

Conceptual Validity Verification of a Narrative-Based User Experience Evaluation Framework: Expert Validation of Freytag's Pyramid and PAD-Heatmap Integration

Sunyoung Jang*, Hokyoung Ryu**

*Ph.D. Student, Graduate School of Technology and Innovation Management, Hanyang University, Seoul, Korea

**Professor, Graduate School of Technology and Innovation Management, Hanyang University, Seoul, Korea

[Abstract]

UX evaluations of public digital services rely on single-point measurements taken after the fact, failing to capture changes in usability at each stage of the user journey. Furthermore, they are limited by the fact that, during retrospectives, the overall experience is distorted in memory due to the Peak-End Rule. This study proposes a narrative-based UX evaluation framework that applies Freytag's Pyramid, comprising Exposition, Rising Action, Climax, Falling Action and Dénouement to the public digital service user journey. This framework integrates Pleasure-Arousal-Dominance(PAD) self-report measurements and heatmap behavioral data through triangulation. Conceptual validity was examined through semi-structured interviews with seven UI/UX experts and deductive qualitative content analysis following Elo and Kyngäs. Expert validation showed unanimous agreement on practical value over existing methods and supported the conceptual validity of the Dominance dimension as a critical indicator for capturing perceived control in mandatory public service contexts. As an exploratory study examining the conceptual validity of the framework, this study establishes a theoretical foundation for integrating narrative theory into UX evaluation methodology and provides concrete grounds for the measurement design of subsequent empirical studies. However, operational variables for the framework, such as measurement interface design, particularly the influence of popup-based PAD measurement on dimension-specific scores, the operational definition of stage boundaries, and PAD-heatmap interpretation criteria, require empirical refinement through subsequent experimental research.

▶ **Key words:** Narrative-based UX evaluation, Freytag's Pyramid, PAD model, Heatmap triangulation, Public digital services, Administrative burden

-
- First Author: Sunyoung Jang, Corresponding Author: Hokyoung Ryu
 - *Sunyoung Jang (sunny112@hanyang.ac.kr), Graduate School of Technology and Innovation Management, Hanyang University
 - **Hokyoung Ryu (hryu@hanyang.ac.kr), Graduate School of Technology and Innovation Management, Hanyang University
 - Received: 2026. 03. 10, Revised: 2026. 05. 12, Accepted: 2026. 05. 13.

[요 약]

공공 디지털 서비스 UX 평가는 사후 단일 시점 측정에 의존하여 이용 과정의 단계별 사용성 변화를 포착하지 못하며, 회고 시 피크-엔드 법칙(Peak-End Rule)에 따라 전체 경험이 왜곡되어 기억되는 한계를 지닌다. 본 연구는 이를 극복하기 위해 Freytag의 5단계 구조(Freytag's Pyramid: 발단-상승-절정-하강-결말)를 공공 디지털 서비스 사용자 여정에 적용하고, 각 단계에서 PAD (Pleasure-Arousal-Dominance) 자기보고 측정과 히트맵 행동 데이터를 삼각검증(triangulation)하는 내러티브 기반 UX 평가 프레임워크를 제안한다. 제안한 프레임워크의 개념적 타당성 검증을 위해 UI/UX 분야 전문가 7인을 대상으로 반구조화 인터뷰(semi-structured interview)를 실시하고 Elo와 Kyngäs의 연역적 질적 내용분석(deductive content analysis)을 적용하였다. 전문가 검증 결과, 기존 방법 대비 실무적 가치에서 전문가 전원이 찬성을 제시하였고, 통제감(Dominance) 차원은 이탈이 불가능한 공공 서비스 맥락에서 통제감 변화를 포착하는 핵심 지표로서의 개념적 타당성이 지지되었다. 본 연구는 프레임워크의 개념적 타당성을 검증하는 탐색적 연구로서, 내러티브 이론을 UX 평가 방법론에 접목하는 이론적 기반을 마련하고 후속 실험 연구의 측정 설계 방향에 구체적 근거를 제공한다. 다만 측정 인터페이스 설계에서 팝업 방식의 PAD 측정이 차원별 측정값에 미치는 영향, 단계 경계의 조작적 정의, PAD-히트맵 해석 기준 등 프레임워크 운용을 위한 세부 설계 변수는 후속 실험 연구를 통한 실증적 정교화가 필요하다.

▶ **주제어:** 내러티브 기반 UX 평가, Freytag 5단계 구조, PAD 모델, 히트맵 삼각검증, 공공 디지털 서비스, 행정 부담

I. Introduction

한국 정부는 2022년 '디지털플랫폼정부 추진계획'을 수립하여 대국민 서비스 전달체계 혁신, AI·데이터 기반 과학적 국정운영 등 4대 과제를 제시하였다[1]. 이 중 서비스 전달체계 혁신의 일환으로, 공공 웹·앱의 직관적이고 일관성 있는 디자인을 위한 'UI/UX 혁신체계' 확립을 추진하고, UI/UX 표준모델 개발과 디자인 평가제 도입 등의 로드맵을 포함하였다[1]. 이를 구체화한 2023년 '디지털플랫폼정부 실현계획'에서는 기관별 시스템 분절과 아날로그 중심의 행정 절차가 국민에게 불필요한 시간 낭비(time-tax)를 야기한다는 문제를 진단하였으며[2], 이에 따라 정부는 2024년 '디지털 정부서비스 UI/UX 가이드라인' 및 디자인 시스템을 배포하였다[3]. 그러나 이러한 정책적 노력에도 불구하고, 서비스 이용 과정에서 사용자가 경험하는 어려움을 체계적으로 진단하는 평가 방법론에 대한 논의는 상대적으로 부족하다.

현행 UI/UX 평가는 서비스 이용 후 "서비스 이용이 편리했나요?", "색상간의 조화가 잘 이루어지고 있나요?"와 같은 포괄적 문항에 대한 5점 척도 응답으로 구성되어 있으며[4], 과업 완료 시간과 오류 횟수 등 효율성·효과성 중심의 전통적 사용성 평가에 머물러 있다[5]. 이러한 평가

방식은 "불편했다"는 결과는 파악할 수 있으나, "어느 지점에서 왜 불편했는지"는 식별할 수 없다는 근본적 한계를 지닌다. 디지털 서비스 이용 경험은 진입, 탐색, 수행, 완료 등 여러 단계를 거치는 시간적 여정임에도[6], 현행 평가는 서비스 이용을 모두 마친 후 경험 전체를 하나의 점수로 압축하도록 요구한다[4]. 이처럼 사후 회상에 의존하는 측정 방식은 피크-엔드 법칙(Peak-End Rule)으로 대표되는 회고적 편향의 영향을 받기 쉬우며, 가장 강렬한 순간과 마지막 순간을 중심으로 평가하는 경향은 중간 과정에서 발생한 문제를 기억에서 희석시킬 수 있다[7].

이러한 평가의 한계는 공공 디지털 서비스의 구조적 특수성을 고려할 때 더욱 심화된다. Moynihan et al.[8]은 공공 서비스 이용 과정에서 시민이 경험하는 행정 부담을 학습 비용, 순응 비용, 심리적 비용의 세 차원으로 구분하며, 심리적 비용은 자율성 상실과 무력감, 그리고 불확실성으로 인한 스트레스를 포함한다고 설명한다[8]. 민간 서비스와 달리 공공 서비스에서는 이용자가 불만족 상태에서 과업을 중단하거나 대안을 선택할 수 없으며[9], 예측하지 못한 추가 서류 요구나 난해한 행정 용어가 과업의 흐름을 반복적으로 차단할 때 이러한 심리적 비용은 서비

스 이용 전 과정에 걸쳐 단계적으로 누적된다[8]. 즉 공공 서비스 이용 경험은 단순한 편의성의 문제가 아니라, 감정적 긴장이 고조되고 해소되는 동적 과정으로 이해되어야 한다. 그러나 현행 사후 단일 시점 평가는 이러한 감정적 긴장의 동적 흐름을 포착하지 못한다는 근본적 한계를 지닌다.

본 연구는 이러한 한계를 극복하기 위해 내러티브 이론을 사용자 경험 평가에 도입한다. 내러티브는 인간이 시간적으로 전개되는 경험을 조직하고 의미를 부여하는 근본적 인지 양식으로[10], 경험의 감정적 긴장이 고조되고 해소되는 과정을 시간축 위에 단계적으로 구조화하는 특성을 지닌다. 사용자 경험 연구에서 대표적으로 활용되는 고객 여정 지도(Customer Journey Map, CJM)는 서비스 접점의 전환을 기준으로 단계를 구분하는데[6], 이는 앞서 확인한 감정적 긴장의 방향 변화를 단계 경계로 포착하기에는 충분하지 않다. 이에 본 연구는 긴장의 고조와 해소를 시간축 위에 단계적으로 배치하는 Freytag[11]의 5단계 구조(Freytag's Pyramid, 이하 Freytag 구조)를 채택한다. Freytag 구조는 전통적 서사 분석을 넘어, 데이터 기반 스토리텔링 연구에서도 정보의 시간적 전개를 조직하는 체계적 틀로 활용한 바 있다[12]. 본 연구는 이 구조를 공공 디지털 서비스 사용자 여정에 적용하고, 각 단계에서 Mehrabian과 Russell[13]의 PAD(Pleasure-Arousal-Dominance) 감정 모델(이하 PAD 모델)에 기반한 자기보고식 측정과 히트맵 행동 데이터를 내러티브 단계라는 공통의 시간축 위에서 삼각검증(triangulation)[37]하는 평가 프레임워크를 제안한다. PAD 모델의 통제감(Dominance)차원은 이탈이 구조적으로 제한되는 공공 서비스 맥락에서[9] 사용자가 서비스 진행에 대해 느끼는 통제감의 변화를 포착하는 핵심 지표로 기능할 수 있다.

본 연구의 목적은 위와 같이 제안된 프레임워크의 개념적 타당성을 전문가 타당성 검증을 통해 확인하는 것이다. 이를 위해 UI/UX 분야 전문가 7인을 대상으로 반구조화 인터뷰(semi-structured interview)를 실시하고, Elo와 Kyngäs[14]의 연역적 질적 내용분석 절차에 따라 사전 설정된 4개 검증 영역, 즉 Freytag 구조의 공공 디지털 서비스 적용 타당성, PAD 모델의 측정 설계 적합성, 히트맵 행동 데이터와의 통합 타당성, 실무적 적용 가능성을 분석 범주로 설정하여 전문가의 판단을 체계적으로 분석한다. 본 논문은 제II장 이론적 배경, 제III장 연구 방법, 제IV장 결과 분석, 제V장 결론으로 구성된다.

