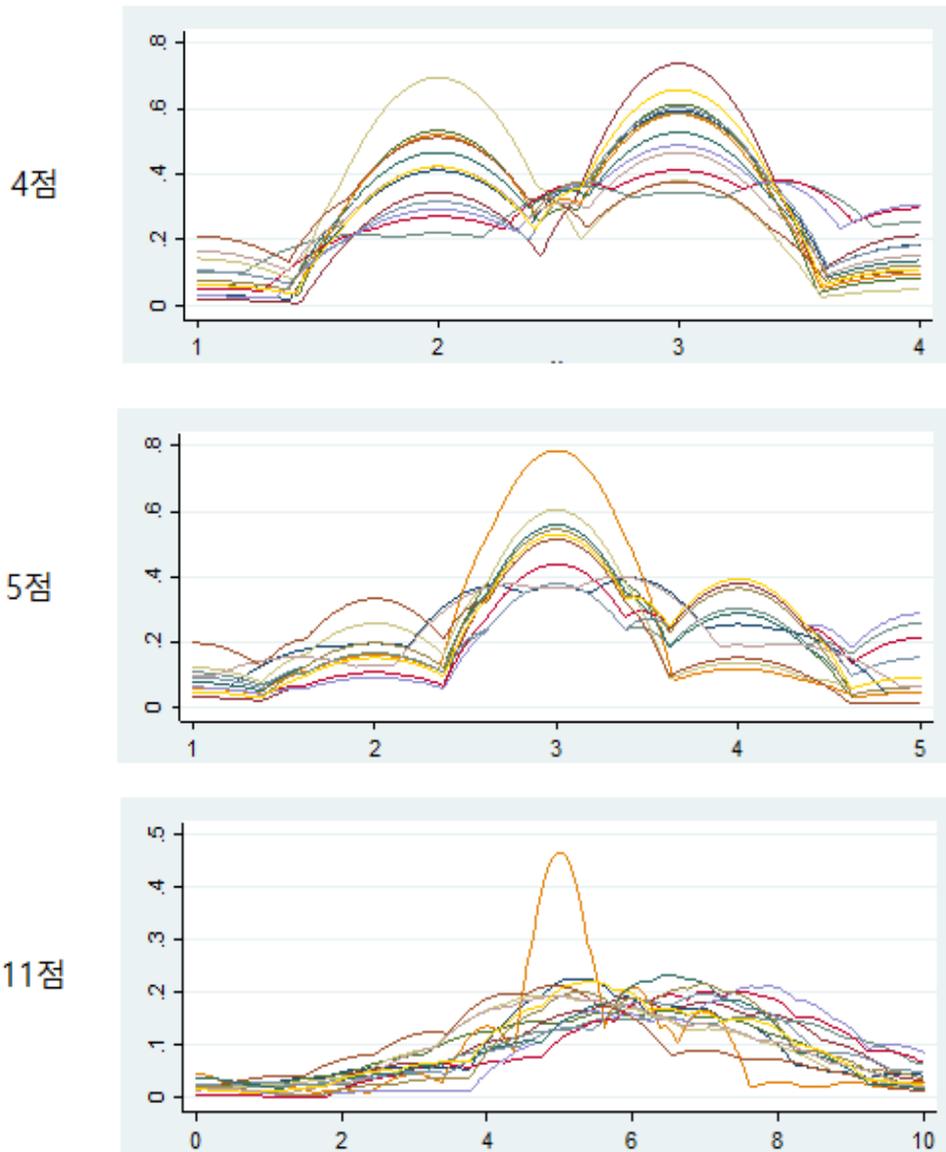
그렇다면 기존 연구들에서 제시한 환산방법들이 왜 동일한 평균값을 나타내지 않을까? 앞서 언급한 바와 같이 평균은 응답강도와 응답빈도의 함수이며 환산방법은 이 중 각 척도의 응답강도를 변형시켜 101(0~100)점 척도에 대응하게끔 하는 방식이다. 이때 환산방식에 따라 그 변형정도에 왜곡이 발생할 여지가 있다. <그림 1>은 4점과 11점 척도 문항이 100점에 대응할 때 응답강도가 어디에 위치하는지를 각각의 환산방법에 따라 나타낸 것으로서, 환산과정에서 척도의 선택지에 따라 과잉 혹은 과소 대표되는 현상을 보여주고 있다. 좀 더 구체적으로 보면, 환산방법1을 사용하는 경우 4점 척도 문항의 1점은 25점이, 2점은 50점이 되는데 환산된 점수를 중심으로 응답강도가 정규분포를 이룬다고 가정해 보면, 4점 척도 문항에서 1점을 선택한 응답자의 경우 100점으로 환산했을 때 0~37.5점까지의 범위에 해당하게 되며 이는 전체 100점이라는 응답강도 중 37.5%에 해당한다. 반대로 4점 척도 문항의 4점은 100점 중 87.5~100점까지이고 이는 전체 응답강도 중 12.5%만을 반영하게 된다. 이렇게 4점 척도 문항 내에서 각각의 선택지가 100점으로 환산되는 과정에서 지나치게 넓거나 좁은 영역을 대표하게 되어 실제(참) 응답강도를 왜곡하는 현상을 낳을 수 있다는 것이다.



