

점수의 불일치는 왜 발생하는가: 가중치 성향과 예비타당성조사 평가

최근혁*

김봉환**

〈目 次〉

- | | |
|---------------------|-----------|
| I. 서론 | IV. 분석 결과 |
| II. 이론적 배경과 선행연구 검토 | V. 결론 |
| III. 연구설계 | |

〈요 약〉

본 연구는 예비타당성조사에서 평가 점수의 불일치 원인을 분석하고, 항목별 가중치 성향이 평가 점수에 미치는 영향을 실증적으로 검토한다. 기존 연구들은 사업의 통과 여부나 제도적 측면에 초점을 맞추는 경우가 많았으나, 점수에 영향을 주는 개별 가중치에 대한 미시적 분석은 상대적으로 부족하였다. 본 연구는 점수 불일치의 대표적 지표인 최고점수와 최저점수를 종속변수로 설정하고 가중치 항목과의 관계를 분석하였다. 분석 결과, 경제성 가중치가 낮고 정책성 가중치가 높을수록 최고점수가 높게 나타났다. 반면에, 경제성 가중치가 높고 정책성 가중치가 낮을수록 최저점수가 높게 나타났다. 본 연구는 이러한 결과를 평가 항목의 성격 차이와 관련된 판단 방식의 차이로 해석할 수 있으며, 행동경제학의 이중처리 이론을 토대로 가중치 기반 패턴에서 추론 가능한 가능성을 제시한다. 이를 통해 예비타당성조사 제도의 평가 과정에서 발생할 수 있는 판단 편차를 이론적으로 설명하고, 평가의 공정성과 일관성 확보를 위한 개선 방향을 제시한다.

【주제어: 예비타당성조사, 가중치 성향, 판단 편차, 행동경제학】

* 주저자, 서울대학교 행정대학원 박사수료(geunhyuk@snu.ac.kr)

** 교신저자, 서울대학교 행정대학원 교수, 한국행정연구소 겸무연구원(kimbong@snu.ac.kr)

논문접수일(2025.11.26), 수정일(2026.1.29), 게재확정일(2026.2.9)

I. 서론

우리나라의 국책사업인 예비타당성조사는 기획재정부가 객관적이고 중립적인 기준에 따라 시행하는 제도로, 총사업비 500억 원 이상의 대규모 국가 예산이 투입되는 사업의 사전 타당성을 검증하는 데 목적이 있다. 이 제도는 예산 낭비와 사업 부실화를 방지하고, 한정된 재정자원의 효율적 분배와 국가 재정운영의 생산성을 제고하기 위해 1999년에 도입되었다. 제도 시행 이후 1999년부터 2015년까지 약 298조 원 규모의 전체 사업비 중 42%의 예산 절감 효과를 거둔 것으로 보고되었으며, 이는 대규모 공공투자사업의 타당성 검증이 국가 정책 결정의 효율성을 실질적으로 향상시켰음을 보여준다. 특히 총사업비 500억 원 이상의 대형 프로젝트의 경우, 예비타당성조사 결과가 사업의 추진 여부를 사실상 결정짓는 핵심 평가 절차로 기능하고 있다(공공투자관리센터, 2015; 2019).

예비타당성조사는 다기준의사결정(Multi-Criteria Decision Making, MCDM) 기법 중 하나인 계층화분석법(AHP: Analytic Hierarchy Process)을 활용하여 사업의 경제성, 정책성, 지역균형발전 등의 항목을 종합적으로 평가한다(기획재정부, 2022). 이 과정에서 각 항목별 가중치가 평가자의 판단에 따라 산출되며, 도출된 AHP 점수는 계량화된 형태로 종합되어 제시된다. 일반적으로 AHP 점수가 0.5 이상이면 사업의 타당성이 인정되어 추진이 가능하다는 기준이 적용된다.

그러나 그동안의 예비타당성조사 연구는 개별 평가 점수보다는 사업 통과 여부와 같은 결과 변수에 초점을 맞추어 왔다. 기존 문헌들은 제도의 제도적 특성이나 사업 면제 요건, 정치적 영향 요인 등 거시적 요인을 중심으로 연구를 진행하였고, 평가 점수가 무엇에 기인하고 있는지에 관한 개인적 판단 차이나 인지적 특성에 기초한 불일치 현상에 대해서는 상대적으로 소홀하였다. 현재 예비타당성조사 사업은 동일한 데이터와 기준을 바탕으로 평가가 이루어지더라도, 개별 점수의 편차가 적지 않게 발생하는 경우가 있다. 이는 자료를 해석하고 판단하는 과정에서 각각의 평가자들이 서로 다른 인지적 처리방식과 가치지향성을 적용하고 있기 때문일 수 있다. 이러한 불일치는 단순한 개인차에 그치지 않고, 국가 정책결정의 공정성과 일관성에 직결되는 중요한 이슈로 확장될 수 있다(Kahneman & Klein, 2009; Fiske & Taylor, 2020). 따라서 이러한 점수 불일치의 원인은 단순한 통계적 변동이 아니라 인지적·행동적 요인에 기인할 가능성이 있으며, 이에 대한 학문적 분석이 요구된다.

본 연구는 이와 같은 문제의식을 바탕으로, 행동경제학적 관점에서 예비타당성조사 평가 점수에 관한 판단행동을 분석하고자 한다. 점수 부여 과정은 단지 객관적 자료 분석의 결과가 아니라, 개인의 인지적 편향(cognitive bias), 가치 판단, 그리고 정보처리 전략 등에 의해

영향을 받을 수 있다(Slovic, 1995; Stanovich & West, 2000; Gigerenzer & Gaissmaier, 2011). 특히, 대니얼 카네만(Daniel Kahneman)의 이중처리이론(Dual Process Theory)은 이러한 평가의 차이를 설명하는 유용한 틀을 제공한다. 즉, 정성적이고 가치적인 판단인 직관적인 방식(시스템 1)에 의존하는지, 아니면 체계적이고 분석적인 방식(시스템 2)에 의존하는지에 따라 동일한 정보라도 다른 해석과 가중치 부여로 다른 평가가 이어질 수 있다.

예비타당성조사에서 점수 산정의 핵심은 각 항목에 얼마만큼의 가중치를 부여하느냐에 달려있다. 따라서 평가가 이루어질 때, 인지적 처리유형이 가중치 부여 경향에 어떠한 방식으로 작용하는지를 분석하는 것은 예비타당성조사 제도의 공정성과 신뢰도를 제고하는 데 매우 중요한 시사점을 제공할 것이다. 본 연구는 행동경제학의 이중처리 이론을 토대로, 가중치 성향과 판단 방식이 실제 부여된 점수 간의 불일치에 어떤 영향을 미치는지를 실증적으로 검증한다.

그동안 예비타당성조사 관련 연구는 제도의 효율성이나 절차적 측면에서는 상당한 발전을 이루었으나, 개별 평가 결과와 가중치 항목에 대한 실증 연구는 거의 부재하였다. 이러한 한계는 제도적 보안성과 평가자 익명성이라는 구조적 제약에서 기인한다. 따라서 본 연구는 제한된 데이터를 활용하되, 행동경제학의 관점을 접목하여 평가자의 인지처리유형, 가중치 부여 성향, 그리고 점수 불일치 현상 간의 관계를 실증적으로 분석함으로써, 기존 연구가 다루지 못했던 평가자의 인지적 결정구조를 추론하는 데 의의가 있다. 다만, 개별 평가자의 배경 및 특성을 확인할 수 없는 한계에 따라 본 연구의 목적은 인지 작동을 검증하는 연구가 아니라 가중치 기반 패턴에서 추론 가능한 가능성을 제시하는 연구임을 명시한다. 이를 통해 본 연구는 예비타당성조사 평가 과정에서의 판단 편차를 인지적 관점에서 설명함으로써 제도의 한계를 보완하고, 궁극적으로 공정하고 일관된 평가체계 구축을 위한 정책적 개선 방향을 제시하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 예비타당성조사 제도와 관련한 이론적 배경 및 선행연구를 검토하고, 3장에서 이를 토대로 연구가설을 설정한다. 4장에서는 연구설계와 분석방법을 제시하며, 5장에서 실증분석 결과를 보고한다. 마지막으로 6장에서는 본 연구의 주요 결과를 요약하고 정책적 함의 및 향후 연구 과제를 제시한다.

II. 이론적 배경 및 선행연구 검토

1. 예비타당성조사제도

예비타당성조사(Pre-feasibility Study)는 대규모 공공투자사업의 타당성을 사전에 검토하여 사업의 효율성과 필요성을 검증하기 위한 제도로써, 국가 재정운영의 건전성과 효율성 확보를 목적으로 한다(김봉환 외, 2024). 1999년 도입된 이래, 예비타당성조사는 예산 낭비를 방지하고 사업의 부실화를 예방하는 국가재정관리의 핵심적 절차로 자리 잡았다. 제도의 시행 이후 다수의 사업에서 불필요한 예산 투입이 사전에 차단되었으며, 결과적으로 공공투자사업의 실행 타당성 판단에 과학적 근거를 제공하고 있다(김강수, 2016).

이와 같은 중요성으로 인해 예비타당성조사 제도의 제도적 효과(김태운·조예진, 2018), 면제 및 통과 여부(박민정, 2020), 그리고 정치적 영향 요인(김봉환·최근혁, 2022)에 관한 다양한 연구가 축적되어 왔다. 과거에는 사업의 ‘통과 여부’ 또는 ‘면제 여부’ 그 자체가 주요한 정책 논의의 초점이었다. 그러나 2019년 제도 개편 이후에는 사업의 통과율이 전반적으로 상승하면서, 쟁점이 미시적 관점에서의 ‘점수 산정의 과정과 평가자의 판단’으로 이동하는 추세를 나타내고 있다. 실제로 제도 개편 이후 6년간 예비타당성조사 사업의 통과율은 수도권 78%, 비수도권 88%로 매우 높게 나타났으며(정동호, 2024; 최근혁·김봉환, 2025), 이에 따라 단순한 통과 여부의 결과보다 평가 점수의 구체적 구성과 평가자 간 점수 차이에 관한 연구 질문이 떠오르고 있다.

예비타당성조사의 평가 체계는 점차 정교해졌다. 제도 도입 초기인 2000년에는 연구진 3명과 관리기관 1명 등 4명의 평가자로 구성되었으나, 2005년에는 연구진 5명, 관리기관 2명, 외부검토위원 1명으로 확대되었고, 2017년에는 외부 전문가 2인을 추가하여 총 10인 체계를 구축하였다. 이후 2019년 제도 개편에서는 외부 전문가 중심의 구조로 전환되어 연구진 1명, 재정사업평가 분과위원 7명으로 구성되었으며, 2022년 12월 기준으로는 연구진 3명, 분과위원 9명 등 총 12인 체계로 운영되고 있다(정동호, 2024). 이들은 각 부처가 수립한 사업계획의 타당성을 검토하고, 사업의 추진 시기 및 규모를 조정하며, 합리적 의사결정을 위한 근거를 제공한다(기획재정부, 2019).