II. Theoretical Background

2.1 Theoretical Limitations of UX Evaluation in Public Digital Services

Moynihan, Herd와 Harvey[8]는 시민과 정부 간 상호작용에서 발생하는 부담을 행정부담(administrative burden)으로 개념화하고, 이를 학습비용, 심리비용, 순응비용의 세 가지 유형으로 체계화하였다. 학습비용은 서비스의 존재와 자격 요건을 파악하는 과정에서, 순응비용은 서류 제출·양식 작성 등 행정적 요구를 이행하는 과정에서 발생하며, 심리비용은 자율성 상실, 무력감, 결과의 불확실성에서 비롯되는 정서적 부담을 의미한다[8]. 이 세 가지 비용은 서비스 이용의 특정 시점에 일괄 발생하는 것이 아니라, 서비스 접근에서 완료에 이르는 전 과정에 걸쳐 단계적으로 발생하고 누적되는 특징을 지닌다[8]. 그러나 현행 공공 디지털 서비스 평가는 서비스 이용을 마친 후 단일 시점에서 전체 경험을 하나의 점수로 측정하는 방식에 의존하고 있어, 이용 과정의 각 단계에서 차별적으로 작동하는 부담의 성격과 강도를 식별하기 어렵다는 한계를 지닌다.

이러한 단계별 부담의 누적적 특성은 측정 방식 자체의 한계와도 결합된다. Kahneman[15]은 매 순간의 경험을 느끼는 경험 자아와 경험 종료 후 이를 회상하는 기억 자아를 구분하였으며, Redelmeier와 Kahneman[7]은 회고적 평가가 경험 중 가장 강렬했던 순간과 마지막 순간에 의해 주로 결정되고 전체 지속 시간은 거의 반영되지 않는다는 피크-엔드 법칙(Peak-End Rule)을 실증하였다. 즉 사후 단일 시점 평가는 이용 과정 전반에 걸쳐 단계적으로 누적된 부담을 이미 왜곡된 기억을 통해 측정한다는 이중적 한계를 지닌다. Vermeeren et al.[16]은 96개의 사용자 경험 평가 방법을 분석한 결과, 대다수가 단일 에피소드의 순간적 평가에 집중되어 있으며 경험의 시간적 전개를 포착하는 방법은 부족하다고 보고하였다. 이는 공공 서비스 이용 경험처럼 여러 단계에 걸쳐 부담이 누적되는 맥락에서, 이용 과정의 각 단계에서 감정 반응을 측정하고 이를 시간적 순서에 따라 구조화하는 평가 방법이 필요함을 시사한다.

2.2 Structural Characteristics of Public Digital Service Experience

이러한 구조적 한계는 공공 디지털 서비스와 민간 디지털 서비스의 경험 평가 조건이 근본적으로 다르다는 점에서 비롯된다. 민간 디지털 서비스 환경에서 사용자는 경험의 질을 판단할 때 두 가지 인지적 자원을 활용한다. 하나

Table 1. Comparison of Experience Evaluation Structures between Private and Public Digital Services

Comparative Attribute	Commercial Digital Services	Public Digital Services	Comparative Basis
Usage Motivation	Voluntary selection	Mandatory/compulsory use	[8]
Alternative Availability	Multiple alternatives, free switching	Monopolistic provision, no alternatives	[9]
Exit Possibility	Exit possible upon dissatisfaction	Exit unavailable, task completion mandatory	[9], [20]
Usage Frequency and Script Formation	Cognitive script accumulation through repeated use	Limited script formation due to infrequent and irregular use	[18], [19]
Reference Point Formation	Relative standards through inter-service comparison	Confusion in reference point formation due to absence of comparative targets	[17]
Primary Sources of Experience Evaluation Bias	Peak-End Rule + contrast effect against reference points + expectation bias from script violation	Loss of perceived control + compounded Peak-End Rule + sustained uncertainty after task completion leading to ambiguous experience endpoint	[7], [9], [17], [18]

는 동종 서비스 간 비교를 통해 형성되는 상대적 준거점(reference point)이다[17]. 사용자는 복수의 대안 서비스를 경험하면서 각 서비스의 품질을 상대적으로 판단할 수 있는 평가 기준을 내적으로 구성한다[17]. 다른 하나는 단일 서비스의 반복 이용을 통해 축적되는 인지적 스크립트(cognitive script)에 기반한 순차적 기대 구조이다[18]. 사용자는 반복된 이용 경험으로부터 서비스 절차에 대한 내적 표상을 형성하며, 이를 통해 다음 단계의 상호작용을 예측하고 현재 경험의 적절성을 판단한다[18].

그러나 공공 디지털 서비스는 이러한 평가 조건이 구조적으로 상이하다. Table 1은 양 서비스 유형 간 경험 평가 구조의 핵심 차이를 정리한 것이다. 공공 디지털 서비스는 독점적으로 제공되어 비교 대상이 부재하며, 사용자의 이탈이 구조적으로 제한된다[9][20]. 비정기적 이용으로 인해 인지적 스크립트가 형성되지 않아 매 이용 시 높은 내재적 인지부하(intrinsic cognitive load)에 노출된다[18][19]. 이러한 조건에서 추가 서류 요구, 난해한 행정 용어, 사전 처리 절차 등 예측 불가능한 과업 중단이 반복되면서 학습 비용, 순응비용, 심리비용이 동시에 작용한다[8]. 이러한 구조적 제약이 결합되면, 이탈 불가능성과 과업 중단의 반복이 사용자의 통제감을 점진적으로 약화하고, 부정적 감정이 해소되지 못한 채 누적되는 경험 전개 구조를 형성할 수 있다[8].

이러한 조건은 경험 평가의 왜곡 방식에도 차이를 만든다. 민간 디지털 서비스에서는 준거점 기반 대비 효과와 스크립트 위반에 따른 기대 편향이 평가를 왜곡하는 주된 요인으로 작용하며, 피크-엔드 법칙에 의해 절정과 종료 시점의 경험이 전체 평가를 대표하게 된다[7][17][18]. 반면 공공 디지털 서비스에서는 이탈 불가능성으로 인한 통제감 상실이 피크-엔드 법칙과 복합적으로 작용한다. 또한

과업 완료 후에도 처리 결과의 불확실성이 지속되는 경우, 하강과 결말 단계의 경계가 모호해져 경험의 종결점 자체가 불분명해지는 추가적 왜곡이 발생할 수 있다[7][9].

민간 서비스에서는 사용자가 준거점과 인지적 스크립트를 바탕으로 경험의 전개를 예측하며, 불만족 시 이탈을 통해 부정적 경험을 중단할 수 있다. 반면 공공 서비스에서는 이러한 예측 자원이 부재한 상태에서 부담이 누적되더라도 과업 완수까지 경험이 강제로 지속된다[9][20]. 이는 두 서비스 유형 간 경험의 시간적 전개 과정 자체가 상이함을 시사한다. 특히 공공 디지털 서비스에서는 이탈 불가능성과 과업 완수의 강제성으로 인해 부정적 감정이 누적되고 해소되지 못하는 긴장의 고조와 이완 과정이 경험 전반에 걸쳐 나타난다. 따라서 단순히 시간적 순서를 기록하는 것을 넘어, 감정적 긴장의 고조와 해소라는 동적 흐름을 구조화할 수 있는 분석적 틀이 요구된다.

2.3 Narrative Theory for Structuring Temporal Experience

내러티브 이론은 인간이 시간적으로 전개되는 경험을 조직하고 의미를 부여하는 인지적 기제를 설명한다는 점에서, 공공 디지털 서비스의 경험 평가에 요구되는 감정적 긴장의 동적 흐름을 구조화하는 분석적 틀로 기능할 수 있다. Bruner[21]는 시간 순서에 따라 사건들을 인과적으로 연결하여 의미를 구성하는 내러티브 양식을 패러다임적 양식과 구분하였으며, Ricoeur[22]는 비정형적 시간 경험을 이해 가능한 형태로 전환하는 것이 서사적 구조화(emplotment)라 하였다. Adaval과 Wyer[23]는 동일 정보를 내러티브 형식으로 제시했을 때 목록 형식보다 호의적 평가가 유도됨을 실증하였으며, Pennington과 Hastie[24]는 사람들이 복잡한 정보를 시간적 순서에 따라

서사적으로 조직하여 판단함을 보여주었다.

서비스 경험의 시간적 흐름을 구조화하는 기존 도구로는 고객여정지도(Customer Journey Map, 이하 CJM)가 대표적으로 활용되어 왔다[6]. CJM은 사용자가 서비스와 접촉하는 터치포인트를 기준으로 단계를 구분하고 각 접점의 행동과 감정을 기록한다[6]. 그러나 터치포인트 기반의 단계 구분은 경험의 위치를 식별할 수 있을 뿐, 감정적 긴장이 어느 방향으로 변화하는가라는 방향성을 구조적으로 포착하기 어렵다. 동일한 터치포인트라도 긴장이 고조되는 국면에 속하는지 이완되는 국면에 속하는지에 따라 경험의 성격은 근본적으로 달라지기 때문이다.

내러티브의 시간적 전개를 단계화하는 대표적 모델은 Freytag 구조이다. 이 모델은 발단(exposition), 상승(rising action), 절정(climax), 하강(falling action), 결말(dénouement)¹⁾의 5단계로 구성되며, 본 연구는 이를 감정적 긴장의 방향이 전환되는 지점을 기준으로 단계 경계가 형성되는 구조로 해석한다[29]. CJM이 터치포인트를 기준으로 단계를 구분하는 것과 달리, Freytag 구조는 긴장의 고조와 해소라는 방향성을 단계 경계의 기준으로 삼는다는 점에서 차별된다. 또한 하나의 서비스 이용 경험 안에서 Freytag 구조가 한 번이 아니라 여러 차례 나타날 수 있다. 앞 절에서 논의한 바와 같이[8][9], 공공 디지털 서비스에서는 긴장이 고조되었다가 부분적으로 해소되는 작은 긴장 주기가 여러 차례 반복된다. 이는 하나의 서비스 이용 경험이 단일 피라미드가 아니라 복수의 Freytag 파형이 연속되는 구조로 나타날 수 있음을 시사하며, 단계 경계는 페이지 전환이 아니라 긴장 방향의 전환 여부에 의해 결정된다. 따라서 CJM의 터치포인트 기반 단계 구분만으로는 이러한 반복적 긴장 구조를 포착하기 어렵다.