〈그림 1〉 각 환산방법에 따른 4점 척도 문항과 11점 척도 문항의 응답강도 대응

환산방법2와 환산방법3 역시 유사한 문제점을 안고 있다. 환산방법1의 경우 4점 척도 문항의 1점이 4점에 비해 반영범위를 과다 대표하면서 분포가 전체적으로 우측으로 밀려올라간 양상이라면, 환산방법2의 경우에는 4점 척도 문항의 2점과 3점이 1점과 4점보다 과잉 대표하며, 반대로 환산방법3에서는 1점과 4점이 2점과 3점보다 과잉 대표하는 현상이 발생하고 있다. 즉, 이러한 과잉, 과소 대표 문제는 척도 문항의 선택지 수가 적을수록 실제(참) 응답강도를 반영하지 못할 위험성이 크고 오차의 크기가 커질 수 있음을 의미하게 된다.

환산과정에서 척도의 과잉, 과소 대표 문제에 더해 응답분포의 문제 역시 환산을 통한 왜곡을 유발할 수 있다. 평균은 응답강도와 응답빈도의 곱으로 표현되는데 이때 응답분포는 응답빈도들의 종합으로서 4점 척도 문항이든, 5점 척도 문항이든, 또는 11점 척도 문항이든 모두 동일한 응답분포를 보인다면 척도 문항 간 비교 문제는 응답강도에 한정되겠지만, 실제의 경우 응답자는 척도 문항에 따라 달리 반응하는 경향이 있다. 즉, 동일 설문에 대해서도 사용하는 척도 문항에 따라 응답분포도 달라질 수 있다는 것이다.



〈그림 2〉 14개 행정서비스 분야에 대한 만족도 분포(Kernel 밀도 함수)¹⁴⁾

〈그림 2〉는 14개 행정서비스 분야에 대한 실제 응답분포로서 4점 척도 문항에 대한 응답은 2점과 3점을 중심으로 정규분포를, 5점 척도 문항에 대한 응답은

14) 5점과 11점 척도 문항에서 중앙값의 가장 높은 봉우리는 농·어업 행정서비스 만족도에 대한 응답이다. 설문이 진행된 창원 지역의 응답자 중 해당 분야를 접하고 있는 응답자들이 다른 영역에 비해 상대적으로 적기 때문에 모름에 대한 응답이 중앙값으로 나타난 것으로 추정된다.

3점을 중심으로 정규분포 형태를 띠지만, 11점(0~10점) 척도 문항에 대한 응답은 중앙값인 5를 기준으로 대칭을 이루지 않고 오른쪽으로 치우친 모습을 보이고 있고, 특히 1점과 같은 최솟값 근처에서의 빈도는 매우 낮았다. 즉, 응답자들이 11점 척도 문항으로 응답하는 경우 4점, 5점 척도 문항에 반응하는 것보다 상대적으로 더 높은 수준을 중심으로 응답하고 있었으며, 이는 동일한 설문에 대해서도 어떤 척도 문항을 이용하느냐에 따라 응답행태가 다를 수 있음을 의미한다.

종합하면, 환산방법에 의해 응답정보의 왜곡이 일부 발생할 수 있을 뿐 아니라 척도 문항의 선택지 수에 따라 달라지는 응답 분포 문제가 결합되면 단순한 산술계산으로는 비교 가능한 환산 값을 산출하기는 어렵다는 점이다. 따라서 리커트형 척도 문항 간 비교를 위해서는 단순 산술식을 이용한 변환이 아니라 각 척도에 반응하여 나타난 실제 응답 분포들을 고려하여 상호 대응하는 응답강도를 추정하여 보정하는 방식이 필요한 것으로 보인다. 다음 절에서 이에 대해 살펴본다.

2. 응답강도 추정

실제 응답 분포를 반영한 상태에서 각 척도의 응답강도를 추정하는 식은 (3)으로 유도되었다. 이를 이용하여 먼저, 11점 척도 문항을 중심으로 4, 5점 척도 문항의 선택지가 대응하는 응답강도를 추정하고자 하였다. 참값을 확인할 수 없는 상황에서 4, 5, 11점 척도 문항의 측정 결과(평균) 중 어떠한 결과가 참값에 근사한지 알 수 없으나, 이러한 상황에서도 서로 동일한 결과를 보인다면 각 척도 간 상대 비교는 가능하기 때문이다. 두 번째로 앞서 언급하였던 가상의 참값을 이용하여 간접적으로나마 각 척도 문항의 응답강도를 추정하고자 하였다. 11점 척도 문항을 기준으로 한 첫 번째 추정은 상대적인 응답강도만을 알려줄 뿐 절대적인 응답강도를 예상할 수 없다. 따라서 가상의 참값을 통해 각 척도의 실제(참) 응답강도를 확인하고 각 척도 문항의 특성을 파악하고자 하였다.