평가 방법론은 경제성 분석, 정책성 분석, 지역균형발전 분석 등 세 가지 항목으로 구분된다. 첫째, 경제성 분석은 비용-편익분석(Cost-Benefit Analysis)에 기반을 두며, 예상 사업 수요를 추정하여 편익을 산정하고, 총사업비와 운영 경비 등을 포함한 총비용과 비교함으로써 경제적 효율성을 평가한다. 둘째, 정책성 분석은 정책목표 부합성, 사회적 파급효과,

재원 조달 가능성 등을 정량적·정성적으로 평가하며, 필요할 경우 문화재 가치나 재원 조달의 위험 등 사업별 특성을 반영한 별도 항목이 추가된다. 셋째, 지역균형발전 분석은 사업이 지역 간 격차를 완화하고 지역 발전에 미치는 영향을 평가하기 위한 것으로, 지역낙후도 및 지역경제 파급효과 등 다면적 요소를 고려한다. 특히 지역낙후도는 1점에서 9점 범위 내에서 평가되며, 사업의 지역적 형평성을 검증하는 기준으로 활용된다.

이와 같은 단일 항목별 분석이 모두 완료된 후에는 계층분석법(AHP: Analytic Hierarchy Process)을 활용해 세 항목의 평가 결과를 종합하고, 그 상대적 중요도를 수치화하여 최종 AHP 점수를 산출한다. 일반적으로 AHP 점수가 0.5 이상이면 사업의 타당성이 인정되며, 이는 사업추진 여부 결정의 핵심 지표로 활용되고 있다(기획재정부, 2022).

그러나 동일한 사업을 평가하더라도 각 평가자가 부여하는 가중치의 차이에 따라 최종 AHP 점수가 달라질 수 있다. 다시 말해, 어떤 항목에 더 큰 가중치를 두느냐가 평가 결과를 결정짓는 주요 요인으로 작용할 수 있다는 것이다. 본 연구는 바로 이 점에 주목하고자 한다. 즉, 동일한 정보를 제공받더라도 평가자마다 정책성·경제성 항목에 부여하는 가중치 성향이 상이하며, 이로 인해 점수의 불일치가 발생할 수 있다. 본 연구는 이러한 불일치 현상을 행동경제학의 이중처리이론(System 1과 System 2)을 기반으로 해석함으로써, 평가자의 인지전략이 실제 점수 산정 과정에 반영되고 있음을 실증적으로 확인하고자 한다.

2. 평가자 행태에 관한 연구 동향

앞서 논의한 것과 같이 예비타당성조사 제도는 우리나라 대규모 국책사업의 예산 타당성을 사전에 검증하는 제도로서, 정책결정의 핵심 도구로 기능하고 있다(고길곤, 2007). 그러나 제도적 중요성에도 불구하고, 그동안 예비타당성조사 평가자의 행태나 판단 특성에 대한 연구는 거의 이루어지지 않았다. 기존 연구는 대부분 제도의 효율성(엄근용, 2019; 박승준, 2022), 정치·제도적 영향요인(김봉환·최근혁, 2022; 정동호, 2024) 등에 초점을 두며 거시적 접근에 머물렀다. 외부연구진을 통해 간접적으로 평가자와 점수의 관계를 파악하려는 시도가 있었지만(김정숙 외, 2017) 제한적이었다.

이러한 현상의 근본적인 이유는 우리나라 예비타당성조사의 제도적 특수성에 기인한다. OECD 국가들 중에서 중앙정부가 모든 대규모 국책사업의 예비타당성조사를 직접 관리하는 구조를 유지하는 곳은 대한민국이 사실상 유일하다(김태일, 2019). 이와 같은 강력한 중앙집중형 제도는 재정의 효율성을 확보하는 데에는 유리하지만, 동시에 이해관계가 복잡하게 얽히는 구조를 형성한다. 대형 사업의 평가 결과는 수천억 원 규모의 예산 배분과 직결되기 때문에, 평가자를 둘러싼 지대추구 행위 및 정치적 외압 가능성이 지속적으로 제기되어

왔다(신가희·하연섭, 2015; 박민정, 2020).

이러한 문제를 방지하기 위해 우리나라의 예비타당성조사 제도는 ‘블라인드(Blind) 평가 체계’를 채택하고 있다. 즉, 전공, 학력, 소속 기관, 출신 지역, 경력 등 평가자의 신상에 관한 정보는 공식적으로 비공개이며, 담당 부처인 기획재정부만 해당 정보를 보유하고 있다. 평가자들 간의 상호 확인도 제한되어 있으며, 보고서상의 익명처리 원칙이 엄격히 유지된다. 이러한 제도적 특성으로 인해 예비타당성조사 평가자 관련 자료를 획득하기가 구조적으로 불가능에 가깝다.

그 결과, 학계에서는 평가자의 특성을 분석하거나 인지적 판단 메커니즘을 탐색하기 위한 자료 접근이 현실적으로 어려웠다. 연구자는 평가자가 누구인지조차 알 수 없으며, 평가자의 전문분야, 가치관, 분석 성향 등 미시적 요인을 변수화하기 어렵다. 따라서 기존 연구들은 예비타당성조사에서 발생하는 점수의 불일치나 판단 편차를 ‘제도적 문제’ 혹은 ‘가중치 설정 방식의 기술적 한계’로만 해석하였으며, 평가자의 사고 구조나 인지 과정을 탐구하지 못했다.

그러나 평가 점수는 결국 개인적 판단의 집합적 표현이다. 동일한 데이터와 지침을 제공받더라도, 평가자 간 점수가 크게 다르게 나오는 이유는 정보 해석, 위험 인식, 가치 판단 등 인지적 차이(cognitive variation)에 의해 발생할 가능성이 크다. Kahneman(2011)과 Stanovich & West(2000)는 인간의 판단과 행동이 직관적 체계(System 1)와 분석적 체계(System 2)의 상호 작용에 의해 결정된다고 보았으며, 이러한 인지 처리 방식의 차이가 일관성 있는 판단을 어렵게 한다고 설명하였다. 이러한 이론은 예비타당성조사 평가자 간 점수 불일치를 이해하는 강력한 분석 틀을 제공한다.

특히 본 연구는 데이터의 한계를 인정하면서도, 공개된 예비타당성조사 자료를 활용하여 평가자들의 점수 데이터를 분석하는 새로운 접근을 시도한다. 각 평가자별 점수 중 ‘최고점과 최저점’을 중심으로 경제성과 정책성 항목 간 가중치 부여 성향을 분석하고, 이를 행동경제학의 이중처리이론(Dual-Process Theory) 관점에서 해석한다. 이를 통해 평가자의 인지 전략(cognitive strategy)이 실제 평가 결과에 반영되고 있음을 실증적으로 검증하고자 한다.

3. 이중처리 이론

본 연구는 대니얼 카네만(Daniel Kahneman)의 이중처리 이론(Dual-Process Theory)에 주목한다. 이 이론은 인간의 사고 체계를 두 가지 시스템, 즉 직관적인 사고인 시스템 1(System 1)과 분석적이고 논리적인 사고인 시스템 2(System 2)로 구분한다. 각각의 시스템은 의사결정의 과정에서 상이한 방식으로 작동하며, 상황과 개인의 성향에 따라 어느 한 시

시스템이 우세하게 작동하게 된다.

이중처리 이론은 인간의 판단과 선택이 상이한 두 가지 정보처리 체계, 즉 직관적 처리와 분석적 처리의 상호작용으로 이루어진다는 설명틀이다. 일반적으로 시스템 1은 경험, 연상, 감정, 패턴 인식과 같은 자동적 처리 방식에 기반하여 판단을 형성하는 경향이 있다. 이러한 판단은 필연적으로 직관적 요소를 포함하지만, 이는 단순한 즉흥적 반응을 의미하는 것이 아니라 개인이 축적해 온 가치관·규범 인식·경험적 판단 틀 등이 정성적으로 활성화되는 것을 포괄한다(Kahneman, 2003; Stanovich & West, 2007). 반면 시스템 2는 계산, 규칙 적용, 정보 검토 등 노력이 필요한 인지작용을 통해 판단을 교정하거나 분석적으로 정교화하는 역할을 한다.

두 체계의 특징은 처리 방식, 의식적 통제 가능성, 감정 의존성에서 뚜렷이 구분된다. 시스템 1은 가치·규범 판단이 자동적이고 확신 기반으로 작동하는 영역으로써, 정성적 요소가 개입되어 직관적 판단이 강화되는 기능을 한다. 특히 위험·공공정책 분야에서도 감정적 직관에 기반하여 판단을 내린다는 연구가 존재하고 있으며 가치에 기반한 직관적 판단을 일으키는 대표적인 요소이다(Tversky & Kahneman, 1974; Kahneman & Tversky, 1984; Slovic, 1995). 반면 시스템 2는 계산, 규칙 적용, 정보 검토 등 노력이 필요한 인지작용을 통해 판단을 교정하거나 정교화하는 역할을 한다. 구체적으로는 비용-편익 분석, B/C, NPV 등 경제적 판단을 요구하는 특성이며 정량적 분석을 필요로 하는 판단 요소이다(Kahneman, 2003; Weber & Johnson, 2009).

중요한 점은 어떤 체계가 우세해지는지가 맥락 요인에 좌우된다는 것이다. 시간 압박, 정보의 모호성, 정성적 함의, 사회적 규범의 소환 등은 시스템 1의 영향력을 강화한다. 반대로 높은 책무성(accountability), 명확한 평가기준, 계산가능한 정보구조, 외부검증 가능성 등은 시스템 2의 개입을 촉진한다(Brehmer, 1994; Keren & Teigen, 2004). 즉, 동일한 대상이라도 과업 설계·정보 제시 방식·평가 환경에 따라 두 체계의 가중치는 달라지고, 그 결과 판단값이 달라질 수 있다.

다기준 평가 맥락에 적용하면, 시스템 1은 가치지향·규범적 고려(공익성, 형평·분배, 전략적 시의성 등)를 직관적으로 전면화하고, 시스템 2는 화폐화 가능한 편익·비용과 방법론적 타당성(자료 신뢰도, 모수·모형의 적합성 등)을 체계적으로 점검한다. AHP와 같이 가중치를 통해 최종 점수를 산출하는 제도에서 두 체계의 상대적 개입 정도는 곧 ‘어떤 항목에 더 큰 가중을 부여하는가’로 관찰될 수 있다. 따라서 가중치 성향의 체계적 차이는 이중처리의 심리적 기반을 반영하는 표지로 해석될 여지가 있다.

본 연구는 이러한 이론적 논의에 기반하여, 예비타당성조사에서의 가중치 부여 성향이

판단 방식의 차이를 부분적으로 반영할 가능성에 주목한다. 즉, 특정 항목이 요구하는 정보 처리 방식의 특성이 평가자 간 가중치 차이에 일정 부분 나타날 수 있으며, 이는 점수의 불일치 현상을 이해하는 하나의 관찰 지점이 될 수 있다. 다만 본 연구는 평가자의 인지과정을 직접 측정하는 연구가 아니므로, 이중처리 이론을 가중치·점수 관계의 '설명적 가능성'을 제시하는 틀로 활용하며, 이를 인과적으로 단정하지 않는다는 점을 명확히 한다.