Freytag 구조는 데이터 스토리텔링 연구[12] 및 UX 실무에서도 활용 가능성이 제안되었고[26], 내러티브 이론 자체를 제품 디자인[25] 분야에 적용한 연구도 있다. 그러나 이를 감정 변화를 단계별로 측정하는 평가 프레임워크로 전환하려는 시도는 전산 서사학(computational narratology) 분야의 연구들에서 확인된다. Reagan 등[27]은 감성 분석을 통해 내러티브의 감정 곡선을 정량적으로 추출하였고, Boyd 등[28]은 Freytag 구조의 단계적 패턴이 다양한 장르에 걸쳐 언어적 수준에서 재현됨을 확인하여 구조적 보편성을 입증하였다. Schmidt와 Winkler[29]는 플롯 단위별로 정서(Valence), 각성

(Arousal), 정서 강도를 측정하여, 서사의 구조적 단계가 정서 측정의 시간적 단위(temporal unit)로 직접 활용될 수 있음을 방법론적으로 입증하였다.

이상의 논의는 Freytag 구조가 사용자 경험에서 감정적 긴장의 고조와 해소라는 동적 흐름을 단계화하는 분석적 틀로 기능할 수 있음을 시사한다. 그러나 이러한 틀이 실제 측정 프레임워크로 작동하려면, 각 단계에서 사용자의 정서적 반응을 체계적으로 포착할 수 있는 감정 측정 모델의 선정이 선행되어야 하며, 이는 서사 구조와 정서 반응의 상관관계 연구를 통해 뒷받침된다[29].

2.4 Emotion Measurement Models and PAD Dimensions

사용자 경험 연구에서 감정 측정 접근법은 이산적 접근(discrete approach)과 차원적 접근(dimensional approach)으로 구분된다. 이산적 접근은 인간의 감정이 질적으로 구별되는 기본 범주들로 구성된다는 관점에 기반하며, Ekman[30]은 문화적 보편성을 가진 여섯 가지 기본 감정을, Plutchik[31]은 여덟 가지 기본 감정과 그 조합에 의한 파생 구조를 제안하였다. 이산적 접근은 특정 감정 상태를 명확하게 식별할 수 있다는 장점이 있으나, 이산적 감정 각각이 고유한 신경체계에 의해 뒷받침된다는 증거가 비일관적이며[32], 감정의 연속적 변화나 복합적 상태를 표현하는 데 한계가 있다. 차원적 접근은 모든 감정 상태가 소수의 연속적 차원 위에 위치한다는 관점을 취한다. Russell[33]의 감정 차원 모델(circumplex model of affect)은 정서(Valence)와 각성(Arousal)의 두 차원으로 감정 공간을 구성하며, Mehrabian과 Russell[13]은 여기에 통제감(Dominance) 차원을 추가한 PAD 모델을 제안하였다. 감정의 연속적 변화를 시간 축 위에서 추적하고 단계별로 비교하기 위해서는 차원적 접근이 적합하며, 그 중에서도 PAD 모델을 채택한다.

PAD 모델이 Russell의 2차원 모델과 구별되는 핵심은 통제감(Dominance) 차원의 포함 여부이다. 2차원 모델에서 동일한 부정적 정서(Valence)와 높은 각성(Arousal)의 조합으로 기술되는 감정 상태가, PAD 모델에서는 통제감(Dominance) 수준에 따라 질적으로 변별된다. 통제감(Dominance)이 높으면 상황을 통제할 수 있다고 느끼는 감정 상태로, 통제감(Dominance)이 낮으면 상황에 압도되어 무력한 감정 상태로 구별될 수 있다[13]. 다만 PAD 모델의 세 차원이

1) Freytag의 원서(1863)에서 5단계는 독일어 Katastrophe(파국)로 표기되어 있다. 이는 비극 구조 분석을 전제로 한 용어로, 본 연구에서는 공공 디지털 서비스의 사용자 여정에 적용하는 맥락을 고려하여 영미권 학술 문헌에서 통용되는 Dénouement로 대체하여 사용한다.

모든 맥락에서 동등한 설명력을 갖는 것은 아니다. 민간 디지털 서비스에서는 선택과 비교가 가능하므로 즐거움 (Pleasure) 차원이 만족도 평가의 주된 지표로 기능할 수 있다. 반면 앞 절에서 논의한 공공 디지털 서비스의 구조적 특성[8][9]에서는, 서비스가 즐거운가보다 내가 원하는 대로 진행되고 있는가와 긴장이 얼마나 고조되어 있는가가 경험의 질을 결정하는 핵심 요인이 된다. 이러한 맥락에서 사용자의 감정은 단순한 즐거움(Pleasure)의 하락이 아니라, "내가 원하는 대로 진행되고 있는가, 아니면 예측할 수 없는 상황에 무력하게 놓여 있는가"라는 통제감 (Dominance)의 변동으로 더 정확하게 포착될 수 있다. 따라서 공공 디지털 서비스 맥락에서는 각성(Arousal)과 통제감(Dominance) 차원이 즐거움(Pleasure)보다 상대적으로 높은 설명력을 가질 것으로 예측된다.

측정 도구 설계에 있어, PAD 모델의 대표적 측정 도구인 SAM(Self-Assessment Manikin)[34]은 각 차원을 비언어적 그림으로 표현하여 시각적 직관성이 높으나, 본래 5단계 또는 9단계의 이산적 척도로 설계되어 중심화 경향 편향의 문제를 내포한다. 본 연구에서는 SAM의 시각적 직관성을 유지하면서, 이모지 기반 양 끝점 표시와 시각 아날로그 척도(Visual Analogue Scale, 이하 VAS)를 결합한 인터페이스를 설계한다. VAS는 중심화 경향 편향을 완화하고 등간 척도 데이터를 생성하여 모수적 통계 기법의 적용에 적합하다[35]. 또한 측정 시점의 개입이 서비스 이용 흐름을 방해하여 감정 반응 자체에 영향을 줄 수 있다는 반응성(reactivity) 문제를 고려하여, 측정 인터페이스는 Freytag 구조 단계 전환 시점에 최소한으로 개입하

는 방식으로 설계한다.

이상의 직접 측정과 별도로, 감정의 간접 측정은 사용자의 행동 데이터로부터 감정 상태를 추론하는 방법으로, 표정 분석, 음성 분석, 행동 로그 분석 등이 이에 해당한다. 그 중 히트맵으로 대표되는 행동 로그 데이터는 사용자에게 추가적 부담 없이 연속적으로 수집할 수 있다는 강점을 가지지만, 동일한 행동 패턴이 상이한 감정 상태에서 발생할 수 있다는 해석적 모호성을 내포하며, 이는 사용성 측정 전반에서 발생하는 도전 과제이다[36][41]. 이러한 모호성은 행동 데이터만으로는 사용자의 감정 상태를 신뢰성 있게 추론하기 어렵다는 방법론적 한계를 드러낸다. 특히 동일한 히트맵 패턴이 상이한 감정 상태에서 발생할 수 있다는 점에서[36][41], PAD 직접 측정 데이터와의 결합 없는 행동 데이터의 해석 방향을 확정하기 어려울 수 있다. 이는 직접 측정과 간접 측정을 공통의 시간적 단위 위에서 통합하는 삼각검증의 필요성을 시사하며[37], 본 연구에서는 Freytag 구조를 그 통합 축으로 활용한다.

2.5 Structural Foundation of Narrative-Based UX Evaluation Framework

앞선 논의를 종합하면, 공공 디지털 서비스의 사용자 경험 평가에는 세 가지 방법론적 요구가 도출된다. 첫째, 공공 서비스의 구조적 특수성으로 인해 경험의 시간적 전개 과정에서 감정적 긴장이 고조되고 해소되는 전환 시점을 포착할 수 있는 구조적 틀이 필요하다. 둘째, PAD 모델의 검토에서 확인한 바와 같이 각 단계에서의 감정 변화를 포착하되, 특히 통제감(Dominance)의 변동을 측정할 수 있

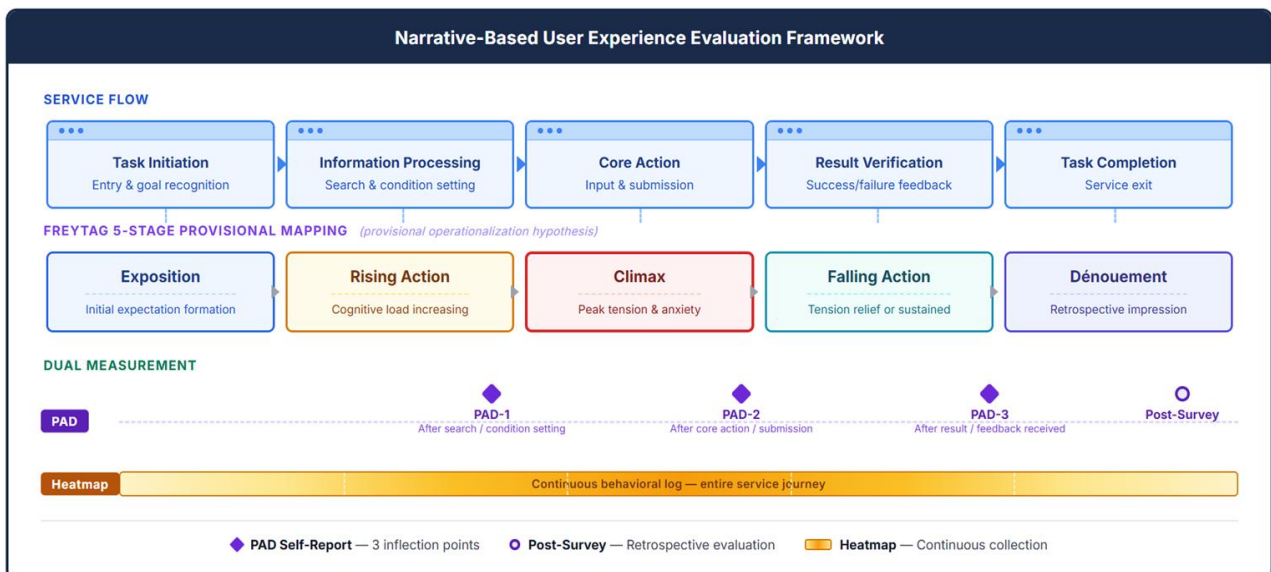


Fig. 1. Application Structure of the Narrative-Based UX Evaluation Framework (Freytag stage mapping represents a provisional operationalization hypothesis)

어야 한다. 셋째, 간접 측정의 해석적 모호성을 고려할 때 주관적 감정 데이터와 객관적 행동 데이터의 삼각검증이 요구된다.

이러한 요구에 대응하는 구조적 기반은 공공 디지털 서비스 사용자 여정에서 찾아볼 수 있다. 디지털 서비스의 사용자 여정은 서비스 진입에서 과업 완료까지 시간적으로 전개되는 순차적 과정이며[6], 본인인증 완료나 신청서 제출 등 성공과 실패가 결정되는 전환점이 존재한다. 이러한 속성은 긴장의 고조와 해소를 시간축 위에 단계적으로 배치하는 Freytag 구조와 구조적으로 대응한다.