먼저 식 (3)을 이용한 4점, 5점, 11점 척도 문항 간 상대적인 응답강도에 대한 추정 결과는 <표 6>에 제시되었는데, 이는 11점 척도 문항 혹은 5점 척도 문항을 등간척도로 가정하였을 때 각각에 대해 4점과 5점 척도 문항의 응답 강도들을 회귀식을 이용해 추정한 것이다. 구체적으로 보면 4점 척도 문항의 경우 1점을

1.64(=141.4/80)로, 2점을 4.84(=1.64+3.195)로, 3점을 5.96(=1.64+4.322)로, 4점을 10.18(=1.64+8.533)로 치환하였을 때 4점 척도 문항과 11점 척도 문항이 동일한 결과(평균)를 갖는다는 것이다. 5점 척도 문항의 경우에도 유사한 방식으로 11점 척도 문항과 동일한 평균값을 갖도록 치환할 수 있고, 4점 척도 문항이 5점 척도 문항과 같은 결과를 갖기 위해서는 1점이 0.94(=75.55/80)로, 2점이 2.53(=0.94+1.587), 3점이 3.43(=0.94+2.482), 4점이 4.93(0.94+3.986)로 치환되어야 한다.

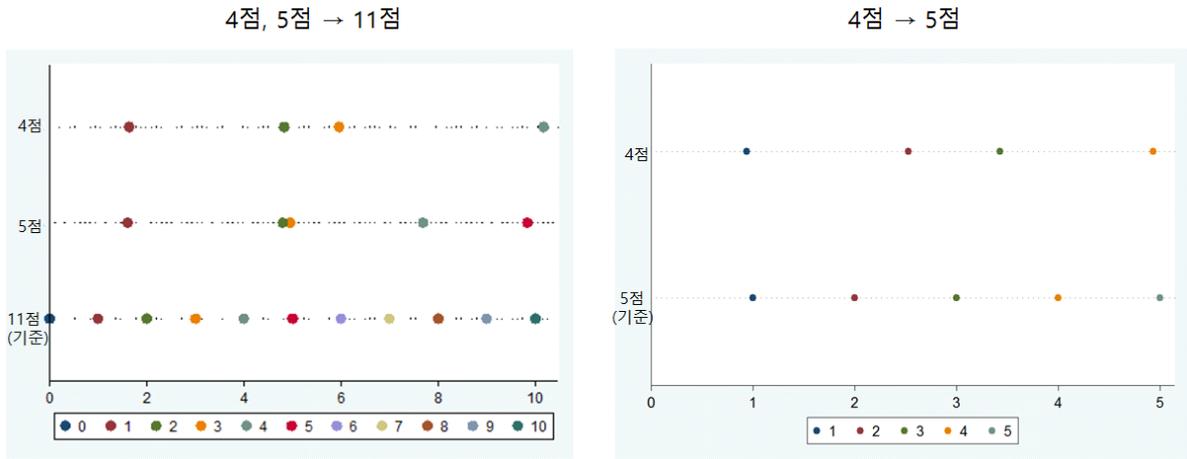
<표 6> 상호 대응 응답강도 추정¹⁵⁾

변수명	추정 계수	4점→11점	5점→11점	4점→5점
총점(y)		11점 총점	11점 총점	5점 총점
2 번 선택지 응답자(x2)	응답강도b - 응답강도a	3.195 ** (2.362)	3.188 (3.31)	1.587 (1.167)
3 번 선택지 응답자(x3)	응답강도c - 응답강도a	4.322 (1.511)	3.335 (2.084)	2.482 *** (0.747)
4 번 선택지 응답자(x4)	응답강도d - 응답강도a	8.533 *** (2.161)	6.069 *** (1.92)	3.986 *** (1.068)
5 번 선택지 응답자(x5)	응답강도e - 응답강도a		8.219 *** (2.569)	
상수	80*응답강도a	131.4 (137.9)	129.4 (166.5)	75.55 (68.16)
관측치		16	16	16
R-squared		0.863	0.941	0.860

*** p<0.01, ** p<0.05, * p<0.1, 괄호 안은 표준오차

15) 여기에서 유의확률은 가장 낮은 선택지로부터 얼마나 멀리 떨어지는가에 대한 검증 결과이기 때문에 당연히 높은 선택지일수록 유의확률이 낮아지게 된다. 따라서 유의확률이 제공하는 정보의 가치보다는 추정된 계수가 제공하는 정보의 가치에 보다 주목할 필요가 있다.

또한 이를 토대로 각 척도의 응답강도를 기준척도에 적용한 것이 <그림 3>이다.



선택지	추정계수	4점 → 11점	5점 → 11점	4점 → 5점
1	응답강도 a	1.64	1.62	0.94
2	응답강도 b	4.84	4.81	2.53
3	응답강도 c	5.96	4.95	3.43
4	응답강도 d	10.18	7.69	4.93
5	응답강도 e		9.84	

<그림 3> 상호 대응 응답강도 분포

이러한 추정결과로 확인할 수 있는 것은 하나의 척도 문항으로 측정된 값을 다른 척도 문항으로 전환하는 것이 단순히 동일한 값을 곱해 주거나 더하는 즉, 선형화를 통해 해결하기 어렵다는 점이다. <그림 3>에서 볼 수 있듯이 11점 척도 문항을 등간으로 가정할 때 4점 척도 문항에 대한 응답행태는 상대적으로 2점과 3점의 차이가 좁고 3점과 4점의 차이가 넓어 등간으로 보기 어렵다. 특히, 11점 척도 문항과 대응하는 5점 척도 문항의 경우 2점과 3점은 차이가 드러나지 않을 만큼 근접해 있는 실정이다. 이러한 상황에서 등간 또는 그 비율을 유지하고자 하는 환산은 적절하지 않음을 알 수 있다. 물론 5점 척도 문항을 등간으로 가정하는 경우 4점 척도 문항 역시 2점과 3점의 차이가 좁지만 상대적으로 그 차이가

그리 크지 않다는 점도 알 수 있다.