4. 연구가설의 설정

점수 불일치는 평가자 개인의 판단 방식과 가치 성향에 기인할 수 있다. 예비타당성조사에서는 평가자가 항목별 가중치를 해당 범위 안에서 자율적으로 설정하고 이는 평가자가 중요하게 여기는 판단 기준을 반영하는 지표로 해석될 수 있다. 즉, 각 평가자의 사고방식이나 정보 처리방식에 따라 동일한 정보를 받아들여도 상이한 판단이 이루어질 수 있다는 것을 의미한다.

이러한 현상을 이해하기 위해 본 연구는 행동경제학의 이중처리 이론이 제시하는 논의를 하나의 해석틀로 활용한다. 이중처리 이론은 판단 과정에서 정성적·경험적 판단과 분석적·속고적 판단이 상호작용한다고 설명하며, 정보 유형이나 판단 환경에 따라 두 처리 방식의 비중이 달라질 가능성을 제시한다. 다만 본 연구는 평가자의 인지 체계를 직접 측정할 수 없으므로, 이 이론을 가중치와 점수 패턴과의 관계를 해석하는 보조적 설명틀로 연구에 사용하였다. 직접적 인과 설명이 아닌 점수 및 가중치 부여 패턴을 이해하는 틀로 활용하면서, 항목별 정보 특성이 평가자의 점수 산정 경향성과 어떤 관련성을 갖는지를 실증적으로 검증하고자 한다. 즉, 경제성과 정책성 항목의 내용적 특성이 최고점수와 최저점수에 서로 다른 방향의 영향을 미치는지를 확인하는 것이 본 연구의 핵심 목적이다.

본 연구는 예비타당성조사의 첫 번째 항목인 경제성 항목이 시스템 2의 작동 맥락과 밀접하게 연관되어 있다고 본다. 경제성 항목은 비용-편익비(B/C), 순현재가치(NPV), 내생적 수요 추정 등 정량적 근거를 중심으로 사업의 효율성을 검토하는 영역이다. 이러한 항목은 계산 가능성과 객관적 자료에 근거하기 때문에 평가자의 분석적·계산적 사고를 요구하며, 이는 시스템 2의 작동을 촉발한다(Slovic, 2000; Reyna, 2004). 평가자는 정량적 기준을 통해 사업의 타당성을 검토하며, 수익성 부족이나 낮은 편익 추정에 보다 민감하게 반응하게 된다.

이와 같은 시스템 2 기반의 평가는 보수적 성향을 강화하며, 평균보다 낮은 점수를 산출할 가능성이 크다. 전망이론(prospect theory)에서 보여주듯, 손실 회피적 사고는 낮은 B/C 비율이나 재무적 불확실성을 더욱 부정적으로 평가하게 만든다(Hodgkinson et al, 2008; Weber & Johnson, 2009). 따라서 경제성에 높은 가중치를 두는 평가는 효율성과 시장 논리

에 기반한 보수적 판단으로 이어지며, 결과적으로 점수의 하향 조정으로 귀결될 가능성이 높다(Fischhoff et al, 1978 ; Frank et al, 1993).

가설 1. 경제성 항목은 정량적 분석 비중이 높기 때문에, 경제성에 더 큰 비중을 두는 평가자는 보수적인 접근에서 낮은 점수를 부여할 가능성이 있다.

두 번째 항목인 정책성 항목은 시스템 1의 작동 맥락과 연관되어 있다고 본다. 예비타당성조사는 사업의 효율적 운영 및 적절성을 분석하고 평가하는 역할을 하지만, 분석적 환경 속에서도 가치·규범 판단의 개입 가능성은 존재하며 이 부분이 정책성 항목에서 상대적으로 크다고 볼 수 있다. 정책성은 사회적 필요성, 정부 정책 방향과의 일치성, 생활 여건 개선 등 정량화하기 어려운 요소를 포함한다. 이러한 요소들은 평가자의 가치지향성이나 직관적 판단에 크게 영향을 받을 수 있으며, 이는 시스템 1이 강조하는 정성적·가치 기반 판단과 밀접한 관련이 있다.

따라서 정책성 항목은 평가자의 직관과 신념, 사회적 가치 지향성이 더욱 강하게 작동하는 영역이다. 기존 연구에 따르면, 정성적 직관은 대체로 긍정적 평가를 강화하고 관대화 편향(leniency bias)으로 이어질 가능성이 크다(Hogarth & Einhorn, 1992; Mitchell et al, 1993; Sunstein, 2000). 이 때문에 시스템 1에 기인하는 평가는 정책적 이상이나 공공재로서의 의미를 크게 강조하며, 이는 상대적으로 높은 점수 부여로 연결될 수 있다.

가설 2. 정책성 항목은 가치·정성적 판단 요소가 크기 때문에, 정책성을 더 중시하는 평가자는 높은 점수를 부여할 가능성이 있다.

앞선 두 가설에서 제시된 바와 같이, 평가자의 점수 부여는 인지처리체계의 작동 방식에 따라 구조적으로 달라질 수 있다. 그러나 실제 평가 과정에서는 한 시스템에 과도하게 의존하는 편향된 인지전략(cognitive bias)이 관찰될 수 있으며, 이러한 경향은 점수의 ‘극단화(score extremity)’로 이어질 가능성이 있다.

시스템 1은 감정적 확신과 직관적 판단에 근거해 신념 일관성을 강화하려는 경향이 있으며, 이는 긍정적 혹은 부정적 판단의 강도를 확대시키는 효과를 낳는다. 한편 시스템 2 역시 분석적 정합성을 확보하려는 과정에서 과도한 논리적 확신(logical certainty)의 함정에 빠지며, 복잡한 상황을 단일 값으로 단정하려는 과잉 확신(overconfidence)을 보인다(Simon, 1972; Mackie & Preston, 1998; Nickerson, 1998; Gilovich et al. 2002)

이처럼 어느 한 체계에 대한 의존성이 커질수록, 판단은 중간적 균형점을 벗어나 더 극단

적인 결과를 초래한다. 평가자가 경제성과 같은 정량적 항목에 모든 판단을 집중할 경우, 효율성을 중시하는 시스템 2의 영향이 강화되어 낮은 점수가 과도하게 산출될 수 있다. 반대로 정책성과 같은 정성적 항목에 치우칠 경우, 가치적 정당성을 강조하는 시스템 1의 직관이 강화되어 높은 점수를 부여할 가능성이 높아진다.

이는 행동경제학의 “인지적 일관성 편향(cognitive consistency bias)” 및 “확증편향(confirmation bias)” 이론과도 맞닿아 있다. Kunda(1990)는 사람들이 자신이 중시하는 신념과 일관된 해석을 선호하며, 이를 통해 심리적 안정감을 확보한다고 설명하였다. 예비타당성조사 평가자 역시 경제성 혹은 정책성 중 하나의 판단 틀을 기준으로 모든 정보를 해석함으로써, 복합적 판단 대신 단일화된 방향으로 점수를 부여할 수 있다.

가설 3. 가중치가 한쪽으로 많이 치우칠수록 점수가 더 극단적으로 부여될 것이다.

아래 <그림 1>은 앞서 논의한 가설을 토대로 평가자의 가중치 성향과 점수의 관계를 나타낸 내용이다.

<그림 1> 가중치 성향과 점수의 관계

	최고점수	최저점수
경제 가중치	-	+
정책 가중치	+	-

III. 연구설계

1. 자료, 변수 및 기술통계량

본 연구는 예비타당성조사에서 항목별 가중치 성향이 평가자의 점수에 어떠한 영향을 미치는지 확인하는 것을 목표로 한다. 분석을 위하여 수집이 가능한 공개데이터 중에서 예비타당성조사 수행기관인 KDI의 공공투자관리센터 홈페이지에 게재된 각 사업의 예비타당성조사 보고서를 이용하였다. 자료는 2013년부터 2024년까지 총 12년의 사업을 분석단위로 활용하였으며 242개의 사업이다.

연구에 활용한 종속변수는 예비타당성조사 결과에 실질적으로 영향을 미친 8인의 평가자 점수 중 최대값(max)과 최소값(min)을 종속변수로 활용하였다.¹⁾ 평균이나 중앙값은 개별 평가자의 판단 편차를 통계적으로 흡수·희석시켜, 관대한 평가자와 엄격한 평가자 간의 실질적 차이를 충분히 드러내기 어렵다. 반면, 최고점과 최저점은 집단 내 판단 분포의 양극단을 포착함으로써 평가자의 인지적 극단성을 정밀하게 반영한다. 이는 전문가 집단 내 판단 편차를 연구하는 데 적합한 접근으로, 의사결정 연구에서 관찰되는 “극단값 기반 행동분석(extreme value-based behavior analysis)”의 논리와 같은 맥락을 지니고 있다(Wang et al, 2024).

또한, 최고점과 최저점은 평가자의 전략적 행동과 인지적 편향이 나타나는 지점을 가장 민감하게 반영하는 지표로 볼 수 있다. 예비타당성조사와 같이 복잡한 요인과 전문적 지식을 고려하면서 평가하는 상황에서는, 평가자들이 정책적 가치나 재정 효율성에 대한 자기신념과 확신에 따른 차별적 평가를 수행할 가능성이 높다. 이러한 비대칭적 판단은 평균값에는 거의 영향을 남기지 않지만, 극단값에서는 명확히 드러난다. 실제로 Slovic(1995)은 전문가 판단의 다양성을 분석하는 과정에서 '평균적 경향보다 극단 점수에 집중하는 것이 가치기반 판단을 탐지하는 데 더 유효하다'고 지적하였다. 따라서 본 연구는 최고점·최저점

1) 각 사업의 평가자 점수 중 최댓값(max)과 최솟값(min)을 종속변수로 활용하였다. 단순한 이상치(outlier)가 아닌 정책적 의미를 지닌 점수인지 확인하기 위해, 본 연구는 각 사업별 평균점수(8인 기준)로부터 최댓값과 최솟값이 떨어진 정도를 평균 수렴 지수(Mean Pull Index, MPI)로 측정하였다:

$$MPI_i = \frac{|S_{\max} - \bar{S}| - |S_{\min} - \bar{S}|}{|S_{\max} - \bar{S}| + |S_{\min} - \bar{S}|}$$

MPI가 0이면 평균이 두 극단값의 중간에 위치함을, 음수이면 평균이 최고점에 더 가깝고 양수이면 최저점에 더 가깝다는 뜻이다. 분석 결과, 평균은 다소 최고점 쪽으로 치우치는 경향이 있었으나 전반적으로 대칭적 분포를 보여, 특정 평가자의 극단값이 평균을 왜곡한다는 귀무가설을 기각하였다(회귀계수: -.4296, $p < .001$).

변수의 활용을 통해 행동경제학적 편향과 가중치 성향의 불균형을 보다 명확히 포착하고자 한다.