본 연구는 행정안전부의 '디지털정부 서비스 UI/UX 가이드라인'[3]에서 정의한 방문, 검색, 로그인, 신청, 정책정보확인 등 다섯 가지 서비스 패턴을 분석하여, 과업 시작, 정보 처리, 핵심 행위, 결과 확인, 과업 종료라는 공통 여정 구조를 도출하였다. 이 공통 구조는 Freytag의 5단계와 잠정적으로 대응하는 것으로 가정하며, 서비스 유형에 따라 일부 단계가 압축되거나 확장될 수 있다[10].

이상을 바탕으로, 본 연구는 Freytag 구조를 공공 디지털 서비스 사용자 여정의 시간적 전개를 단계화하는 측정틀로 활용하고, 각 단계에서 PAD 자기보고 측정과 히트맵 행동 데이터 수집을 이 틀 위에서 통합 수행하는 내러티브 기반 사용자 경험 평가 프레임워크를 제안한다. Fig. 1은 본 프레임워크의 이론적 도출 구조를 나타낸다. PAD 측정은 상승, 절정, 하강 3개 변곡점에서 자기보고 방식으로

수행되어 단일 시점이 아닌 복수 시점에서의 감정 변화를 포착하며[38], 과업 종료 후 사후 설문을 통해 회고적 인상을 수집한다. 히트맵 행동 로그는 전 과정에서 연속적으로 수집되어, PAD 자기보고 데이터와 내러티브 단계라는 공통의 시간축 위에서 삼각검증된다[37]. 이 프레임워크에 대한 전문가 타당성 검증은 다음 장에서 수행된다.

III. Research Method

3.1 Framework Design

본 연구는 행정안전부의 '디지털정부 서비스 UI/UX 가이드라인'[3]에서 정의한 5가지 서비스 패턴(방문, 검색, 로그인, 신청, 정책정보확인)의 이용 플로를 분석하여 공통 여정 구조를 도출하고, 이를 Freytag의 5단계 구조와 잠정적으로 대응시켰다. Table 2는 도출된 공통 구조와 Freytag 5단계 간의 잠정적 대응 관계를 정리한 것이다. 본 프레임워크는 각 내러티브 단계에서 주관적 감정 측정과 객관적 행동 측정을 병행하는 이중 측정 구조를 채택한다. 감정 측정은 PAD 모델에 기반하며, 상승-절정, 절정-하강, 하강-결말의 3개 변곡점에서 자기보고 방식으로 수행된다. 행동 측정은 히트맵 데이터(마우스 클릭, 이동 경로, 체류 시간 등)를 전 과정에 걸쳐 연속적으로 수집한다. 두 데이터를 내러티브 단계라는 공통의 시간축 위에서 삼

Table 2. Provisional Mapping between Common Service Structure and Freytag's Five Stages

Freytag Stage	Task Characteristics	Emotional Dynamics
Exposition (Task Initiation)	Service entry, goal recognition	Initial expectation formation
Rising Action (Information Processing)	Information search, condition setting	Cognitive load & tension gradually increasing
Climax (Core Action)	Input, decision, confirmation, submission	Peak of concentration, anxiety & tension
Falling Action (Result Verification)	Success/failure feedback	Gradual tension relief or sustained tension
Dénouement (Task Completion)	Service exit	Retrospective impression formation

* Stage-to-service-flow correspondence represents a provisional operationalization hypothesis subject to empirical refinement in subsequent research.

Table 3. Expert Profiles

ID	Degree / Major	Expertise	Career (yr)	Position	Interview
E1	Master's / Digital Design	UI/UX	24	Director	In-person
E2	Master's / Cognitive and Perceptual Psychology	UI/UX	10	Lead	In-person
E3	PhD / Industrial Engineering	HCI, UI/UX	8	Senior Product Manager	Remote
E4	PhD / Visual Design	UX Design	20	CEO	Remote
E5	PhD / Industrial Engineering	UI/UX Planning	20	Director	In-person
E6	Master's / Business Administration	UI/UX Planning	25	CEO	In-person
E7	PhD / Cognitive Science	HCI, UI/UX	25	CSO	Remote

각검증함으로써, 주관적 감정 측정의 회고적 왜곡과 객관적 행동 측정의 해석적 모호성을 상호 보완한다[37].

3.2 Expert Selection

프레임워크의 개념적 타당성을 검증하기 위해, 목적적 표집(purposive sampling)을 통해 UI/UX 분야 전문가 7인을 선정하였다. 선정 기준은 디지털 서비스의 UX 설계, 디자인, HCI 분야에서 실무 경력 8년 이상을 보유하고, 관련 분야 석사학위 이상의 학력을 갖춘 전문가로 설정하였다. 7인의 전문가 패널은 개념적 타당성의 탐색적 검증을 목적으로 하는 본 연구의 핵심 쟁점을 식별하기에 적절한 규모이다[39]. 질적 연구에서 적정 표본 수는 통계적 확률이 아닌 '정보의 포화'를 기준으로 하며, 선행 연구에 따르면 고도의 전문성을 갖춘 집단의 경우 6~12명 사이에서 주제적 포화가 발생함이 입증된 바 있다[40]. Elo와 Kyngäs[14]는 내용분석의 타당성 확보를 위해 전문가 패널의 활용을 방법론적 요건으로 제시하였으며, 본 연구는 검증 영역 4개·하위 범주 9개가 인터뷰 실시 전 사전 확정된 연역적 설계를 채택하고 있어, 귀납적 주제 탐색 연구 대비 추가 인터뷰를 통한 새로운 범주 출현 가능성이 구조적으로 낮다. 또한 합의 이론(consensus theory)에 따르면, 특정 도메인 역량이 높은 참여자들로부터는 5~7명의 소수 인원으로도 .80 이상의 높은 신뢰도를 도출할 수 있다 [40]. Table 3은 참여 전문가의 프로필을 정리한 것이다.

3.3 Data Collection

자료 수집은 반구조화 인터뷰(semi-structured interview)를 통해 이루어졌다. 사전에 구조화된 인터뷰 가이드에 기반하되, 전문가 응답에 따라 심화 질문(probing question)을 추가하는 유연한 진행 방식을 채택하였으며, 동일한 인터뷰 가이드를 사용하여 대면과 원격 화상 회의를 병행하였다.

인터뷰는 2026년 2월 23일부터 3월 4일까지 진행되었으며, 1인당 소요 시간은 약 60~90분이었다. 인터뷰 실시 3~5일 전, 프레임워크 개요서(Freytag-PAD 통합 구조, 서비스 패턴별 매핑 표, PAD-히트맵 매핑 로직)와 인터뷰 질문지를 사전 배포하였다. 인터뷰 당일에는 참여자 동의 하에 전 과정을 녹음하였으며, 연구 목적 및 프레임워크 개요 설명 후 4개 검증 영역(Freytag 단계 매핑 타당성, PAD 측정 설계, 히트맵 통합 가치, 실용적 구현 가능성)에 대해 순차적으로 질문하였다. 영역별 질문 시에는 사전에 준비한 히트맵 예시 화면, PAD 측정 인터페이스 샘플 등 자극 자료를 제시하였으며, 마지막으로 프레임워크 전반에 대한 개방형 질문으로 마무리하였다.

3.4 Data Analysis

인터뷰 데이터는 Elo와 Kyngäs[14]의 연역적 질적 내용분석(deductive qualitative content analysis) 절차에 따라 분석하였다[14]. 본 연구는 검증 영역이 사전에 설정

Table 4. Structure of Categorization Matrix

Main Category	Sub-category	Definition
D1. Validity of Applying Freytag's Pyramid to Public Digital Services	D1-1. Validity of Five-Stage Mapping	Logical validity of mapping Freytag's Pyramid onto the public digital service user journey
	D1-2. Validity of Common Journey Abstraction	Representativeness of abstracting user journeys across five service patterns into a common structure
D2. Appropriateness of PAD Model Measurement Design	D2-1. Appropriateness of Dominance Dimension	Whether the Dominance dimension is appropriate for measuring user experience in mandatory-use public service contexts
	D2-2. Sufficiency of Measurement Timing	Whether measurements at three stage-transition points are sufficient to capture emotional changes
	D2-3. Feasibility of Measurement Method	Whether a visual analog scale based on emoji is practically feasible for actual users
D3. Validity of Heatmap Behavioral Data Integration	D3-1. Value of PAD-Heatmap Triangulation	Whether integrating subjective PAD measurements with objective heatmap behavioral data provides additional value over single-metric measurement
	D3-2. Appropriateness of Heatmap Indicators	Whether the selection of six heatmap indicators (Click, Move, H2C, Path, Dwell, Scroll) is appropriate for capturing user behavior
D4. Practical Applicability	D4-1. Practical Value over Existing Methods	Distinctive practical value of the proposed framework compared to existing post-task survey and task-completion evaluations
	D4-2. Implementation Barriers and Improvement Priorities	Anticipated barriers and priority improvements upon practical adoption (open-ended)

되어 있으므로, 이를 분석 범주로 활용하는 연역적 접근을 채택하고 구조화된 분류 매트릭스(structured categorization matrix)를 개발하였다. 분류 매트릭스의 4개 주범주(D1~D4)와 9개 하위 범주(D1-1~D4-2)는 3.1 절에서 제시한 프레임워크의 구성요소에 근거하여 도출되었다. 구체적으로, 프레임워크의 세 가지 핵심 구성요소인 Freytag 5단계 구조[11, 12], PAD 3차원 모델[13], 히트맵 행동 데이터 통합[37]은 각각 D1(Freytag 적용 타당성), D2(PAD 측정 설계 적절성), D3(히트맵 통합 타당성)으로 매핑되었으며, 프레임워크의 실제 적용 가능성을 검증하기 위한 영역으로 D4(실용적 적용 가능성)를 추가하였다. 각 주범주 내 하위 범주는 해당 구성요소에서 검증이 요구되는 세부 타당성 쟁점을 바탕으로 세분화하였다. 이 과정은 Elo와 Kyngäs[14]가 제시한 연역적 내용분석의 준비 단계(preparation phase) 절차에 따라 인터뷰 실시 전에 완성되었으며, 이후 인터뷰 질문도 본 분류 매트릭스의 각 영역에서 파생하여 구성하였다. 따라서 본 분류 매트릭스는 선행 연구 기반의 프레임워크 설계에서 표준 방법론 절차를 통해 연역적으로 도출된 구조이며, 그 범주 구조는 Table 4에 정리하였다.