앞서 환산방법3으로 4점, 5점 척도 문항을 비교하였을 때 두 척도의 결과(평균)가 16개 모든 영역에서 유사하다는 실증 결과를 얻었다. 이때 환산방법3은 4점 척도의 1점, 2점, 3점, 4점을 5점 척도의 1.2점, 2.4점, 3.6점, 그리고 4.8점에 대응하는 환산을 의미하는 것이었다. 다른 방식의 추정이지만 실제 4점 척도 문항의 선택지가 5점 척도 문항에 어떻게 대응하는지를 <그림 3>을 통해서도 살펴볼 수 있는데, 약간의 차이를 제외하고는 추정결과와 환산방법3의 결과와 근접함을 알 수 있다. 결과적으로 이러한 응답강도의 근접함이 평균 차이를 미미하게 한 것이고 통계적 오차 범위 내에 위치하게끔 한 것이라 할 수 있다. 이에 반해 일부의 평균차이가 발생하였던 환산방법1의 경우, 4점과 5점 척도 문항의 선택지가 11점 척도 문항에 대응하는 과정에서 4점 척도 문항의 2점과 3점이 11점 척도의 5점과 7.5점에, 5점 척도 문항의 2점과 3점이 4점과 6점에 대응해야 하지만 실제 응답강도 추정 결과와는 큰 차이가 존재하였다. 반대로 5점 척도 문항에 대응하는 4점 척도 문항의 경우에는 환산방법1과 응답강도 추정 결과가 별반 큰 차이를 나타내지 않음을 알 수 있어 환산방법1로 환산 가능하다고 볼 수 있다.

이상의 논의를 종합하면, 일부의 사례를 제외하고는 현재 제시되는 100점 환산 과정을 통해 다양한 리커트형 척도 문항으로 측정된 결과를 비교하는 것은 다소 위험하다는 점을 알 수 있다. 이는 현재의 환산방식이 선형식을 이용한 비율과 간격의 유지에 초점을 맞추고 있는 데 반해, 실제 응답행태는 어떤 척도 문항을 이용하는지에 따라 응답강도와 응답빈도(분포)가 달라질 수 있기 때문이다. 일부 환산 가능한 사례의 경우조차 일반적이기보다는 예외적인 것으로 볼 수 있다는 점에서 서로 다른 척도를 이용한 측정값 간의 비교는 매우 제한적인 상황에서만 이루어져야 할 것으로 판단된다.

한편, 가상의 참값을 활용하여 4, 5, 11점 척도의 응답강도를 추정한 결과는 <표 7>과 <그림 4>인데, 척도 문항 간 상대 비교결과와는 다른 양상을 보이는 것을 알 수 있다. 참고로 가상의 참값을 활용하여 각 척도 문항의 응답강도를 추정하는 과정은 식(4)에 제시된 가정된 참값을 종속변수로 하여 4점 척도 문항의 경우에는 1~4에 대한 응답강도를, 5점 척도 문항의 경우에는 1~5, 11점 척도 문항의 경우에는 0~10에 대한 응답강도를 회귀분석을 활용하여 추정하는 것이다. 이때,

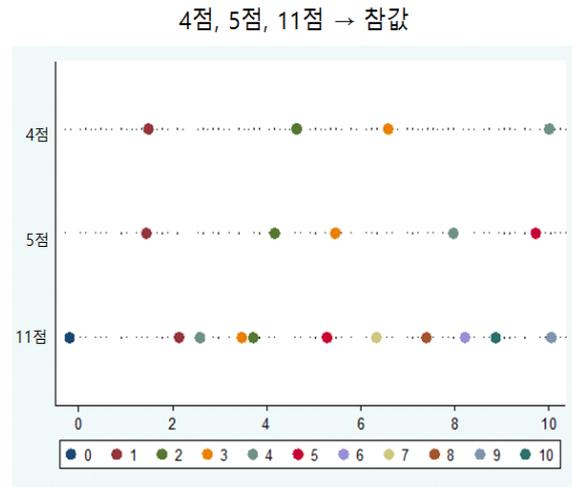
<표 7> 가상의 참값에 대응하는 4, 5, 11점 척도 문항의 응답강도 추정