본 연구의 첫 번째 독립변수는 각 사업별로 가장 높은 점수를 준 평가자와 가장 낮은 점수를 준 평가자가 설정한 항목별 가중치이다. 점수 불일치가 발생한 경우, 그 극단에 위치한 평가자들이 어떠한 판단 기준을 중시했는지를 확인함으로써, 점수 불일치의 구조적 원인을 파악하고자 한다. 특히 평가자 간 가치관이 가장 뚜렷하게 갈리는 항목인 경제성, 정책성 항목에 주목하였다.²⁾ 경제성 가중치는 비용 대비 편익을 중심으로 한 정량적 판단 기준에 대한 평가자의 선호를 반영한다. 높은 경제성 가중치는 수익성, 효율성, B/C 비율 등 계량적 성과를 중시하는 성향을 나타내며, 이는 보수적·위험회피적 평가 태도와 연관될 수 있다(Kahneman & Tversky, 1984; Gigerenzer & Gaissmaier, 2011). 반대로 낮은 경제성 가중치는 경제성 평가를 상대적으로 경시하거나, 다른 정성적 항목을 더 중시하는 경향을 시사한다. 정책성 가중치는 공공성, 분배 정의, 사회적 필요성 등 정성적 기준에 대한 평가자의 가치관을 반영한다. 정책성에 높은 가중치를 둔 평가는 사업의 사회적 의미나 정책 목표의 정당성을 강조하며, 보다 관대한 점수 부여로 이어질 가능성이 있다. 지역균형발전 가중치는 사업이 지역 간 격차 해소에 기여하는 정도를 중시하는 평가자의 태도를 나타내며, 특히 비수도권 또는 낙후 지역을 고려하는 정책적 감수성과 연결된다.

본 연구의 두 번째 독립변수는 가중치 편향에 관한 변수이다. 크게 2개의 변수로 나뉘어 살펴봤으며 가중치 차이는 각 경제성 가중치와 정책성 가중치의 차이를 절대값으로 측정한 변수이다. 가중치 비율은 정책 가중치를 경제 가중치로 나눈 값으로 가중치의 불균형이 평가자의 극단성을 높이는지 확인할 수 있는 변수로 설정하였다.

또한, 다양한 요인들을 통제하여 종속변수와 독립변수 간의 관계를 더 정확하게 추정하고자 하였다. 우선 각 사업의 규모(log) 변수를 통제하여 금액에 따른 편동을 줄였고 사업의 주요 지표로 볼 수 있는 종합평가 점수인 ahp 점수를 통제함으로써 분석결과와 정확도를 높이고자 하였다. 정치적 요인으로 볼 수 있는 여·야 국회의원의 수를 통제하였고 지역의 수도 통제하였다. 시·도 광역단체를 기준으로 예비타당성조사 사업이 복수의 지역에 걸쳐서 시행되는 지역의 수가 몇 개인지 파악하는 것이 중요하며 이것을 통제해 연구를 진행하였다. 다음으로는 분석단위가 사업인 것을 고려하여 사업에 영향을 주는 요인들을 고정효과로 통제하였다. 첫 번째는 연도를 통제하여 각 연도별 차이에 따른 효과를 통제하였고 사업 부문 변수도 활용하였다. 사업 부문 변수는 김봉환·최근혁(2022) 논문에서의 기준을 토대로 사

2) 지역균형발전 항목은 이중처리이론을 활용한 본 연구의 실증분석을 수행하기 위해 제외하였다. 다만, 실증분석에서 비수도권 여부를 고정효과로 통제하여 지역균형발전 항목으로 인해 발생할 수 있는 한계를 해소하였다.

업의 부문을 12개의 분야로 나누어 분석을 진행하였다.³⁾ 마지막으로는 비수도권 여부도 통제변수로 활용하였다. 수도권 지역인 서울, 경기, 인천을 포함하지 않은 사업의 경우 1로 설정하였고 그 외의 사업을 0으로 설정하여 비수도권 더미변수를 사용하였다. 다만, 수도권 중 접경·도서, 농산어촌 지역(경기도 김포시, 파주시, 양주시, 동두천시, 포천시, 연천군, 가평군, 양평군, 인천광역시 강화군, 옹진군)은 비수도권으로 분류하였다(기획재정부, 2019). 아래 <표 1>은 가설검증을 위한 변수의 정의를 정리한 내용이다.

<표 1> 가설검증을 위한 변수의 정의

구분	변수명	조작적 정의	비고
종속 변수	최고점수	각 사업 점수의 최고점과 최저점	
	최저점수		
독립 변수	경제 가중치	각 사업의 경제, 정책 가중치 점수	
	정책 가중치		
	가중치 차이	$\left\{ \begin{array}{l} \text{경제 가중치} - \text{정책 가중치} \\ \text{경제 가중치} \\ \text{경제 가중치} \end{array} \right.$	
	가중치 비율		
통제 변수	사업 규모	예타 사업 예산의 로그값	
	ahp	예타 사업의 최종점수	
	여당 의원 수	각 지역 연도별 여·야 소속 국회의원 수	19-22대 국회의원 선거 기준
	야당 의원 수		
	지역 수	예타 사업이 진행되는 시·도 광역단체 지역의 수	
	사업 부문	예타 사업의 부문별 변수	정부의 예산안 개요 및 국가재정운용계획에서 국가재정을 12대 분야로 나눈 것을 기준
	연도	예타 사업이 기록된 연도	KDI 공공투자관리센터에 보고된 일자 기준
	비수도권 더미	예타 사업 중에서 수도권 지역에 해당하지 않은 사업은 1, 나머지는 0	수도권 지역 (서울, 경기, 인천)

<표 2>는 본 논문에서 활용된 주요 변수들의 기술통계량을 보여준다. 본 연구는 총 242개

- 3) 총 12가지의 사업 부문으로 나누어서 분석을 진행하였으며 분류 기준은 다음과 같다. 1. SOC(철도, 도로, 공항, 항만, 상하수도, 방파제), 2. 교육, 3. 복지(보건, 복지, 고용), 4. 문화(문화, 체육, 관광), 5. 국방, 6. 안전, 7. R&D(과학, 기술, 개발), 8. 외교(외교, 통일, 국제통상), 9. 환경(환경, 에너지, 신재생), 10. 행정(디지털-정보, 통신, 스마트, 빅데이터, 네트워크), 11. 산업(산업, 산업단지, 클러스터), 12. 건축(신·증축, 시설, 이전, 리모델링, 건립)

사업 사례를 분석 대상으로 하였으며, 각 사업은 8인의 평가자가 부여한 AHP 점수를 바탕으로 구성되었다. 주요 변수에 관한 기술통계량을 중점적으로 설명하면, 종속변수인 최고점수와 최저점수, 그리고 독립변수로서 경제적·정책적 가중치, 통제변수로서 사업 규모·사업 부문·연도 등이 있다.

먼저, 종속변수인 최고점수와 최저점수의 평균은 각각 0.588과 0.486으로 확인되었다. 이는 평균적으로 평가자 간 평균 점수의 차이가 약 0.1 이상 존재하고 있으며 일부 사업에서는 평가자 간 판단이 크게 엇갈렸음을 보여준다. 최고점수의 표준편차는 0.116, 최저점수의 표준편차는 0.113으로, 비슷한 수준으로 나타나고 있다. 첫 번째 독립변수로 활용된 경제 가중치와 정책 가중치는 최고점수에서는 평균값이 0.424, 0.347로 나타나고 있고 최저점수에서도 0.457, 0.316으로 비슷한 수준을 확인할 수 있다. 표준편차와 최소값, 최대값에서도 크게 차이가 발생하고 있지 않다는 것을 알 수 있다. 두 번째 독립변수인 가중치 차이와 가중치 비율은 평균값이 0.881, 0.725로 나타내고 있다. 주목해서 볼 점으로는 절대값인 가중치 차이 변수를 확인할 수 있다. 평균값이 0.881로 전반적인 불일치 수준이 높지 않고 대부분의 관측치에서 차이가 1점 미만으로 나타나고 있다. 하지만, 최소값이 0.4이고 최대값이 4라는 점은 점수 차이가 상황에 따라 크게 변동될 수 있음을 의미한다. 이는 평균적으로는 작은 편인 차이를 보이지만, 특정 조건이나 상황에서는 두 값의 판단이 크게 갈릴 수 있음을 시사한다. 총사업비에 해당하는 사업 규모의 평균은 약 6786.4억 원이며, 표준편차가 3910.3으로 나타나 사업 간 규모 차이가 매우 크다는 점을 알 수 있다. 사업부문 변수는 부문이 총 1~11로 나타나 분석 대상이 특정 부문에 치우치지 않고 다양한 분야에 걸쳐 있음을 보여준다. 사업이 이루어지고 있음을 보여주는 지역 수 변수는 최대값 9, 최소값 1로 차이가 있지만 평균값이 1.243에 가까운 것으로 나타나고 있어 대다수 사업이 시·도 1개의 지역에서 진행되고 있음을 알 수 있다. 연도 변수는 2011년부터 2022년까지의 자료를 포함하고 있으며, 표본 내에서는 2019년 이후 데이터가 다수를 차지하였다. 또한 비수도권 더미변수의 평균값은 0.67 수준으로, 약 2/3의 사업이 비수도권 지역에서 수행되었음을 나타낸다.

〈표 2〉 기술통계량

변수	표본크기	평균	표준편차	최소값	최대값
최고점수	242	0.588	0.116	0.265	0.851
경제 가중치	242	0.424	0.094	0.200	0.700
정책 가중치	242	0.347	0.084	0.200	0.800
최저점수	242	0.486	0.113	0.183	0.737
경제 가중치	242	0.457	0.087	0.300	0.700

정책 가중치	242	0.316	0.070	0.200	0.700
가중치 차이	242	0.881	0.416	0.400	4.000
가중치 비율	242	0.725	0.271	0.400	2.333
사업 규모	242	6786.397	14715.42	480	124,284
ahp	242	0.539	0.114	0.226	0.814
여당의원 수	242	31.247	46.191	0	180
야당의원 수	242	23.566	36.415	0	127
지역수	242	1.243	0.741	1	9
사업 부문	242	4.024	3.125	1	11
연도	242	2018.124	3.272	2013	2024
비수도권 더미	242	0.665289	0.472867	0	1

아래의 <표 3>은 분석에서 활용된 예비타당성조사 부문별 규모를 나타낸 자료이다. 예비타당성조사 사업 중에서 가장 많은 부문을 차지하고 있는 것은 SOC 부문이다. SOC 부문은 전체 242건 중 149건으로 전체의 약 61.6%를 차지하였으며, 평균 사업비는 0.843조 원으로 확인되었다. 최소 사업비는 0.051조 원, 최대 사업비는 12.126조 원으로 나타나, 규모의 변동 폭이 다른 부문에 비해 크게 나타났다. 건축, 교육, 문화 부문은 각각 30건, 20건, 18건으로 파악되었으며, 평균 사업비는 0.675조 원, 0.524조 원, 0.586조 원으로 확인되었다. 교육 부문은 최소 0.213조 원에서 최대 11.907조 원까지 폭넓은 규모 분포를 보였으며, 문화 부문은 평균 사업비 수준이 유사하나 편차가 상대적으로 커, 사업의 특성에 따라 예산 규모가 크게 달라지는 경향을 예상할 수 있다. 국방과 환경 부문은 사업 수가 상대적으로 적으나, 평균 사업 규모가 크고 집중도가 높게 나타났다. 국방 부문은 9건이고 평균 사업비는 0.917조 원으로 확인되었다. 환경 부문은 16건으로 나타났으며, 평균 사업비는 0.633조 원으로 나타났다. 산업, 안전, R&D(연구개발) 부문은 상대적으로 사업 수가 적고, 사업비 규모 또한 다른 부문에 비해 작은 편으로 나타났다. 산업 부문은 9건으로 평균 사업비가 0.328조 원, 안전 부문은 4건으로 평균 0.334조 원, R&D 부문은 4건으로 평균 0.273조 원으로 확인되었다.⁴⁾

4) R&D 부문의 건수 및 규모가 작게 나온 것은, 본 연구의 데이터 구성에 기인하고 있다. 이번 연구는 KDI 공공투자센터에서 발간한 예비타당성조사보고서의 데이터를 활용해서 연구를 진행하였으며 주요 R&D 사업을 담당하는 KISTEP의 자료는 사용하지 않았다.