분석 절차는 3단계로 진행하였다. 첫째, 준비(Preparation) 단계에서 인터뷰 녹음을 텍스트로 변환(전사, transcription)한 후, 전사본을 반복적으로 읽어 데이터 전체에 대한 이해를 확보하였다. 둘째, 조직화(Organizing) 단계에서 Table 4의 9개 하위 범주를 사전 범주로 설정한 분류 매트릭스를 기반으로, 전사본에서 각 범주에 해당하는 의미 단위(meaning unit)를 식별하고 코드를 부여하였다. 각 의미 단위는 해당 범주에 대한 전문가의 판정(찬성, 조건부 찬성, 반대)과 함께 매트릭스에 배

치하였다. 셋째, 보고(Reporting) 단계에서 범주별 전문가 합의 수준과 핵심 의견을 종합하고, 대표 인용문을 선정하여 결과를 구성하였다.

범주별 타당성 판정은 사전에 확정된 규칙에 따라 수행되었다. 구체적으로, (1) 찬성(○) 판정이 전문가 7인 중 4인 이상으로 수렴될 경우 해당 범주의 개념적 타당성이 인정된 것(Support)으로, (2) 반대(×) 판정이 복수로 수렴될 경우 타당성 불인정 또는 근본적 재설계 필요로, (3) 위 두 조건에 해당하지 않으면서 찬성(○)과 조건부 찬성(△)이 혼재할 경우 운용 기준 구체화를 전제로 한 조건부 타당성 인정(Conditional Support)으로 판정하였다. 이러한 판정 기준은 도메인 전문성을 갖춘 소수 전문가 집단에서도 높은 신뢰도의 합의 수준을 도출할 수 있다는 합의 이론[40]에 근거한다. 이 판정 규칙은 분석 전에 사전 확정되었으며, 각 전문가의 개별 판정과 종합 판정(Overall) 결과는 Table 5에 투명하게 공개되어 독자가 규칙의 적용 과정을 직접 추적할 수 있다.

Table 4-1은 위 분석 절차에서 원문 발언이 의미 단위와 코드를 거쳐 사전 확정된 하위 범주 및 주범주로 배치되는 추상화 경로를 전문가 발언 두 사례로 예시한 것이다. 서로 다른 판정 유형(찬성, 조건부 찬성)을 보이는 두 사례를 병치함으로써, 동일한 분석 매커니즘이 판정 유형에 관계없이 일관되게 적용됨을 드러낸다.

본 연구자는 검증 대상 프레임워크의 설계자이며, 이에 따른 확증 편향(confirmation bias) 위험을 최소화하고 연구의 신뢰성(trustworthiness)[41]을 확보하기 위해 다음의 절차를 수행하였다.

첫째, 분석 범주(분류 매트릭스 4개 주범주, 9개 하위 범주)를 인터뷰 실시 전에 사전 확정하고, 분석 과정에서

Table 4-1. Examples of Abstraction Path Based on the Categorization Matrix

Component	Example 1	Example 2
Raw Statement	[E1] "PAD seems appropriate in that it allows us to observe users' emotions multidimensionally across Pleasure, Arousal, and Dominance. In particular, measuring Dominance in public service contexts is so critical that without it, this study would have little meaning."	[E3] "If you define the page where cognitive load is highest and errors occur most frequently as the climax, then the mapping would be valid."
Meaning Unit	In the public service context, the Dominance dimension functions as a key indicator that complements the limitations of conventional two-dimensional emotion measurement	If the climax stage is operationally defined as the point of highest cognitive load and error concentration, the Freytag-based journey mapping is valid
Code	Multidimensionality of PAD measurement / Centrality of Dominance in public services	Proposed operational criterion for the climax stage / Conditional acceptance
Sub-category	D2-1. Dominance dimension	D1-2. Journey abstraction
Category	D2, PAD Measurement	D1, Freytag Validity
Judgment	○ Support	△ Conditional support

범주의 추가·수정을 배제하는 연역적 분석 프로토콜을 채택하였다[14]. 이는 분석자의 사후 재구성에 의한 편향 개입을 구조적으로 차단하기 위한 것이다.

둘째, 전문가의 조건부 찬성(△) 및 반대(×) 발언을 찬성(○)과 동등한 비중으로 데이터에 포함하고, 프레임워크의 한계나 문제점을 지적하는 발언을 선별적으로 누락하지 않도록 하였다.

셋째, Lincoln과 Guba[41]의 신뢰성 기준에 따라 참여자 확인(member checking)을 실시하였다. 인터뷰 종료 후 인용한 핵심 발언과 해석 내용을 해당 전문가에게 이메일로 송부하였으며, 7인 전원으로부터 수정 요청 없이 해석에 동의한다는 회신을 확인하였다.

넷째, 분석 절차 및 판정 기준을 논문 내 Table 4와 Table 5에 투명하게 공개하고, 범주별 전문가 판정의 근거가 되는 발언 원문을 대표 인용문으로 병기하여 독자가 분석 경로와 판정 근거를 함께 추적할 수 있도록 하였다.

다섯째, 9개 사전 범주에 포함되지 않는 새로운 관점은 '기타 의견' 범주를 추가하여 별도 수집하였다. 이는 첫째의 연역적 분석 프로토콜을 유지하면서도, 사전 범주 체계에 포함되지 않는 발언의 존재 여부를 투명하게 기록하기 위한 감사 장치(audit trail)이며, Elo와 Kyngäs[14]가 제시한 연역적 내용분석 절차에서 명시적으로 허용하는 확장 장치이다.

IV. Results and Analysis

4.1 Overview of Expert Validation Results

7인의 UI/UX 전문가(E1-E7)를 대상으로 실시한 반구조화 인터뷰와 연역적 질적 내용분석을 통해 9개 하위 범주에 대한 개념적 타당성을 검증하였다. 본 장에서 제시하는 전문가 판정 분포와 핵심 의견은 프레임워크의 개념적 타당성에 대한 전문가 합의 수준을 보여주는 것이며, 프레임워크의 측정 정확도나 편향 감소 효과에 대한 실증적 근거를 제시하는 것은 아니다. Table 5는 범주별 전문가 판정 분포를 요약한 것이다.

D4-1(기존 방법 대비 실무적 가치)에서 전문가 7인이 전원 찬성(○)으로 가장 강하게 수렴되었으며, D3-1(PAD-히트맵 삼각검증 가치)에서도 4인 찬성(○)으로 높은 지지가 확인된 반면, D1과 D2 범주에서는 이론적 타당성을 인정하면서도 프레임워크 운용 기준 구체화 및 측정 시점 설계 정확화를 조건으로 제시하는 조건부 찬성(△)이 우세하였다.

4.2 Validity of Freytag Five-Stage Mapping (D1)

5단계 매핑 적절성(D1-1)에 대해 전문가 1인이 찬성(○), 나머지 6인이 조건부 찬성(△)을 제시하였다. 전문가들은 공공 서비스 이용자가 경험하는 긴장의 흐름이 Freytag의 5단계 구조와 구조적으로 부합한다는 점에서 기본 타당성을 인정하였다. 행안부의 '디지털정부 서비스 UI/UX 가이드라인'[3] 개발 및 정부24 UI/UX 컨설팅에 직접 참여한 E1은 실무 경험을 근거로 이를 확인하였으며, E4는 공공 디지털 서비스 프로젝트에서 직접 관찰한 이용자의 심리적 긴장 흐름으로 뒷받침하였다.

[E4] "사용자는 이걸 써야 돼 말아야 돼, 내가 지원금을 못 받으면 어떡하지 하는 그런 긴장감을 가지면서 서비스 유입부터 결말까지 가는 것 같거든요. 그런 측면에서 이 구조는 맞다는 생각이 들어요."

E6 또한 긴장 고조와 해소의 흐름이라는 관점에서 구조적 타당성을 인정하였다.

[E6] "심리적으로 긴장이 고조됐다가 해소되는 일련의 과정들을 설명하기에 어느 정도 적절한 틀이지 않겠나라는 생각은 했습니다."

다만 두 가지 공통 조건이 제기되었다. 첫째, 절정을 '인지 부하 최고·오류 집중 구간'으로 정의한 것은 현행 공공 디지털 서비스 수준을 전제로는 타당하나, 서비스 품질이 민간 수준으로 향상될 경우 절정 위치의 재설정이 필요하다(E4). 둘째, E5는 신청형 서비스에서 하강과 결말의 경계가 모호해지는 경우가 많아 실질적으로는 4단계에 가까울 수 있다고 지적하였다. E3는 다중 시점 측정의 가치는 인정하면서도, 서비스 패턴별로 행동 단계를 감정적 긴장 수준으로 추상화하는 과정의 표준화가 어렵다는 점을 반복적으로 지적하였다.

공통 여정 추상화 타당성(D1-2)에 대해서는 E4가 찬성(○), 나머지 전문가들이 조건부 찬성(△)을 제시하였다. 추상화의 개념적 타당성은 공통으로 인정하면서도, 서비스 패턴 단위 설정 기준(E1·E2)과 방문·로그인 패턴처럼 목표 강도가 낮은 서비스에서 각 단계의 경계를 실제 서비스에서 어떻게 구분할 것인지에 대한 기준(E5)을 추가로 명확히 해야 한다는 의견이 제기되었다. E3는 절정을 인지 부하와 오류가 집중되는 지점으로 정의하는 방식에 동의하였다.

[E3] "그 과업이 가장 인지 부하가 많이 발생하고 오류가 많이 발생하는 페이지를 절정으로 정의를 했다면 그러면 맞을 것 같아요."

E7은 Freytag 구조를 사용자 여정 지도의 대체제가 아닌, 여정의 각 단계에서 감정 변화를 구조화하는 보조 도구로 재개념화할 것을 제안하였다. 아울러 E4·E5·E6은 독립적으로 Freytag 구조의 중첩과 반복 패턴을 확인하였다. E4는 고용24에서 인증서 등록과 신청서 제출이 각각 독립된 긴장 피라미드를 형성하는 사례를 제시하였고, E5는 공공 서비스가 민간보다 반복 구조가 더 빈번하다고 판단하였으며, E6은 자격 조회와 서류 작성 단계에서도 각각 독립적인 상승-절정 흐름이 반복될 수 있음을 지적하였다.

4.3 PAD Measurement Design (D2)

통제감(Dominance) 차원 적절성(D2-1)에 대해 E1이 찬성(○), 나머지 6인이 조건부 찬성(△)을 제시하였다. E1은 PAD의 3차원 측정이 기존 이모지 기반 평면적 측정의 한계를 극복한다고 강하게 지지하였다.