변수명	추정계수 (4, 5점 척도)	추정계수 (11점 척도)	4점→참값	5점→참값	11점→참값
총점(y)			가상의 참값	가상의 참값	가상의 참값
1 번 선택지 응답자(x1)		응답강도 b - 응답강도 a			2.333 (5.393)
2 번 선택지 응답자(x2)	응답강도 b - 응답강도 a	응답강도 c - 응답강도 a	3.146* (1.544)	2.73 (2.309)	3.903 (3.039)
3 번 선택지 응답자(x3)	응답강도 c - 응답강도 a	응답강도 d - 응답강도 a	5.094*** (0.988)	4.010** (1.454)	3.667 (2.647)
4 번 선택지 응답자(x4)	응답강도 d - 응답강도 a	응답강도 e - 응답강도 a	8.518*** (1.413)	6.521*** (1.34)	2.771 (2.671)
5 번 선택지 응답자(x5)	응답강도 e - 응답강도 a	응답강도 f - 응답강도 a		8.273*** (1.793)	5.474** (2.109)
6 번 선택지 응답자(x6)		응답강도 g - 응답강도 a			8.408** (2.485)
7 번 선택지 응답자(x7)		응답강도 h - 응답강도 a			6.518** (2.305)
8 번 선택지 응답자(x8)		응답강도 i - 응답강도 a			7.591** (2.198)
9 번 선택지 응답자(x9)		응답강도 j - 응답강도 a			10.24*** (2.23)
10 번 선택지 응답자(x10)		응답강도 k - 응답강도 a			9.064** (3.468)
상수	80*응답강도 a	80*응답강도 a	119.2 (90.19)	115.8 (116.2)	-15.2 (167.2)
관측치			16	16	16
R-squared			0.943	0.973	0.986

*** p<0.01, ** p<0.05, * p<0.1, 괄호 안은 표준오차

위에서 확인한 상대적인 응답강도에 대한 추정의 경우와 동일하게 4, 5, 11점 척도의 가장 낮은 선택지는 상수로 추정된다. 가령, 4점 척도 문항의 경우에는 상수 $119.2/80=1.49$, 5점 척도 문항의 경우에는 상수 $115.8/80=1.4475$, 11점 척도 문항

의 경우에는 상수 $-15.2/80=-0.19$ 로 추정되는데, 이를 바탕으로 각 척도 문항의 여타 선택지들의 대응 응답강도는 회귀분석 추정치에 상수를 통해 추정된 응답강도를 더해서 결정된다. 이에 대한 결과는 <그림 4>의 표에서 확인 가능하다.



선택지	4점 → 참값	5점 → 참값	11점 → 참값
0			-0.19
1	1.49	1.4475	2.143
2	3.146	4.1775	3.713
3	5.094	5.4575	3.477
4	7.028	7.9685	2.581
5		9.7205	5.284
6			8.218
7			6.328
8			7.401
9			10.05
10			8.874

<그림 4> 가상의 참값에 대응하는 4, 5, 11점 척도 문항의 응답강도

먼저 척도 문항 간 상대 비교결과를 나타낸 <그림 3>에서는 11점 척도 문항에 대응하는 5점 척도 문항의 2점과 3점이 응답강도상 차별성이 거의 존재하지 않았지만 이를 가상의 참값에 대응시키면 <그림 4>에서 볼 수 있듯이 구분 가능할 정도로 떨어져 위치한다. 하지만 <그림 4>에서 가장 주목할 점은 4, 5, 11점 척도 문항 모두 등간을 인정하기는 곤란해 보일 뿐 아니라 특히 11점 척도 문항의 응답강도 분포가 가장 일그러져 있다는 것이다. 4점과 5점 척도 문항의 경우에는 각 선택지의 서열 정보가 그대로 유지되어 있고 그 간격 역시 대체로 인지할 정도이지만, 11점 척도 문항의 경우에는 0과 1 사이의 간극이 크고, 1~4점의 차이는 매우 작으며, 6~10점의 경우 응답강도의 서열조차 유지되지 않고 있다.

이러한 사실은 11점 척도 문항을 마주하는 응답자의 응답강도에 대한 인식이 매우 혼재하거나 제약적일 수 있음을 의미한다. 다시 말하자면, 선택지 수가 많아지면서 일부 선택지들의 응답강도에 대한 차별성을 식별하지 못하거나 오류를 범할 수 있다는 것이다. 응답자는 1점보다는 2점, 2점보다는 3점이 높은 응답강도를 지닌다고 생각하지만, 척도의 길이가 늘어나 예를 들어, 30점 척도에서는 일부 응답자는 21, 22, 23, 24점의 차이를 인지하지 못하거나, 때에 따라서는 일부 응답자가 생각하는 24점의 응답강도가 다른 응답자가 생각하는 20점과 동일할 수 있다. 이러한 상황은 기존의 연구들이 척도의 선택지 수를 고려할 때 즉, 몇 점 척도 문항을 이용할지를 고려할 때 주로 타당성, 신뢰성, 집단 간 식별력, 응답시간 등을 주요 고려요인으로 상정하였지만, 본 연구의 결과에 비추어 볼 때 선택지 간 식별력¹⁶⁾ 역시 고려요인이 되어야 함을 의미한다. 4점 척도 문항에서 3점에 대해 응답자들이 인식하는 응답강도를 선택지의 수가 늘어날수록 동일한 응답강도를 일부는 6점으로 혹은 7점 또는 8점으로 간주할 수도 있다는 것이다. 따라서 이러한 척도 선택지에 대한 식별력 문제는 척도의 선택지 수 증가에 대한 제약요인으로 작용할 것으로 보인다.

16) Preston & Colman (2000)이 언급한 척도 식별력(discriminating power)은 환산방법2를 활용하여 응답자의 하위 그룹(0~25%)과 상위 그룹(80~100%) 간 평균 차이의 크기를 언급한 개념으로 본 연구에서 제기하는 식별력과는 다르다.