〈표 3〉 예비타당성조사 부문별 규모

(단위: 개수, 조원)

부문	사업 수	평균액	최저값	최대값	사업 총액
R&D	4	0.273	0.054	0.624	1.093
SOC	149	0.843	0.051	12.428	125.603
건축	10	0.196	0.053	0.629	1.961
교육	4	3.052	0.081	11.907	12.209
국방	2	0.123	0.053	0.194	0.247
문화	19	0.130	0.052	0.455	2.479
복지	8	0.256	0.085	0.531	2.045
산업	14	0.543	0.048	2.916	7.598
안전	2	0.234	0.133	0.335	0.468
행정	14	0.374	0.118	0.740	5.233
환경	16	0.331	0.063	2.496	5.297
합계	242	0.679	0.048	12.428	164.231

2. 분석모형

본 연구는 예비타당성조사 사업의 최고점과 최저점을 활용하여 평가자의 가중치 부여 경향이 점수에 어떠한 영향을 미치는지 실증적으로 확인하는 것을 목표로 한다.

가설 1을 검정하기 위한 회귀분석의 연구모형은 아래와 같다.

$$\begin{aligned} \text{평가점수 (최고점수, 최저점수)} = & \beta_0 + \beta_1 \text{가중치 항목 (경제, 정책)} + \beta_2 \text{사업 규모}_{\ln} + \beta_3 \text{ahp} \\ & + \beta_4 \text{여당 의원수} + \beta_5 \text{야당 의원수} + \beta_6 \text{지역 수} \\ & + \text{연도 FE} + \text{부문 FE} + \text{비수도권 FE} + \epsilon \end{aligned}$$

가설 2를 검정하기 위한 회귀분석의 연구모형은 아래와 같다.

$$\begin{aligned} \text{평가점수 (최고점수, 최저점수)} = & \beta_0 + \beta_1 \text{가중치 편향} + \beta_2 \text{사업 규모}_{\ln} + \beta_3 \text{ahp} \\ & + \beta_4 \text{여당 의원수} + \beta_5 \text{야당 의원수} + \beta_6 \text{지역 수} \\ & + \text{연도 FE} + \text{부문 FE} + \text{비수도권 FE} + \epsilon \end{aligned}$$

IV. 분석 결과

본 연구는 평가자의 가중치 성향이 점수 결과에 미치는 영향에 대해 다중선형회귀분석(OLS)을 수행하였다. 결과는 아래 표에서 확인할 수 있다.

〈표 4〉는 가중치 항목과 최고점수와의 관계를 실증분석한 결과이다. 분석 결과, 경제 가중치의 회귀계수는 모형 (1), (2)에서 모두 음(-)의 방향으로 추정되었으며, 통계적으로 유의하게 나타났다(Model 1: $\beta_1 = -0.1107$, $p < 0.01$; Model 2: $\beta_1 = -0.0849$, $p < 0.01$). 이는 경제성 항목의 중요도를 높게 인식할수록, 평가자가 상대적으로 최고점수를 더 적게 부여한다는 것을 의미한다. 이러한 결과는 가설 1의 내용과 같은 결과를 보여주며 경제성 항목에 더 큰 주의를 기울이는 평가자가 비용·효율성 측면에서 상대적으로 더 엄격한 판단을 내릴 가능성이 크다는 점을 의미한다.

반면에 정책 가중치는 모형 (3), (4)에서 모두 양(+)의 방향으로 추정되었으며, 통계적으로 유의하게 나타났다(Model 3: $\beta_1 = 0.1072$, $p < 0.01$; Model 4: $\beta_1 = 0.1228$, $p < 0.01$). 이는 정책적 필요성이나 공익성에 주목하는 평가자일수록 최고점수를 더 많이 부여하는 것으로 해석할 수 있다. 이는 정책적 판단은 경제성과 달리 직관적·가치지향적 요소(system 1적 판단)에 기초하는 경향이 있어 점수에 관대함이 나타날 수 있다는 행동경제학적 논의와도 연결된다. 즉, 동일한 사업이라도 어떤 기준을 더 중시하는지에 따라 평가자의 점수 부여 방향이 뚜렷하게 달라짐을 보여준다.

통제변수와 고정효과를 포함한 전체 모형의 설명력(r^2)은 매우 높은 수준이며(0.957~0.967), 주요 변수들의 부호는 고정효과 포함 여부와 관계없이 일관되게 나타나고 있다. 이를 통해 경제 가중치의 부정적 효과와 정책 가중치의 긍정적 효과가 단순한 모델 특정 문제에 의해 발생한 것이 아니라, 평가자의 판단 구조와 실제 점수 산정 과정에서 안정적으로 확인되는 패턴임을 알 수 있다.

〈표 4〉 실증분석 결과 1

	coef.	(1)	(2)	(3)	(4)
		종속변수: 최고점수			
경제 가중치	β_1	-0.1107*** (0.0167)	-0.0849*** (0.0197)		
정책 가중치	β_1			0.1072*** (0.0210)	0.1228*** (0.0282)

사업 규모 (ln)	β_2	0.0029** (0.0014)	0.0038*** (0.0015)	0.0031** (0.0014)	0.0041*** (0.0015)
ahp	β_3	0.9639*** (0.0137)	0.9578*** (0.0141)	0.9769*** (0.0140)	0.9599*** (0.0141)
여당의원 수	β_4	0.0003*** (0.0001)	0.0002*** (0.0001)	0.0001 (0.0001)	0.0000 (0.0001)
야당의원 수	β_5	-0.0001 (0.0001)	0.0000 (0.0001)	0.0001 (0.0001)	0.0000 (0.0001)
지역수	β_6	-0.0014 (0.0021)	-0.0010 (0.0021)	-0.0004 (0.0022)	0.0010 (0.0022)
상수항	β_0	0.0870*** (0.0157)	0.0746*** (0.0212)	-0.0041 (0.0157)	-0.0077 (0.0202)
사업 부문 FE		No	Yes	No	Yes
연도 FE		No	Yes	No	Yes
비수도권 FE		No	Yes	No	Yes
r ²		0.959	0.967	0.957	0.967
N		242	242	242	242

주: ()안의 값은 표준오차, p-value: 0.1, 0.05, 0.01

〈표 5〉는 가중치 항목과 최저점수와와의 관계를 실증분석한 결과이다. 분석 결과, 경제 가중치의 회귀계수는 모형 (1), (2)에서 모두 양(+)의 방향으로 추정되었으며, 통계적으로 유의하게 나타났다(Model 1: $\beta_1 = 0.0640$, $p < 0.01$; Model 2: $\beta_1 = 0.0424$, $p < 0.1$). 이는 앞서 살펴봤던 가설 2와 같은 결과로, 경제성 항목의 중요도를 높게 인식할수록 평가자가 상대적으로 최저점수를 더 많이 부여한다는 것을 의미한다.

정책 가중치의 계수는 두 모형 모두에서 음(-)의 값으로 나타나 가설에서 제시한 방향성과 일치한다. 즉, 정책적 고려에 무게를 두는 평가자일수록 상대적으로 낮은 최저점수를 부여하는 경향이 관찰되었다는 점에서 이론적 기대와 부합한다. 그러나 계수의 통계적 유의성이 확보되지 않았기 때문에 이러한 패턴이 표본 내에서 일관된 효과로 해석될 정도의 강한 증거라고 보기는 어렵다(Model 3: $\beta_1 = -0.0251$, $p > 0.1$; Model 4: $\beta_1 = 0.0456$, $p > 0.1$). 그럼에도 계수의 부호가 안정적으로 유지된다는 점은 정책적 판단 기준이 평가자의 극단적 점수 부여 과정에 일정한 방향성을 제공할 수 있음을 시사한다. 다만, 이 효과가 통계적으로 명확하게 드러나지 않는 이유는 평가자 간 정책 판단의 이질성, 점수 부여 과정에서의 상한·하한 제약, 또는 정책성 평가 자체의 상대적 주관성 등으로 인해 신호가 약하게 나타났을 가능성이 있다.

〈표 5〉 실증분석 결과 2

	coef.	(1)	(2)	(3)	(4)
		종속변수: 최저점수			
경제 가중치	β_1	0.0640*** (0.0190)	0.0424* (0.0216)		
정책 가중치	β_1			-0.0251 (0.0254)	-0.0456 (0.0335)
사업 규모 (ln)	β_2	-0.0046*** (0.0014)	-0.0046*** (0.0015)	-0.0042*** (0.0014)	-0.0045*** (0.0015)
ahp	β_3	0.9817*** (0.0142)	0.9910*** (0.0148)	0.9727*** (0.0142)	0.9901*** (0.0149)
여당의원 수	β_4	-0.0003*** (0.0001)	-0.0002* (0.0001)	-0.0002** (0.0001)	-0.0001 (0.0001)
야당의원 수	β_5	0.0002* (0.0001)	0.0001 (0.0001)	0.0001 (0.0001)	0.0000 (0.0001)
지역수	β_6	0.0002 (0.0022)	-0.0005 (0.0022)	-0.0003 (0.0022)	-0.0011 (0.0022)
상수항	β_0	-0.0319* (0.0166)	-0.0221 (0.0218)	0.0068 (0.0163)	0.0120 (0.0218)
사업 부문 FE		No	Yes	No	Yes
연도 FE		No	Yes	No	Yes
비수도권 FE		No	Yes	No	Yes
r2		0.956	0.963	0.954	0.962
N		242	242	242	242