[E1] "PAD 같은 경우에는 쾌불쾌, 긴장감, 통제감 같이 상당히 입체적으로 실제 사용자의 감정을 볼 수 있겠다는 측면에서 적절한 것 같고, 특히 공공 서비스 같은 경우에 통제감(Dominance)을 보는 게, 그게 없으면 사실 이 연구가 좀 의미가 없지 않았나 싶을 정도로 상당히 맞는 것 같아요."

E6 또한 대안제가 없는 공공 서비스 맥락에서 통제감(Dominance)을 핵심 지표로 설정한 것에 강하게 동의하였다.

[E6] "통제감이라는 거를 핵심 지표로 설정하신 거는 듣고 보니까 진짜 그런 거예요. 공공은 대체제가 없는 거죠."

E2와 E3는 공통적으로 통제감(Dominance) 개념의 두 층위, 즉 공공 서비스의 독점 구조에서 비롯되는 구조적 통제감 부재와 서비스 이용 단계별로 변화하는 행위적 통제감을 본문에서 명확히 구분하여 서술할 것을 제안하였다. 이는 통제감(Dominance) 차원의 이론적 정교화를 위한 건설적 제안으로, PAD 3차원 측정 설계 자체의 타당성을 부정하는 것은 아니다. 이 두 층위는 2.2절에서 논의한 공공 서비스의 구조적 제약과 시간적 전개 특성에 각각 대응된다. 구조적 통제감(structural perceived control)은 이탈 불가능성과 대안 부재에서 비롯되어 서비스 이용 전반을 배경 조건으로 제약하며, 행위적 통제감(perceived situational control)은 각 이용 단계에서 과업 진행 상황에 따라 변화한다. 본 프레임워크의 PAD 측정은 이 중 행위적 통제감 층위를 측정 대상으로 삼으며, 이는 특정 내러티브 단계에서 사용자가 지각하는 과업 진행의 통제 가능성을 의미한다.

E4는 'Dominance의 통제감'이라는 한국어 레이블의 수동태적 뉘앙스를 지적하며 '선택 가능성' 또는 '자기 조절감'으로 재명명할 것을 제안하였고, E5는 VAS 양극 레이블('잘해냈다' / '불안하다')이 의미미분법(Semantic Differential)[34]상 반의어 요건을 충족하지 않음을 지적하며 재설계를 요청하였다. E4·E5·E7은 각각 독립적으로 공공 서비스 맥락에서 각성(Arousal)과 통제감(Dominance)에 비해 즐거움(Pleasure)의 측정 중요도가 상대적으로 낮다고 평가하였다. 특히 E7은 Freytag의 감정 고조 흐름과 가장 자연스럽게 대응되는 차원은 통제감(Dominance)보다 각성(Arousal)이라고 제안하였다.

측정 시점 충분성(D2-2)에 대한 전문가 의견은 상반된 방향으로 나뉘었다. E2·E6은 독립적으로 개인차 보정을 위한 발단 단계 기저선(baseline) 측정 추가를 권유하였다. E3도 이에 동의하면서, 특히 고령 사용자의 경우 서비스 진입 초기부터 인지 부하가 높을 수 있어 기저선 측정

Table 5. Summary of Expert Validation Judgments by Category (E1-E7)

Category	Sub-category	E1	E2	E3	E4	E5	E6	E7	Overall
D1, Freytag Validity	D1-1. Stage mapping	○	△	△	△	△	△	△	△ Conditional
	D1-2. Journey abstraction	△	△	△	○	△	△	△	△ Conditional
D2, PAD Measurement	D2-1. Dominance dimension	○	△	△	△	△	△	△	△ Conditional
	D2-2. Measurement timing	○	△	△	△	△	△	△	△ Conditional
	D2-3. Measurement Method	△	△	△	△	△	△	△	△ Conditional
D3, Heatmap Integration	D3-1. Triangulation	○	△	○	○	○	○	△	○ Support
	D3-2. Heatmap Indicators	○	○	N/E	△	△	△	△	△ Conditional
D4, Practical Applicability	D4-1. Value vs. existing Methods	○	○	○	○	○	○	○	○ Support
	D4-2. Implementation barriers	-	-	-	-	-	-	-	-

* ○ Support, △ Conditional support, - Not applicable(open-ended), N/E(Not evaluable)

의 필요성이 더욱 크다고 보았다. 반면 E5는 3회 측정이 현실적으로 과다하다고 보고 절정 후와 종료 시 각 1회로 줄일 것을 권장하였으며, 상승 초기 팝업이 오히려 부정적 감정을 유발할 위험을 지적하고 대규모 표본 환경에서는 측정 시점을 참여자별로 랜덤 분산 노출하는 방식을 대안으로 제안하였다. E7은 5단계 전체를 측정 후 사후적으로 불필요한 시점을 제거하는 것이 방법론적으로 타당하며, 사전에 특정 단계를 임의로 제외할 경우 측정 시점 선택의 자의성으로 인한 데이터 편향 문제를 제기하였다. E3는 측정 횟수보다 각 단계의 경계를 어떻게 정의할 것인가가 더 선행되어야 할 과제라는 점을 강조하였다. 이러한 의견 분기는 측정 시점 최적화가 후속 실험을 통한 실증 검증이 필요한 설계 변수임을 시사한다.

측정 방식 실행 가능성(D2-3)에 대해서는 7인 전문가 모두 팝업 형식의 개선 필요성을 공통으로 제기하였다. E4는 화면 측면 고정 배치를, E5는 드로어(drawer) 방식을, E6은 화면 내 고정 위치 배치를 각각 제안하였다. E3는 팝업이 통제감(Dominance) 측정값 자체에 영향을 줄 수 있다는 방법론적 문제를 명시적으로 지적하였다.

[E3] "통제감에 되게 영향을 많이 미치는 게 팝업인 것 같아서. 무시할 수 있는 것과 아닌 거랑은 큰 차이가 있어서."

그러나 E7은 이용 중 실시간 감정 상태를 포착하기 위해서는 측정 개입 자체가 방법론적으로 불가피하다고 평가하였다. 사용자의 즉시적 감정 반응은 사후 회고 방식으로는 포착할 수 없으며, 팝업 방식 외에 현실적인 대안이 존재하지 않는다는 것이다. 다만 참여자 부담을 최소화하기 위해 PAD 3차원을 매 측정 시점마다 동시에 묻는 대신 시점별로 1차원씩 단축 측정하는 방식을 대안으로 제안하였다. 컬러 적용에 관해서는 연구 환경에서는 모노톤(E1·E2·E4), 실서비스 이용자 대상에서는 컬러(E3·E5·E6)로 의견이 갈려, 적용 맥락에 따른 인터페이스 분기 설계의 필요성이 확인되었다. E6은 이모지와 텍스트 레이블을 병기하는 방식을 추가로 권장하였으며, 고령 사용자의 VAS 조작 어려움도 보완이 필요한 접근성 과제로 지적하였다.

4.4 Heatmap Behavioral Data Integration (D3)

PAD-히트맵 삼각검증 가치(D3-1)는 9개 하위 범주 중 가장 강한 합의를 보인 영역으로, E1·E4·E5·E6이 찬성(○), E2·E3·E7이 조건부 찬성(△)을 제시하였다. 전문가들

은 동일한 히트맵 패턴이 사용자의 감정 상태에 따라 전혀 다른 의미를 가질 수 있다는 점에서 PAD 데이터를 통한 삼각검증의 필요성을 일관되게 지지하였다. E1은 캐러셀 영역의 클릭 집중이 긍정인지 부정인지 히트맵만으로는 판단할 수 없는 실무 사례를 제시하였고, E4는 클릭률이 스트레스의 신호일 수 있다는 점을 실무 경험을 토대로 설명하였다.

[E4] "클릭률이 높다고 좋은 게 아니죠. 스트레스가 많은 구간이니까 다음 버튼을 누르는 거죠. 유입률보다 클릭률이 2~3배 높을 때는 굉장히 문제가 있는 구간이에요."

고용24 서비스 운영 경험을 보유한 E6은 실제 운영 데이터에서 동일한 문제를 확인한 사례를 제시하며 삼각검증의 차별적 가치를 강하게 지지하였다.

[E6] "고용24 민원 신청 유의사항 자격 확인 페이지에서 H2C 전환율을 보면 효율적으로 보이지만, 사용자 심층 인터뷰를 해보면 화면의 내용이 무슨 말인지 모르겠으니까 일단 눈에 띄는 CTA 버튼을 눌렀다는 거죠. 명확하게 차별화된 가치를 제공할 수 있다라고 판단합니다."

아이트래킹 연구 경험이 많은 E5는 온라인 환경에서 행동-의도 연계 해석이 기술적으로 불가능했던 한계를 이 방법이 극복한다고 평가하였다. E3는 주관 데이터(PAD)와 객관 데이터(히트맵)를 결합한다는 연구 목적과 시나리오의 정합성을 인정하였다.

[E3] "그냥 결과치만 봤을 때에는 전체 과정이나 사용자의 감정을 봤을 때 다르게 해석할 수 있기 때문에, 주관과 객관적 데이터를 같이 믹스해서 하신 부분이 목적과 맞는 시나리오고 적합하다고 생각합니다."

E1은 PAD-히트맵 불일치 데이터를 측정 오류가 아닌 후속 연구 질문의 발생 지점으로 재해석하며, FGI 등 심층 방법으로 연계할 가능성을 제안하였다. E6은 불일치 유형을 사전에 분류하는 체계를 수립할 것을 추가로 권고하였다. E7은 스크롤 데이터의 경우 목표 메뉴 탐색 과정과 PAD의 긴장도 데이터가 의미 있게 연계될 수 있다고 평가하였으나, 클릭 결과는 다음 페이지에 나타나는 경우가 많아 현재 페이지의 PAD 값과 직접 연계하는 데 한계가 있다고 지적하였다.

히트맵 지표 적절성(D3-2)에서는 6개 지표 선정 자체의

기본 타당성을 전문가 전원이 인정하였다. E5는 H2C(Hover to Click) 지표가 과업 유형에 따라 긍정 또는 부정 해석이 상반될 수 있어 과업 맥락과 연동한 해석 기준 차별화가 필요하다고 지적하였다. E6은 현재 지표가 PC 웹 환경에 한정되어 있어 모바일 환경 지표와 Dead click이나 Rage click 등 보조 지표의 추가가 필요하다고 제안하였다. E7은 스크롤 지표 해석 시 해당 과업의 중요도를 반영해야 하며 클릭과 스크롤의 해석 기준을 차별화할 것을 권고하였다. E3는 PAD 차원과 히트맵 지표 간의 연계 효과를 실험 전 단계에서는 판단하기 어렵다는 이유로 평가를 유보(N/E)하였다.