V. 결론

본 연구는 공공분야에서 다양한 만족도 조사가 진행되고 있는 상황에서 서로 다른 개수의 선택지를 갖는 리커트형 척도 문항을 비교할 수 있는지 그 가능성을 타진해보고자 시작되었다. 만약 이들 간 비교가 가능하다면, 다양한 선택지 개수의 척도 문항들로 조사된 결과들을 동시에 사용할 수 있기에 조사 결과의 활용 범위가 비약적으로 넓어지게 된다. 하지만 그렇지 않다면 서로 다른 리커트형 척도 문항을 사용한 조사들은 근본적으로 비교 불가능한 독립 조사가 된다. 따라서 리커트형 척도 문항의 비교 가능성에 대한 연구는 한정된 예산하에서 가장 많은 정보를 확보해야 하는 조사의 효율성과 연관되어 있다.

본 연구에서 제시한 행정서비스 만족도 분야에 대한 서베이 실험 결과는 기존의 환산방법으로는 극히 제한적인 상황에서만 비교가 가능하다는 것이다. 이러한 결론이 의미하는 것은 조사 설계 과정에서 척도를 선택할 때 조사의 활용 범위 문제를 염두에 둘 필요가 있다는 것이다. 즉, 조사 결과의 활용 과정에서 함께 비교할 설문문항 혹은 타 조사와 척도를 동일하게 유지하는 것이 척도 문항 비교의 위험성을 피해갈 수 있는 방향이라는 것이다. 특히, 다년도 조사의 경우에는 이후 척도 문항 변경에 따른 비교 가능성이 문제가 될 수 있기 때문에 기존의 척도 문항을 유지하는 것이 효율적이다.

그렇다면 서로 다른 척도 문항에 대한 비교가 불가피한 상황에서 어떻게 하면 척도 비교의 위험성을 그나마 줄일 수 있을 것인가? 첫째, 당연하게도 척도 문항에 따른 응답분포가 유사할수록 유사한 결과를 얻을 가능성이 높아진다. 척도 문항의 응답분포가 유사하다면, 결과 값의 차이는 응답강도로부터 파생된 오차만이 영향을 미친다. 이를 위해서는 척도 간 선택지 수의 차이가 크지 않아야 할 것이다. 앞서, 기존 환산방법을 통해 4, 5, 11점 척도 문항들을 비교하였을 때 4점과 5점에 대한 통계적 차이 빈도는 4점과 11점, 5점과 11점 비교보다 많지 않았다. 즉, 척도 문항의 선택지 수의 차이가 적을 때 두 척도 문항 간 응답분포가 유사할 가능성이 높다는 것이다.

둘째, 다양한 환산방법 중 환산과정에서 척도 문항 선택지가 과잉 혹은 과소 대표되는 현상이 적은 환산방법을 택하는 것이 유리하다. 본 연구에서 제시한 세 가지 환산방법 중 환산방법 그 자체로만 보았을 때에는 환산방법3이 이러한 현상이 적음을 확인하였다. 따라서 응답분포가 유사하다는 전제하에서는 환산방법3을 선택하는 것이 그나마 차이를 줄일 수 있는 방안이라고 볼 수 있다. 그러나 아무리 과잉 혹은 과소 대표되는 현상이 적은 환산방법을 택한다고 해도 응답분포가 상이하다면 측정 결과는 동일하지 않을 것이다.

그러나 본 연구는 다음과 같은 한계를 가지고 있다. 첫째, 본 연구의 대상이 되는 지방행정서비스 만족도는 다양한 변인들에 의해 영향을 받는다. 본 연구는 응답자들을 인구사회학적 변인들로 층화 추출함으로써 집단 간 동질성을 최대한 확보하고자 하였지만, 이 과정에서 행정서비스 만족도에 영향을 미치는 제3의 요인을 간과하였을 가능성이 존재한다. 만약, 행정서비스 만족도에 영향을 미치는 다양한 요인들을 통제할 수 있다면 연구의 정확도가 더욱 향상될 수 있었을 것이다. 둘째, 본 연구에서는 활용빈도가 높은 세 가지 환산방법을 택하여 이를 검증하고자 하였지만 이밖에도 다양한 선형·비선형 환산방법이 존재한다. 따라서 본 연구의 결과를 다른 모든 환산 방법에 확대 적용하는 것은 위험하며 여타 환산방법에 대해서는 별도의 검증 과정이 진행되어야 할 것이다. 셋째, 본 연구에서 논의하지 않았지만 홀수(odd) 선택지와 짝수(even) 선택지의 활용은 중간값(또는 중립 선택지)의 존재 여부에 따른 응답 차이가 발생할 수 있다. 양자 모두 상대적 장·단점을 갖고 있기에 연구 의도에 따라 선택되지만 이로 인한 측정 결과에서의 차이는 결코 간과할 수 없다는 것이 기존 연구의 결과이다. 다만 본 연구에서 주목한 환산 및 비교 가능성에 있어서 이로 인한 차이의 보정 가능성은 또 다른 문제이기에 추후 연구로 남겨 두었다.