주: ()안의 값은 표준오차, p-value: 0.1, 0.05, 0.01

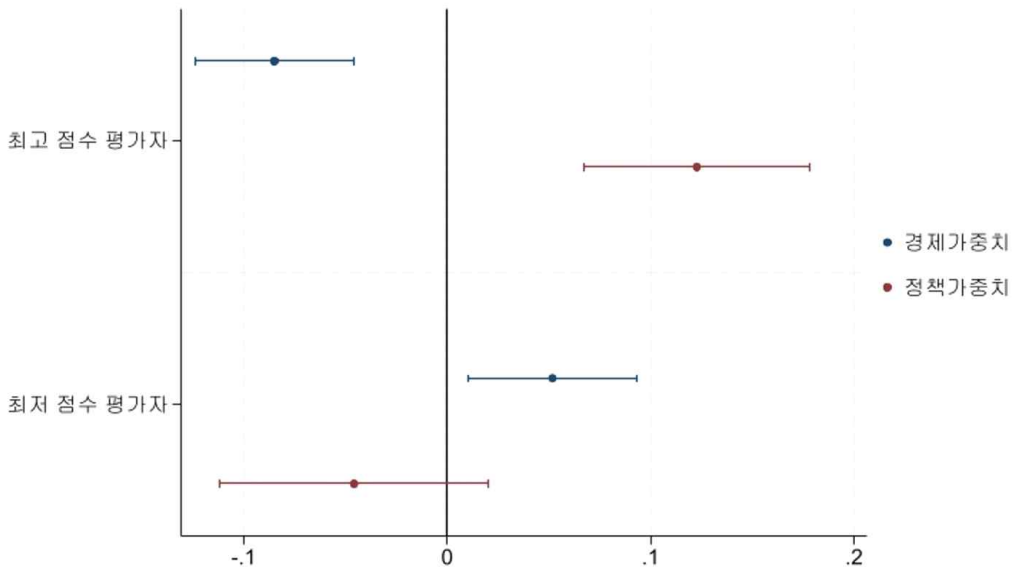
〈그림 2〉는 분석결과 1, 2를 바탕으로 경제 가중치와 정책 가중치가 최고점수 평가자와 최저점수 평가자에게 미치는 영향을 계수와 신뢰구간 형태로 제시하고 있다. 최고점수 평가자의 경우 경제 가중치는 음(-)의 방향으로 나타나고 정책 가중치는 양(+)의 계수로 보이고 있다, 이는 경제성 기준을 중시하는 평가자는 더 적은 최고점수를 부여하고 정책적 요소를 강조하는 평가자는 더 많은 최고점수를 부여하는 경향을 의미한다.

최저점수 평가자에서도 유사한 흐름이 나타나지만 강도와 유의성 측면에서는 차이가 있다. 경제 가중치는 양(+)의 값으로 나타나 경제성을 중시하는 평가자가 최저점수 구간에서 상대적으로 더 높은 점수를 부여하는 경향을 보이나, 효과의 크기는 최고점수 평가자에 비해 제한적이다. 정책 가중치는 음(-)의 방향을 보이며 이론적 예측과는 일치하지만, 신뢰구간이

넓어 통계적으로 확정적 결론을 내리기 어렵다는 점도 그래프에서 직관적으로 드러난다.

전체적으로 그래프는 경제·정책 가중치가 평가자의 극단점 부여 방식에 서로 다른 방향성과 강도로 작용한다는 점을 명확하게 시각화한다. 최고점수에서는 두 변수 모두 뚜렷한 방향성과 유의성을 보이지만, 최저점수에서는 정책 가중치의 효과가 통계적으로 미약하여 상대적으로 불확실성이 크다. 즉, 평가자의 판단 기준이 점수의 상·하한 구간에서 어떻게 다르게 작동하는지를 한눈에 보여주는 시각적 요약이라 할 수 있다.

〈그림 2〉 분석결과 1, 2의 도식화



〈표 6〉은 가중치 편향과 평가 점수 간의 관계를 분석한 결과로, 가중치 차이와 가중치 비율 모두 최고점수에는 양(+), 최저점수에는 음(-)의 방향으로 유의한 영향을 미치는 것으로 나타났다. 이는 특정 항목에 가중치가 과도하게 치우칠수록 평가 결과가 평균으로부터 더 멀어지는 경향이 존재함을 보여준다. 즉, 가중치의 불균형이 평가 범위(최고·최저점수)를 확대시키는 요인이 될 수 있음을 시사하며, 이러한 패턴은 가중치 설정 방식이 점수 산정의 변동성을 높일 수 있다는 본 연구의 기본 가정과도 부합한다. 전체 모형에서 주요 변수의 부호가 일관되게 유지되었다는 점 역시 이러한 경향의 안정성을 뒷받침한다.

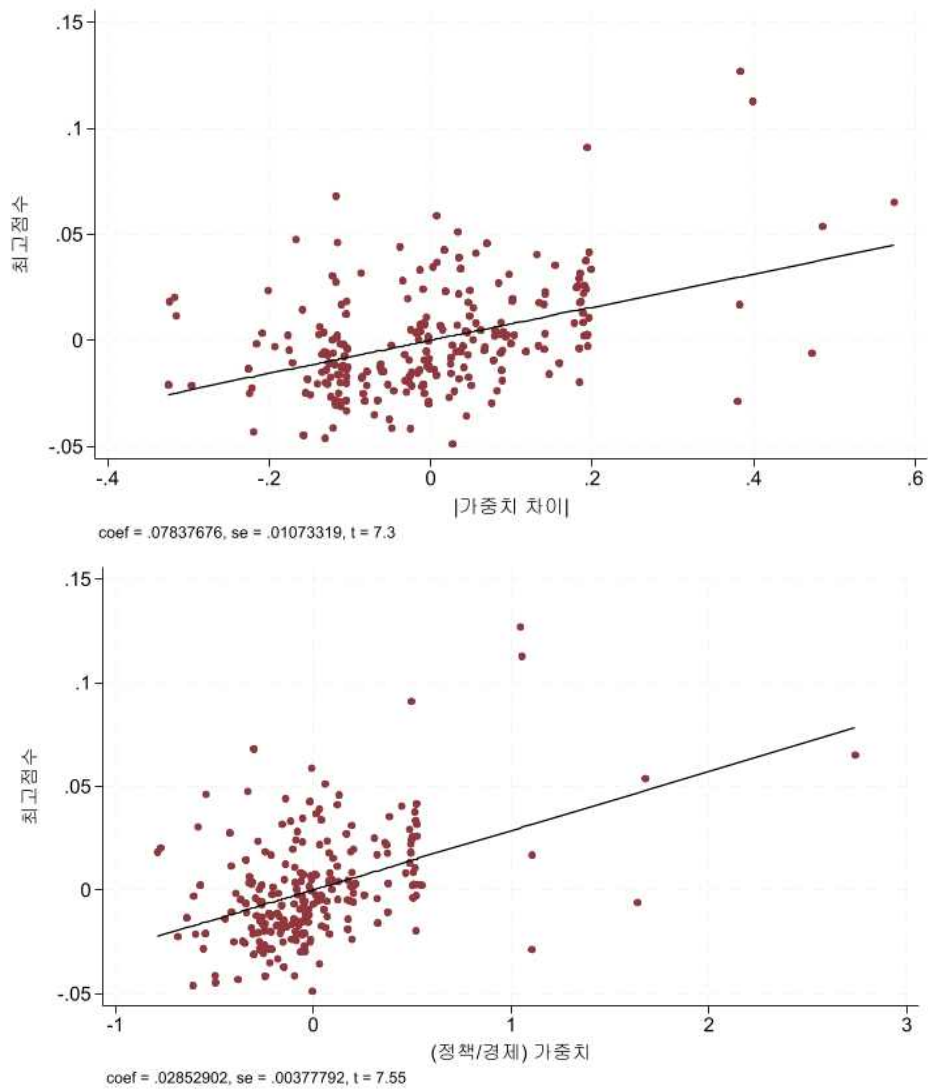
〈표 6〉 실증분석 결과 3

	coef	(1)	(2)	(3)	(4)
		최고점수		최저점수	
가중치 차이	β_1	0.0694*** (0.0134)		-0.0334** (0.0159)	
가중치 비율	β_1		0.0271*** (0.0048)		-0.0159** (0.0074)
사업 규모 (ln)	β_2	0.0039*** (0.0014)	0.0043*** (0.0014)	-0.0046*** (0.0015)	-0.0044*** (0.0015)
ahp	β_3	0.9583*** (0.0139)	0.9585*** (0.0137)	0.9923*** (0.0149)	0.9915*** (0.0148)
여당의원 수	β_4	0.0001* (0.0001)	0.0001 (0.0001)	-0.0001 (0.0001)	-0.0001 (0.0001)
야당의원 수	β_5	0.0000 (0.0001)	0.0000 (0.0001)	0.0001 (0.0001)	0.0001 (0.0001)
지역수	β_6	0.0003 (0.0021)	0.0007 (0.0021)	-0.0008 (0.0022)	-0.0010 (0.0022)
상수항	β_0	0.0455** (0.0184)	0.0153 (0.0181)	-0.0076 (0.0194)	0.0086 (0.0198)
사업 부문 FE		Yes	Yes	Yes	Yes
연도 FE		Yes	Yes	Yes	Yes
비수도권 FE		Yes	Yes	Yes	Yes
r ²		0.968	0.969	0.963	0.963
N		242	242	242	242

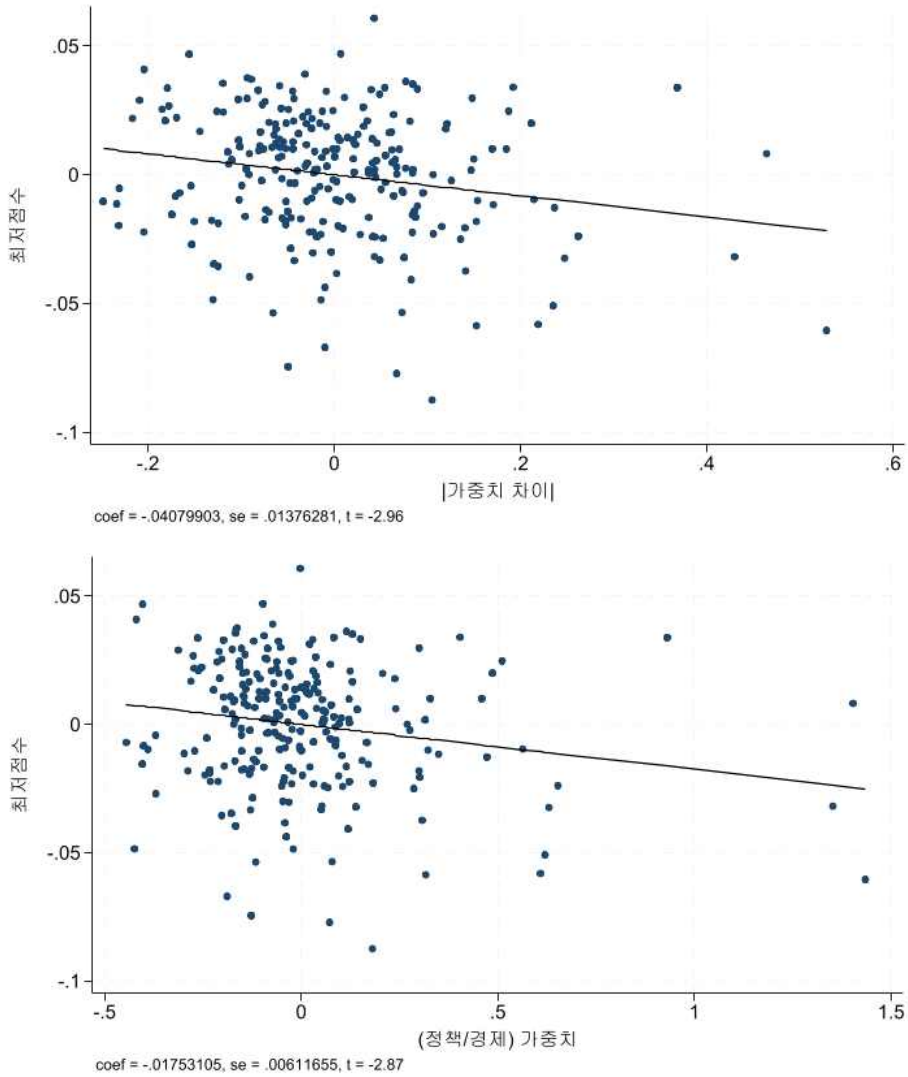
주: ()안의 값은 표준오차, p-value: 0.1, 0.05, 0.01