4.5 Practical Applicability (D4)

기존 방법 대비 실무적 가치(D4-1)에 대해 7인 전문가 전원이 찬성(○)을 제시하였다. 이는 9개 하위 범주 중 PAD-히트맵 삼각검증 가치(D3-1)와 함께 가장 높은 합의 수준을 기록한 결과로, 프레임워크의 실무적 방향성 자체에 대한 전문가 집단의 공통된 인정을 반영한다. E1은 현행 공공 서비스 평가가 전문가 휴리스틱·VOC 등 단면적 방법에 의존한다는 구조적 한계를 지적하며 프레임워크의 차별적 가치를 인정하였고, E2는 정서적 측면 측정이 공공·민간 서비스 평가 전반에서 항상 취약했던 문제를 이 프레임워크가 해소할 수 있다고 평가하였다. E3는 기존 평가 방식에서 연령 등 사용자 유형별 차이가 하나의 집계값으로 축약됨으로써 상세화가 불가능했던 문제를 지적하며, 이 프레임워크가 그간 평가에서 외면되어 온 사용자와 케이스를 발굴할 수 있는 환경을 제공한다는 점에서 가치 있다고 평가하였다.

[E3] "좀 외면됐던 사용자들이나 외면되어 있던 케이스를 좀 더 발굴할 수 있는 환경을 준다는 점에서 가치 있다고 생각합니다."

E5는 소수의 참여자 대상 실험에 의존하던 기존 사용성 평가의 한계를 넘어 실제 서비스에서 더 많은 사용자 데이터를 수집할 수 있다는 점을 추가 가치로 강조하였다. E6은 본 프레임워크가 기존 평가 방식으로는 불가능했던 감정과 행동 통합 진단을 정량적 수치로 표현할 수 있다는 점에서 진단의 객관성과 설득력이 높아지고, 문제 구간을 더 세밀하게 식별할 수 있다고 평가하였다.

[E6] "신청서 작성 단계에서 통제감이 수치로서 82에서 55로 급락한다. 레이저 클릭이 몇 배 증가한다. 이런 식으

로 정량적 수치로 이야기할 수 있는, 진단의 해상도가 올라갈 수 있을 것 같아요."

E7은 공공 서비스 현장에서 사용자 리서치 자체가 부재한 현실을 지적하며, 이 프레임워크가 그 공백을 보완할 수 있다는 점에서 실무적 기여가 크다고 평가하였다.

구현 장벽 및 개선 우선순위(D4-2)의 개방형 수집 결과는 다음과 같이 수렴되었다. 첫째, E4는 공공기관 비전문가의 실제 도입을 위해 분석 절차를 자동화한 수식 내장 템플릿 수준의 실용화 도구 개발이 선행되어야 한다고 제안하였다. 둘째, E5는 과업 방해 최소화 핵심 구현 과제이며, 현재 제안된 어떤 UI 방식도 이를 완전히 해결하지 못하고 있다고 지적하였다. 셋째, E2는 VAS 응답의 개인간 시작점 편차를 보정하기 위한 통계적 분석 방법론을 사전에 수립해야 한다고 제안하였다. 넷째, E6은 정부24·고용24 등 실제 공공 서비스에 측정 스크립트를 삽입하기 위해서는 보안 기관의 행정적 승인 절차를 거쳐야 한다는 현장 고유의 장벽을 신규로 지적하였으며, 분석자 간 결과 편차를 방지하기 위한 표준화 도구의 필요성도 함께 제안하였다. 다섯째, E3는 Freytag 구조의 서비스별 표준화가 가장 큰 구현 장벽이라고 보고, 전체 서비스에 일괄 적용하기보다 이미 문제가 인지된 특정 영역에 집중하여 적용 범위를 좁히는 단계적 도입 전략이 현실적이라고 제안하였다. 여섯째, E7은 이 프레임워크의 지속적 실용화 가능성에 의문을 제기하며, 복잡성을 검증하고 현실 적합성을 확인하기 위한 파일럿 연구의 필요성을 강조하였다.

4.6 Summary and Discussion

전문가 검증 결과를 종합하면, 본 프레임워크는 4개 평가 영역 9개 하위 범주 전반에서 개념적 타당성을 인정받았다. 9개 하위 범주 중 D4-1(기존 방법 대비 실무적 가치)에서 7인 전문가 전원이 찬성(○)을 제시하였고, D3-1(PAD-히트맵 삼각검증 가치)에서도 4인이 찬성(○), 3인이 조건부 찬성(△)으로 가장 강한 합의가 형성되었다. 이는 삼각검증을 통한 해석 모호성 극복이라는 프레임워크의 핵심 전제가 실무 전문가들에게 설득력 있게 수용되었음을 보여준다. 실무 경험을 근거로 한 복수 전문가의 독립적 지지는 본 개념적 타당성 판단의 신뢰성(trustworthiness)[41]을 뒷받침한다. Freytag 5단계 구조의 적용 타당성(D1)에 대해서는 공통 구조의 개념적 타당성을 전원이 인정하면서도, 조건부 찬성이 다수를 차지하였다. Freytag 구조의 중첩·반복 패턴은 복수의 전문가가 독립적으로 확인하였으며, 이는 단일 선형 피라미드가 아닌 복수의 긴장 주기가

공공 서비스 여정 내에 나타날 수 있음을 시사한다.

전문가 검증을 통해 다음과 같은 개선 과제가 도출되었다. 첫째, 통제감(Dominance) 차원의 한국어 레이블 및 VAS 양극 레이블 재설계가 필요하다(E4, E5). 둘째, 신청형 서비스에서의 하강-결말 경계 모호화 문제(E2, E5)와 Freytag 구조 표준화의 어려움(E3)은 각 단계의 조작적 정의를 후속 실험 단계에서 실증적으로 구체화해야 할 과제를 시사한다. 셋째, 발단 단계 기저선 측정의 필요성이 E2·E6에 의해 독립적으로 제기되었고, E3도 고령 사용자의 초기 인지부하를 근거로 이에 동의하였으며, 측정 시점 최적화 전반은 후속 실험을 통한 실증 검증이 필요한 설계 변수로 확인되었다. 넷째, PAD-히트맵 해석 기준표의 케이스 확장은 실험 데이터 확보 후 실증 기반으로 구축해야 할 후속 과제이며, 불일치 유형에 대한 사전 분류 체계 수립이 선행될 필요가 있다(E6). 다섯째, 본 프레임워크는 I장에서 제시한 회고적 편향 극복이라는 연구 목적을 달성하기 위해 이용 중 실시간 자기보고 측정을 채택하였으며, 팝업 형식의 측정 인터페이스는 이를 구현하기 위한 수단이다. 다만 이러한 측정 개입이 Dominance 측정값 자체에 영향을 미칠 수 있다는 반응성(reactivity) 문제가 제기되었으며(E3), 고정 배치 또는 드로어 방식 등 대안적 UI 설계의 실험적 검토가 필요하다. 이 영향의 정량적 확인과 대안 UI의 통제된 비교는 후속 실험 연구를 통해 실증적으로 수행될 것이다. 여섯째, 실제 공공 서비스 환경에 측정 도구를 적용하기 위해서는 보안 기관의 행정적 승인 절차라는 현장 고유의 구조적 제약이 존재하며(E6), 이는 본 연구에서 직접 해소할 수 있는 성격의 과제가 아니다. 후속 연구에서는 프로토타입 기반 사용자 평가를 통해 타당성을 먼저 확보하고 그 결과를 공공 서비스 적용의 근거로 제시하는 방향을 취한다. 일곱째, 프레임워크의 현실 적합성을 검증하기 위한 파일럿 연구의 필요성이 복수의 전문가에 의해 제기되었다(E5, E7).

이상의 결과는 본 프레임워크의 개념적 타당성을 확인하는 동시에, 후속 실험 연구의 설계 방향에 구체적 근거를 제공한다.

V. Conclusions

본 연구는 공공 디지털 서비스 UX 평가의 구조적 한계, 즉 사후 단일 시점 측정과 피크-엔드 법칙으로 인한 경험 왜곡을 극복하기 위해 Freytag의 5단계 구조, PAD 3차원 모델, 히트맵 행동 데이터를 통합한 내러티브 기반 UX 평

가 프레임워크를 제안하였다. Freytag 구조와 서비스 여정 단계 간의 대응은 사전에 확정된 매핑이 아닌 잠정적 조작화 가설로 설정되었으며, UI/UX 전문가 7인을 대상으로 한 반구조화 인터뷰와 Elo와 Kyngäs[14]의 연역적 질적 내용분석을 통해 개념적 타당성을 검증한 결과, 9개 하위 범주 전반에서 타당성이 인정되었다. 본 검증은 프레임워크의 개념적 타당성을 확인하는 탐색적 연구(exploratory study)에 해당하며, 측정 정확도나 편향 감소 효과에 대한 실증적 주장을 포함하지 않는다. 도출된 개선 과제들은 프레임워크의 근본 전제에 대한 부정이 아니라 후속 실험 연구에서 실증적으로 정교화해야 할 설계 변수로 수렴되었다.

이러한 검증 결과를 바탕으로, 본 연구는 UX 평가를 결과 중심에서 과정 중심으로 전환하려는 시도로, 사용자 경험을 시간적·서사적 구조 속에서 재해석한다는 점에서 학술적 의의를 가진다. 본 연구의 학술적 기여는 다음과 같다. 첫째, 내러티브 이론을 UX 평가 방법론에 접목한 이론적 토대를 마련하였다. Freytag 구조를 공공 디지털 서비스 사용자 여정의 분석 틀로 재해석하고, 단계별 감정 긴장의 전환점을 기준으로 서비스 구간을 구분하는 잠정적 조작화 가설을 제시하였다. 특히 전문가 검증 과정에서 복수의 전문가가 독립적으로 Freytag 구조의 중첩·반복 패턴을 확인함으로써, 단일 선형 피라미드가 아닌 복수의 긴장 주기가 공공 서비스 여정 내에 반복적으로 나타날 수 있다는 구조적 특성이 실무 맥락에서 뒷받침되었다. 둘째, 공공 서비스 맥락에 특화된 PAD 차원 적용 방향을 이론적으로 제안하였다. 이탈이 불가능한 공공 서비스에서 통제감(Dominance)은 원하는 대로 진행되지 않는 상황에서의 무력감을 측정하는 핵심 차원으로 기능하며, 기존 PAD 연구에서 세 차원을 동등하게 다루었던 것[13]과 달리, 본 연구는 전문가 검증을 통해 공공 서비스 UX 평가에서는 각성(Arousal)과 통제감(Dominance)이 즐거움(Pleasure)에 비해 상대적으로 더 핵심적인 측정 차원임을 확인하였다. 셋째, 행동 데이터와 감정 데이터의 삼각검증 방법론을 제안하였다. 히트맵 단독 해석의 의미적 모호성을 PAD 측정 데이터로 보완하는 해석 프레임워크를 구체화함으로써, 기존 분석 도구의 데이터를 새로운 관점에서 재해석할 수 있는 실용적 방법론을 제시하였다.