본 연구는 행정서비스 체감 만족도 조사의 활용에 대한 의문에서 시작되었지만, 시사점은 훨씬 광범위하다. 가령, 연구 과정에서 새로운 변수를 만들기 위해 각기 다른 척도들로 조사된 응답들을 환산하여 합산하는 방식이 이미 널리 사용되고 있으며, 또한 각종 선거 과정에서 이루어지는 지지도 조사의 발표 역시 척도에 대한 정보 제공 없이 평균값 중심으로 제시되고 있음은 비일비재하다. 또한 리커트형 척도 문항은 그 어떤 척도보다 빈번히 또한 널리 활용되고 있기에 본질

적 한계에 무감각해지는 경향도 있다는 점에서 그 사용과 활용에 보다 각별한 주의를 기울일 필요가 있다.

참고 문헌

- 고길곤. 2014. “설문순서에 따른 응답오차의 유무와 원인에 대한 연구: 정책이해 및 선호도에 대한 설문조사 사례를 중심으로.” 《한국행정학회 동계학술발표논문집, 2014(단일호)》168-189.
- 고길곤·탁현우·강세진. 2014. “설문조사에서 문항의 역코딩 여부가 응답에 미치는 영향.” 《한국정책학회 동계학술발표논문집, 2014(단일호)》1068-1088.
- 권기현. 2007. “정부개혁의 방향과 과제.” 《국정관리연구》2(2): 31-59.
- 김권현·유동주·김형준·김청택. 2015. “연구논문: 설문지의 길이가 응답의 질에 미치는 영향.” 《조사연구》 16(1): 1-48.
- 송인덕·조성겸. 2013. “서베이 모드와 사회적 바람직함 편향: 민감한 주제에 대한 전화 면접, 대면 면접, 온라인 서베이 응답 비교.” 《조사연구》 14(2): 165-199.
- 앤더슨컨설팅. 2000. 《CRM 고객관계관리》. 대청.
- 유홍준·김월화. 2014. “연구논문: 한국형 직업위세 측정 척도 개발.” 《조사연구》 15(3): 1-29.
- 이상민·신현암·최순화. 2000. 《인터넷 시대의 고객관계관리(CRM)》. 삼성경제연구원.
- 이주현. 2011. “설문조사시 질문순서가 응답결과에 미치는 영향.” 《추계학술대회 발표논문집, 2011(단일호)》 307-308.
- 최병대. 2003. “특집/행정서비스와 고객만족: 고객만족과 신속한 행정서비스.” 《지방행정》 52(594): 37-47.
- Bending, A. 1954. “Transmitted Information and the Length of Rating Scales.” *Journal of Experimental Psychology* 47(5): 303-308
- Garland, R. 1991. “The Mid-point on a Rating Scale: Is It Desirable.” *Marketing Bulletin* 2(1): 66-70.

- Green, P.E., V.R., and Rao. 1970. "Rating Scales and Information Recovery: How Many Scales and Response Categories to Use?" *The Journal of Marketing* 34(3): 33-39.
- Jamieson, S. 2004. "Likert Scales: How to (ab)Use Them." *Medical Education* 38(12): 1217-1218.
- Lehmann, D.R. and J. Hulbert. 1972. "Are Three-point Scales Always Good Enough?" *Journal of Marketing Research* 9(4): 444-446.
- Levine, L.H. and S. Baden. 1978. "The Effect of Scale Size on Obtaining Information in Preference Testing." *Quality and Quantity* 12(2): 169-174.
- Loken, B., P. Pirie, K. Virnig, R.L. Hinkle. and C.T. Salmon. 1987. "The Use of 0-10 Scales in Telephone Surveys." *Journal of the Market Research Society* 29(3): 353-362.
- Matell, M.S. and J. Jacoby. 1971. "Is There an Optimal Number of Alternatives for Likert Scale Items? Study I: Reliability and Validity." *Educational and Psychological Measurement* 31: 657-674.
- Maxell, M.S. and J. Jacoby. 1972. "Is There an Optimal Number of Alternatives for Likert-scale Items?" *Journal of Applied Psychology* 56(6): 506-509.
- Preston, C.C. and A.M. Colman. 2000. "Optimal Number of Response Categories in Rating Scales: Reliability, Validity, Discriminating Power, and Respondent Preferences." *Acta Psychologica* 104(1): 1-15.
- 2006년도 국립춘천병원 고객만족도 조사
http://www.google.co.kr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja&uact=8&ved=0ahUKEwjQsIXEttzQAhWEj5QKHf3ZB3gQFggfMAE&url=http%3A%2F%2Fwww.cnmh.go.kr%2Fcmnh%2Fjsp%2FboardDownload.jsp%3Fboard_id%3DCNMH_BOARD_DATA%26seq%3D262%26idx%3D1&usg=AFQjCNEY-pFefj_KONLv3TiTDIHckiLNRg
- 2015년도 전문연구정보 활용사업 이용자 만족도 조사.
<http://www.mathnet.or.kr/poll/2015%EB%85%84%EB%8F%84%EC%9D%B4%EC%9A%A9%EC%9E%90%EB%A7%8C%EC%A1%B1%EB%8F%84%EC%A1>

%B0%EC%82%AC%EA%B2%B0%EA%B3%BC.pdf

2008년 별정우체국연합회 기타공공기관 고객만족도(PCSI).