〈그림 3〉과 〈그림 4〉는 이러한 관계를 시각적으로 제시한 것으로, 가중치 편향이 커질수록 최고점수는 상승하고 최저점수는 하락하는 모습을 확인할 수 있다. 두 산점도에서 나타나는 우상향·우하향의 단순한 추세는, 가중치가 균형적으로 부여되지 않을 경우 평가자 간 점수 차이가 구조적으로 커질 수 있음을 직관적으로 보여준다. 이는 본문의 실증결과를 보완하는 형태로 제시된 것으로, 가중치 불균형과 점수 극단화 간의 전반적인 연관성을 시각적으로 보여준다.

〈그림 3〉 가중치의 편향과 최고점수와의 관계



〈그림 4〉 가중치의 편향과 최저점수와의 관계



V. 결론

본 연구는 예비타당성조사에서 평가 점수의 불일치 원인을 분석하고, 항목별 가중치 성향이 점수에 미치는 영향을 실증적으로 검토하였다. 예비타당성조사 제도의 핵심은 사업에 관한 점수 측정이며 동일한 정보를 바탕으로 평가하더라도 개별 점수의 불일치가 발생하는

원인을 행동경제학 이론에 기반하여 분석하고자 하였다. 기존 연구들이 주로 사업의 통과 여부 및 면제 여부에 집중하였다면, 본 연구는 평가 점수에 가장 직접적인 영향을 미치는 가중치 항목에 주목함으로써 인지적·심리적 판단 과정의 중요성을 다루고자 하였다. 특히 예비타당성조사 사업에 대한 최고점수와 최저점수를 중심으로 극단값이 어떻게 형성되며, 이 과정에서 시스템 1과 시스템 2가 혼합적으로 작용하고 있다는 점에 초점을 두었다.

본 연구에서는 크게 2가지 이론적 기여가 존재한다. 앞서 예비타당성조사 연구에서 상대적으로 다루기 어려웠던 ‘평가자의 판단 과정’에 접근할 수 있는 새로운 분석적 출발점을 제시했다는 점이 있다. 기존 문헌에서는 평가자의 인지적 판단 구조를 직접적으로 분석할 수 없다는 제약 때문에, 점수 산정 과정의 미시적 편차를 체계적으로 설명하기 위한 틀이 충분히 마련되지 못하였다. 본 연구는 자료의 제약하에서도 경제성·정책성 가중치가 최고점수와 최저점수에 서로 다른 방향으로 작용하는 반복적 패턴에 주목함으로써, 평가 항목의 속성 차이가 판단 방식의 차이로 이어질 수 있다는 가능성을 실증적으로 제시하였다. 이러한 패턴은 결과값을 중심으로 한 간접적 관찰이라는 한계가 있으나, 정보 유형의 차이가 평가자별 판단 경향과 연결될 수 있다는 점을 보여주었다는 점에서 중요한 이론적 의미를 가진다.

또한, 본 연구는 이중처리이론을 평가자의 심리적 기제를 규명하기 위한 ‘확정적 설명’이 아닌, 관찰된 패턴을 이해하는 데 도움이 되는 ‘해석적 틀’로 활용함으로써 행동경제학적 논의를 예타 평가 맥락에 적용할 수 있는 방법론적 가능성을 제시하였다. 이를 통해 가중치-점수 구조가 단순한 통계적 변동이 아니라 판단 기준의 차이를 반영할 수 있다는 점을 보다 정교하게 논의할 수 있었다. 특히 예타 연구에서 이론적 논의가 제도설계·정치적 영향 요인에 집중되어 있는 상황에서, 본 연구는 행동경제학적 접근이 점수 편차 분석에 유의미하게 기여할 수 있음을 보여주며, 향후 평가자 수준의 판단 연구를 확장할 수 있는 토대를 제공한다는 점에서 이론적 기여를 갖는다.

정책적 측면에서는 예비타당성조사 제도의 공정성과 일관성을 강화하기 위해 주목해야 할 지점을 제시한다. 분석 결과는 평가자가 특정 항목에 높은 가중치를 부여할 경우 최고점수 또는 최저점수가 한 방향으로 치우치는 경향이 존재함을 보여주었는데, 이는 가중치 설정의 편차가 최종 점수의 변동성을 키울 수 있음을 의미한다. 예타 제도는 동일한 자료와 지침을 제공하고 있음에도 평가자 간 점수 차이가 발생한다는 점에서, 가중치 설정 방식에 대한 검토와 점수 산정 과정의 투명성 확보가 필요함을 시사한다. 특히 극단값 분석에서 확인된 패턴은 일부 평가자의 판단 기준이 전체 평가 결과의 균형을 흔들 수 있는 가능성을 보여준다는 점에서, 평가자 교육·가중치 설정 절차 점검 등 제도적 고려가 요구된다.

아울러 본 연구가 제시한 결과는 향후 예타 제도 개선 논의에서 평가자 구성, 항목 간 가중

치 범위 설정, 평가 기준의 명확화가 중요한 정책 요소가 될 수 있음을 의미한다. 평가자의 심리적·인지적 특성을 직접 파악하지 못하더라도, 가중치·점수 구조의 반복적 패턴을 통해 판단 과정의 편차 가능성을 사전에 인지하고 대응할 필요가 있다. 장기적으로는 항목별 세부 기준의 정교화, 평가 항목의 성격에 따른 가중치 안내 강화, 그리고 평가 결과의 극단적 변동을 최소화하기 위한 점검 절차를 도입함으로써 예타 제도의 신뢰성과 예측 가능성을 높일 수 있을 것이다. 이러한 정책적 시사점은 본 연구가 제시한 간접적 증거를 기반으로 하며, 향후 자료 접근성이 확대된다면 보다 정교한 정책 개선 방향을 설계하는 데 기초자료로 활용될 수 있을 것이다.

다만 본 연구는 구조적 제약으로 인해 몇 가지 한계를 가진다. 첫째, 예비타당성조사 제도의 특성상 평가자의 개별 성향, 전문 배경, 가치지향 등과 같은 미시적 정보는 일체 공개되지 않기 때문에, 가중치 부여가 어떠한 개인적 요인에 의해 형성되었는지를 직접적으로 확인할 수 없다. 즉, 본 연구가 관찰할 수 있는 것은 가중치와 점수라는 결과값이며, 그 결과값을 통해 판단 방식의 가능성을 간접적으로 추론하는 수준에 머무를 수밖에 없다. 둘째, 데이터의 한계로 인해 평가자별 점수 중 최고점수와 최저점수만을 활용하여 연구를 진행하였다는 점이다. 평가자 전체의 심리적 혹은 인지적 패턴을 정교하게 식별하기보다 극단값을 중심으로 한 패턴 분석에 초점을 두었다는 것이 본 연구의 한계로 볼 수 있다. 이러한 연구의 제약은 이론-가설-실증 간의 연결고리를 약화시키며 이중처리이론을 평가자의 실제 심리구조로 해석하는 것이 아니라, 가중치 부여 패턴을 해석할 수 있는 하나의 설명틀로만 활용하였다는 지적을 받을 수 있다.

그럼에도 본 연구가 이중처리이론을 참고한 이유는, 경제성과 정책성 가중치가 서로 다른 방향의 점수 경향을 반복적으로 형성하는 구조적 패턴을 이해하기 위한 적절한 개념적 틀을 제공하기 때문이다. 본 연구는 이중처리이론을 특정 판단의 ‘원인’으로 단정하거나, 경제성 항목에 시스템 2가, 정책성 항목에 시스템 1이 ‘적용된다’고 주장하려는 것이 아니다. 오히려 정보 속성의 차이가 평가자들의 판단 방식에 서로 다른 접근을 요구할 수 있으며, 이러한 차이가 가중치 부여 성향에 반영될 가능성을 제시하는 정도의 해석적 틀로 활용하였다. 인지적 체계의 실제 작동을 검증하기 위해서는 개인 배경정보와 실험 설계가 필요하며, 본 연구는 그러한 직접적 검증을 수행한 연구가 아니라 가중치·점수 패턴을 설명하기 위한 탐색적 접근임을 본문에서 명확히 하였다.

이러한 한계에도 불구하고 본 연구는 예비타당성조사 과정에서 미시적 판단 요소를 분석할 수 있는 새로운 분석 틀을 제시하였다는 점에서 학술적 의의를 갖는다. 향후 연구에서는 평가자의 성향을 직접 파악할 수 있는 설문조사·실험자료를 활용하거나, 기관별·사업유형

별 자료를 확장하여 분석한다면 보다 정교한 인지적 메커니즘을 확인할 수 있을 것이다. 또한 평가자 익명성 유지 체계를 훼손하지 않는 범위에서, 보다 세분화된 점수 데이터의 공개가 이루어진다면 예타 평가의 공정성과 일관성을 강화하는 정책적 개선 방안을 제시하는 데 기여할 수 있을 것으로 기대한다. 향후에는 다양한 사업 유형과 더 많은 표본을 포함한 확장된 연구가 이루어진다면 더 풍부한 분석이 가능할 것으로 기대한다.