[https://www.google.co.kr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=5&ved=0ahUKEwiX0buHt9zQAhVCpZQKHdeMCroQFgg1MAQ&url=https%3A%2F%2Fwww.popa.or.kr%2Fupfile%2F2008.%25EA%25B3%25A0%25EA%25B0%259D%25EB%25A7%258C%25EC%25A1%25B1%25EB%258F%2584%25EB%25B3%25B4%25EA%25B3%25A0%25EC%2584%259C_01\(0\).ppt&usg=AFQjCNGLQcLFYSIQ1F-gkcpt_rCO4H6Z7g](https://www.google.co.kr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=5&ved=0ahUKEwiX0buHt9zQAhVCpZQKHdeMCroQFgg1MAQ&url=https%3A%2F%2Fwww.popa.or.kr%2Fupfile%2F2008.%25EA%25B3%25A0%25EA%25B0%259D%25EB%25A7%258C%25EC%25A1%25B1%25EB%258F%2584%25EB%25B3%25B4%25EA%25B3%25A0%25EC%2584%259C_01(0).ppt&usg=AFQjCNGLQcLFYSIQ1F-gkcpt_rCO4H6Z7g)

2012년도 통일공원, 임해자연휴양림, 국민체육센터 고객 만족도 설문조사.

[http://www.google.co.kr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwjfitLUt9zQAhVCKZQKHfEjBz4QFgg1MAA&url=http%3A%2F%2Fcleaneye.go.kr%2Fmodules%2Fdownload.asp%3Ffilename%3D2010000011_2012%25B3%25E2%2520%25B0%25ED%25B0%25B4%25B8%25B8%25C1%25B7%25B5%25B5%2520%25C1%25B6%25BB%25E7%2520%25B0%25E1%25B0%25FA\(%25BF%25E4%25BE%25E0\).pdf%26path%3D%2Fupload%2Fboard_data&usg=AFQjCNF-OOmf2ljPvjek3LIAlk69_o4oEg](http://www.google.co.kr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwjfitLUt9zQAhVCKZQKHfEjBz4QFgg1MAA&url=http%3A%2F%2Fcleaneye.go.kr%2Fmodules%2Fdownload.asp%3Ffilename%3D2010000011_2012%25B3%25E2%2520%25B0%25ED%25B0%25B4%25B8%25B8%25C1%25B7%25B5%25B5%2520%25C1%25B6%25BB%25E7%2520%25B0%25E1%25B0%25FA(%25BF%25E4%25BE%25E0).pdf%26path%3D%2Fupload%2Fboard_data&usg=AFQjCNF-OOmf2ljPvjek3LIAlk69_o4oEg)

2012년도 기상업무 국민 만족도 조사.

http://www.google.co.kr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwjHgorpt9zQAhWEX5QKHcUMB-4QFggdMAA&url=http%3A%2F%2Fweb.kma.go.kr%2Fservlet%2FNeoboardProcess%3Bjsessionid%3DTSG14wWq7VUMUuMnH8DUREDCvYtw6VaYBjEByOWZC6hYYGQUX8UoeBdOfGgbMBE5%3Fmode%3Ddownload%26bid%3Ddepart%26num%3D55%26fno%3D1%26callback%3Dhttp%253A%252F%252Fweb.kma.go.kr%252Fnotify%252Finformation%252Fpublication_department_list.jsp%26ses%3DUSERSESSION%26k%3DATC201301220919221_b853d701-f872-4911-9ea2-acb5ec70a562.pdf&usg=AFQjCNED_rmQW3LkH50pUTX1v4mgOI_IKQ

고위공직자 청렴도 평가 계획.

<https://www.google.co.kr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwjUwY75t9zQAhVDm5QKHVlyBMgQFggfMAA>

&url=https%3A%2F%2Fkostat.go.kr%2Fportal%2Fkorea%2Fkor_ip%2F6%2F1%2Findex.board%3Fbmode%3Ddownload%26bSeq%3D%26aSeq%3D330089%26ord%3D1&usg=AFQjCNEHIOz3UZk3WO4L6UjkXvKFeVXnRA

민원수령 및 환류시스템 구축을 위한 여론조사(輿論調査)실시 결과(관세청).

http://www.google.co.kr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=3&cad=rja&uact=8&ved=0ahUKEwirjKK4uNzQAhXBo5QKHWZKABIQFggpMAI&url=http%3A%2F%2Fwww.prism.go.kr%2Fhomepage%2FresearchCommon%2FdownloadResearchAttachFile.do%3Bjsessionid%3DAFF105BAE DD20A746CE3E11A93A6AF73.node02%3Fwork_key%3D000%26file_type%3DCPV%26seq_no%3D002%26pdf_conv_yn%3DY%26research_id%3D1220000-200700001&usg=AFQjCNGD2GzyhgXomX0_fDhqhcsncbfMBQ

2014년도 서울대학교병원 자체감사 인식도조사

http://www.google.co.kr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja&uact=8&ved=0ahUKEwjb4ZrWuNzQAhULjJKHe1PC9QQFggeMAE&url=http%3A%2F%2Fclean.snuh.org%2Fdn.jsp%3Ffilename%3Dclean_22_1.pdf&usg=AFQjCNGjGSu12MyBIdeC8t9CfZWzc3ciiw

<접수 2016/12/07, 수정 2017/01/18, 게재확정 2017/02/15>