향후 연구에서는 평가자의 심리적 판단을 더 정교하게 측정할 수 있는 설문조사나 실험적 기법을 병행하거나, 평가자 특성 데이터와 연계한 분석을 진행한다면 제도의 공정성과 효율성을 제고할 수 있는 정책적 시사점을 제시할 수 있을 것이다. 또한, 정치적 이해관계와 맞물려 있는 예비타당성조사제도 특성상 정치적 압력과 인지 전략의 상호작용 가능성이 존재할 수 있다. 본 연구에서 활용한 정치 변수(여당·야당 의원 수)는 분석의 핵심 설명 변수가 아니라 지역 정치 환경을 통제하기 위한 보조적 요소로 사용되었으며, 일부 모형에서 나타난 통계적 유의성 또한 해석의 중심에 두지 않았다. 그럼에도 불구하고 정치적 맥락이 평가자의 판단 방식이나 가중치 설정에 어떤 간접적 영향을 미칠 수 있는지, 혹은 정치적 압력과 평가자의 인지 전략이 어떠한 방식으로 상호작용할 수 있는지는 향후 보다 풍부한 자료와 연구설계를 통해 탐구할 필요가 있다. 향후 연구에서는 평가자의 특성 정보 또는 지역·정치 환경과 평가 행동 간 관계를 보다 정교하게 분석할 수 있는 조건이 마련된다면, 예비타당성조사의 공정성과 일관성 제고를 위한 보다 실질적인 정책적 함의를 도출할 수 있을 것으로 기대한다. 마지막으로 향후 데이터 처리 여건과 연구 범위가 확대 가능하다면, 극단값의 활용뿐 아니라 개별 점수 전수 분석을 포함한 확장 연구가 이루어져서 더욱더 풍부한 연구가 될 수 있을 것으로 본다.

참고문헌

- 고길근. (2007). 다차원 공공투자 의사결정 문제에 있어서 정책분석가들의 의사결정 행태분석: 가중치 부여행위를 중심으로. 「한국정책학회보」, 16(1), 23-48.
- 공공투자관리센터. (2015). 2015년도 공공투자관리센터 연차보고서. KDI.
- 공공투자관리센터. (2019). 2019년도 공공투자관리센터 연차보고서. KDI.
- 기획재정부. (2019). 예비타당성조사 정책브리핑.
- 기획재정부. (2022). 예비타당성조사 운용지침. 국가법령정보센터.
- 김강수. (2016). 대규모 사업의 예비타당성조사 성과와 과제. 「국토」, 413(3), 12-17.
- 김봉환·최근혁. (2022). 정책이념이 예타면제에 미치는 영향 분석. 「행정논총」, 60(3), 77-100.
- 김봉환·최근혁·정효용·최근혁. (2024). 예비타당성조사의 정치성 연구: 어느 과정에서 어떻게 작

동하는가? 「재정학연구」, 17(2), 35-67.

김정숙·안민우·엄영호. (2017). 정책분석가 구성이 예비타당성조사 사업비 증감에 미치는 영향: 예비타당성조사 (2012-2014) 분석을 중심으로. 「정부학연구」, 23(1), 95-115

김태운·조예진. (2018). 예비타당성조사제도의 타당성에 대한 연구: 예비타당성조사의 지침을 둘러싼 쟁점. 「행정논총」, 56(2), 279-311.

김태일. (2019). 예비 타당성 조사의 쟁점 및 개편안 분석: 건설 사업을 중심으로. 「한국행정학보」, 53(3), 243-268.

박민정. (2020). 국책사업 예비타당성조사 면제를 둘러싼 정치경제학적 논의: 지대 추구론을 중심으로. 「한국자치행정학보」, 34(3), 101-123.

박승준. (2022). 연구개발사업 예비타당성조사 제도 개선효과에 관한 연구. 「한국정책연구」, 22(4), 29-43.

신가희·하연섭. (2015). 예산심의과정에서 예비타당성조사제도의 정치적 수용성에 관한 연구-예산결산특별위원회 회의록 분석을 중심으로. 「한국정책학회보」, 24(2), 527-564.

엄근용. (2019). 경제·사회 환경 변화에 따른 예비타당성조사 제도의 개선 방안. 「이슈포커스」, 2019(1), 1-30.

정동호. (2024). 예비타당성조사 제도개편의 효과 분석 - 2019년 변화를 중심으로. 「한국정책학회보」, 33(2), 163-190.

최근혁·김봉환. (2025). 예비타당성조사제도 변화에 관한 연구: 비수도권 사업의 통과확률이 비대칭적으로 늘었는가?. 「행정논총」, 63(1), 141-168.

Brehmer, B. (1994). The psychology of linear judgement models. *Acta Psychologica*, 87(2-3), 137-154.

Fischhoff, B., et al. (1978). How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. *Policy sciences*, 9, 127-152.

Fiske, S. T. T., & Taylor, S. E. (2020). Social cognition: From brains to culture.

Frank, R. H., et al. (1993). Does studying economics inhibit cooperation? *Journal of economic perspectives*, 7(2), 159-171.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual review of psychology*, 62(2011), 451-482.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge university press.

Hodgkinson, G. P., Langan-Fox, J., & Sadler-Smith, E. (2008). Intuition: A fundamental bridging construct in the behavioural sciences. *British Journal of Psychology*, 99(1), 1-27.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive psychology*, 24(1), 1-55.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics.

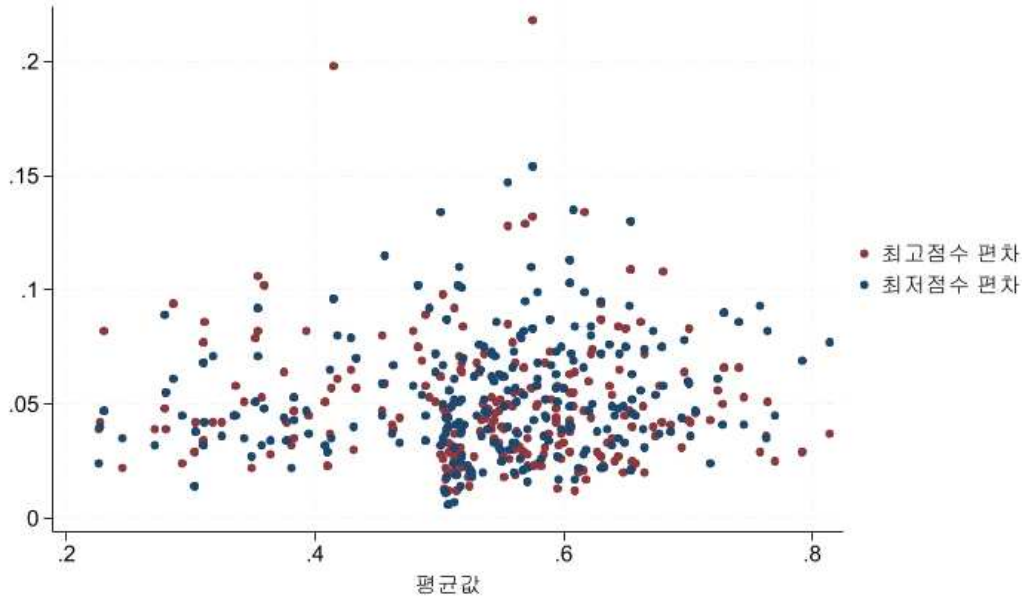
- American economic review*, 93(5), 1449-1475.
- Kahneman, D. (2011). *Fast and slow thinking*. Allen Lane and Penguin Books, New York.
- Kahneman, D. & Klein, G. (2009). Conditions for intuitive expertise: a failure to disagree. *American psychologist*, 64(6), 515.
- Kahneman, D. & Tversky, A. (1984). Choices, values, and frames. *American psychologist*, 39(4), 341.
- Keren, G., & Teigen, K. H. (2004). Yet another look at the heuristics and biases approach. *Blackwell handbook of judgment and decision making*, 89-109.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological bulletin*, 108(3), 480.
- Mackie, P., & Preston, J. (1998). Twenty-one sources of error and bias in transport project appraisal. *Transport policy*, 5(1), 1-7.
- Mitchell, G., et al. (1993). Judgments of social justice: Compromises between equality and efficiency. *Journal of Personality and Social Psychology*, 65(4), 629.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, 2(2), 175-220.
- Reyna, V. F. (2004). How people make decisions that involve risk: A dual-processes approach. *Current Directions in Psychological Science*, 13(2), 60-66.
- Simon, H. A. (1972). Theories of bounded rationality. *Decision and organization*, 1(1), 161-176.
- Slovic, P. (1995). The construction of preference. *American psychologist*, 50(5), 364.
- Slovic, P. (2000). What does it mean to know a cumulative risk? Adolescents' perceptions of short-term and long-term consequences of smoking. *Journal of behavioral decision making*, 13(2), 259-266.
- Stanovich, K. E., & West, R. F. (2000). Advancing the rationality debate. *Behavioral and brain sciences*, 23(5), 701-717.
- Stanovich, K. E. & West, R. F. (2007). Natural myside bias is independent of cognitive ability. *Thinking & Reasoning*, 13(3), 225-247.
- Sunstein, C. R. (Ed.). (2000). *Behavioral law and economics*. Cambridge University Press.
- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, 185(4157), 1124-1131.
- Wang, J., Stelmakh, I., Wei, Y., & Shah, N. (2024). Debiasing evaluations that are biased by evaluations. *Journal of Machine Learning Research*, 25(315), 1-120.
- Weber, E. U., & Johnson, E. J. (2009). Mindful judgment and decision making. *Annual review of psychology*, 60(1), 53-85.

부록

〈부록 표 1〉 평균-극단값 편차의 기술통계

변수	표본크기	평균	표준편차	최소값	최대값
최고점수와 평균값의 차이	242	0.048	0.027	0.012	0.218
최저점수와 평균값의 차이	242	0.053	0.026	0.006	0.154
(최고-최저) 점수의 차이	242	0.101	0.047	0.018	0.372
$\frac{\text{최고 점수} - \text{평균 값}}{\text{평균 값}}$	242	0.096	0.063	0.019	0.477
$\frac{\text{최저 점수} - \text{평균 값}}{\text{평균 값}}$	242	0.102	0.053	0.011	0.318

〈부록 그림 1〉 평균값과 극단값 편차의 분포도



ABSTRACT

Why Do Scores Diverge: Weighting Preferences and Evaluation Disagreement in Preliminary Feasibility Studies

Geunhyuk Choi & Bong Hwan Kim

This study investigates the sources of score disagreement in preliminary feasibility studies and empirically examines how evaluators' weighting preferences across assessment criteria influence evaluation scores. While prior research has largely focused on project approval outcomes or institutional arrangements, relatively little attention has been paid to the micro-level role of individual weighting preferences in shaping evaluation results. To address this gap, this study adopts the highest and lowest scores as dependent variables and analyzes their relationship with criterion-specific weighting preferences. The results indicate that higher maximum scores are associated with lower weights on economic feasibility and higher weights on policy relevance. In contrast, higher minimum scores tend to emerge when greater weight is placed on economic feasibility and less on policy relevance. These findings can be interpreted as reflecting differences in judgment processes rooted in the distinct characteristics of evaluation criteria. This study draws on the dual-process theory in behavioral economics, proposing that systematic patterns of judgment can be inferred from evaluators' weighting preferences. By doing so, it offers a theoretical explanation for judgment bias arising in the evaluation process of preliminary feasibility studies and suggests avenues for improving fairness and consistency in public project appraisal.

【Keywords: preliminary feasibility study, weighting preferences, judgment bias, behavioral economics】