이러한 학술적 기여를 바탕으로, 본 연구의 결과는 공공 디지털 서비스의 실무적·정책적 개선에 다음과 같이 활용될 수 있다. 현행 디지털 정부서비스 UI/UX 가이드라인[3]이 서비스 설계 원칙과 UI 컴포넌트 표준을 제공하는 사전 규범의 역할을 수행한다면, 본 프레임워크는 서비스

운영 이후 단계별 사용자 경험을 진단하는 사후 평가 도구로 상호 보완적으로 기능할 수 있다. 전문가 검증에서 확인된 바와 같이 신청서 작성 단계의 통제감 급락이나 레이저 클릭 증가와 같은 문제 구간을 정량적 수치로 식별할 수 있다는 점은, 가이드라인의 일반 원칙을 특정 서비스의 구체적 개선 과제로 전환하는 연결 지점을 제공하며, UI 개선 우선순위를 설정하는 실증적 근거로 활용될 수 있다. 정책적 측면에서, 본 프레임워크의 단계별 진단 결과는 공공기관이 추진하는 디자인 평가제[1]의 평가 기준을 UI/UX 측면에서 보완하는 방향으로 반영될 수 있으며, 기존 평가에서 하나의 집계값으로 축약되어 외면되어 온 사용자 유형별 경험 차이를 가시화하는 데 기여할 수 있다. 다만 전문가 검증 결과, 실제 공공 서비스 환경에의 적용을 위해서는 분석 절차를 자동화한 실용화 도구의 개발과 보안 기관의 행정적 승인 절차라는 현실적 장벽이 선행적으로 해소되어야 하며, 전체 서비스에 일괄 적용하기보다 이미 문제가 인지된 특정 서비스 구간에 우선 적용하는 단계적 도입 전략이 현실적이다.

이상의 학술적·실무적 기여에도 불구하고 본 연구는 다음의 한계를 가진다. 본 연구는 프레임워크의 개념적 타당성을 전문가 판단을 통해 검증하는 데 초점을 두며, 실제 측정 환경에서의 수렴 타당도 및 예측 타당도 검증은 본 연구의 범위를 벗어난다. 실제 서비스에서 단계 경계를 식별하는 구체적 방법은 실증적으로 검증되지 않았으며, 특히 오류 발생·재시도·반복 탐색 등 서비스 이용 중 발생하는 피드백 과정이 Climax와 Falling Action 중 어느 구간에 귀속되는지는 현 단계에서 명확히 특정하기 어렵다는 한계가 있다. 이 모호성은 마이크로 이벤트 기반 잠정 조작화와 실증 검증을 통해 후속 연구에서 해소할 설계 변수로 설정한다. PAD 측정 도구의 VAS 레이블과 통제감(Dominance) 차원 한국어 용어의 언어적·문화적 적절성, 그리고 팝업 방식 인터페이스가 통제감(Dominance) 측정값 자체에 영향을 미칠 수 있다는 방법론적 문제 역시 실증적으로 검토되지 않은 과제로 남는다.

위의 한계를 극복하기 위한 후속 연구에서는 본 프레임워크를 정부24 민원 신청 및 정책정보 확인 과업을 중심으로 구성된 프로토타입을 제작하고, 공공 서비스 이용 경험을 가진 시민을 대상으로 사용자 평가를 통해 실증적으로 검증한다. 이 과정에서 전문가 검증을 통해 도출된 측정 설계 개선 사항을 반영한 수정 프로토콜을 적용하며, Freytag 구조의 단계 경계와 절정 위치가 어떻게 달라지는지를 실증 데이터를 기반으로 비교 분석하고, 비전문가도 활용 가능한 서비스 패턴 분류 가이드 개발로 이어지는

실용화 방향을 함께 검토한다. 나아가 본 프레임워크는 이탤이 구조적으로 제한되거나 높은 불안을 수반하는 의료, 금융, 법률 등의 고관여 디지털 서비스 UX 평가에도 확장 적용될 수 있는 잠재성을 가진다. 이와 같은 맥락 확장 연구를 통해 내러티브 기반 UX 평가 방법론의 일반화 가능성을 검토하는 것을 장기적 연구 방향으로 설정하고자 한다.

REFERENCES

- [1] Digital Platform Government Implementation Plan, <https://nsp.nanet.go.kr/plan/subject/detail.do?nationalPlanControlNo=PLAN000037879>
- [2] Digital Platform Government Realization Plan, <https://nsp.nanet.go.kr/plan/subject/detail.do?nationalPlanControlNo=PLAN0000038780>
- [3] Digital Government Service UI/UX Guidelines and Design System(KRDS), <https://www.krds.go.kr>
- [4] 2025 Public Web/App Quality Management and UI/UX National Evaluation Briefing Materials, https://www.mois.go.kr/irt/bbs/type001/commonSelectBoardArticle.do?bbsId=BBSMSTR_000000000045&nttlId=120503
- [5] M. Hassenzahl, and N. Tractinsky, "User Experience - A Research Agenda," *Behaviour and Information Technology*, Vol. 25, No. 2, pp. 91-97, March 2006. DOI: 10.1080/01449290500330331
- [6] K. N. Lemon, and P. C. Verhoef, "Understanding Customer Experience Throughout the Customer Journey," *Journal of Marketing*, Vol. 80, No. 6, pp. 69-96, November 2016. DOI: 10.1509/jm.15.0420
- [7] D. A. Redelmeier, and D. Kahneman, "Patients' Memories of Painful Medical Treatments: Real-time and Retrospective Evaluations of Two Minimally Invasive Procedures," *Pain*, Vol. 66, No. 1, pp. 3-8, July 1996. DOI: 10.1016/0304-3959(96)02994-6
- [8] D. Moynihan, P. Herd, and H. Harvey, "Administrative Burden: Learning, Psychological, and Compliance Costs in Citizen-State Interactions," *Journal of Public Administration Research and Theory*, Vol. 25, No. 1, pp. 43-69, January 2015. DOI: 10.1093/jopart/muu0
- [9] P. Herd, and D. P. Moynihan, "Administrative Burden: Policymaking by Other Means," Russell Sage Foundation, pp. 1-360, 2018.
- [10] J. Bruner, "The Narrative Construction of Reality," *Critical Inquiry*, Vol. 18, No. 1, pp. 1-21, October 1991. DOI: 10.1086/448619
- [11] G. Freytag, "Die Technik des Dramas," S. Hirzel, pp. 1-303, 1863.
- [12] L. Yang, X. Xu, X. Lan, Z. Liu, S. Guo, Y. Shi, H. Qu, and N. Cao, "A Design Space for Applying the Freytag's Pyramid Structure to Data Stories," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 28, No. 1, pp. 922-932, January

2022. DOI: 10.1109/TVCG.2021.3114774
- [13] A. Mehrabian, and J. A. Russell, "An Approach to Environmental Psychology," The MIT Press, pp. 1-216, 1974.
- [14] S. Elo, and H. Kyngäs, "The Qualitative Content Analysis Process," *Journal of Advanced Nursing*, Vol. 62, No. 1, pp. 107-115, April 2008. DOI: 10.1111/j.1365-2648.2007.04569.x
- [15] D. Kahneman, "Thinking, Fast and Slow," Farrar, Straus and Giroux, pp. 1-499, 2011.
- [16] A. P. O. S. Vermeeren, E. L. C. Law, V. Roto, M. Obrist, J. Hoonhout, and K. Vaananen-Vainio-Mattila, "User Experience Evaluation Methods: Current State and Development Needs," *Proceedings of the 6th Nordic Conference on Human-Computer Interaction (NordCHI '10)*, pp. 521-530, Reykjavik, Iceland, October 2010. DOI: 10.1145/1868914.1868973
- [17] D. Kahneman, and A. Tversky, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, Vol. 47, No. 2, pp. 263-291, March 1979. DOI: 10.2307/1914185
- [18] R. C. Schank, and R. P. Abelson, "Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures," L. Erlbaum Associates, pp. 1-248, 1977.
- [19] J. Sweller, P. Ayres, and S. Kalyuga, "Cognitive Load Theory," Springer, pp. 1-282, 2011. DOI: 10.1007/978-1-4419-8126-4
- [20] O. James, and P. John, "Testing Hirschman's Exit, Voice, and Loyalty Model: Citizen and Provider Responses to Decline in Public Health Services," *International Public Management Journal*, Vol. 24, No. 3, pp. 378-393, May 2021. DOI: 10.1080/10967494.2021.1878314
- [21] J. Bruner, "Actual Minds, Possible Worlds," Harvard University Press, pp. 1-201, 1986.
- [22] P. Ricoeur, "Time and Narrative (Vol. 1)," University of Chicago Press, pp. 1-274, 1984.
- [23] R. Adaval, and R. S. Wyer Jr., "The Role of Narratives in Consumer Information Processing," *Journal of Consumer Psychology*, Vol. 7, No. 3, pp. 207-245, July 1998. DOI: 10.1207/s15327663jcp0703_01
- [24] N. Pennington, and R. Hastie, "Explaining the Evidence: Tests of the Story Model for Juror Decision Making," *Journal of Personality and Social Psychology*, Vol. 62, No. 2, pp. 189-206, February 1992. DOI: 10.1037/0022-3514.62.2.189
- [25] S. Grimaldi, "Design for Narrative Experience in Product Interactions," Ph.D. thesis, University of the Arts London, pp. 1-180, 2018.
- [26] M. Azarova, "Two Tips for Better UX Storytelling," <https://www.nngroup.com/articles/two-tips-better-ux-storytelling>
- [27] A. J. Reagan, L. Mitchell, D. Kiley, C. M. Danforth, and P. S. Dodds, "The Emotional Arcs of Stories Are Dominated by Six Basic Shapes," *EPJ Data Science*, Vol. 5, No. 1, pp. 1-22, 2016. DOI: 10.1140/epjds/s13688-016-0093-1
- [28] R. L. Boyd, K. G. Blackburn, and J. W. Pennebaker, "The Narrative Arc: Revealing Core Narrative Structures through Text Analysis," *Science Advances*, Vol. 6, No. 32, pp. 1-9, eaba2196, August 2020. DOI: 10.1126/sciadv.aba2196
- [29] M.-L. C. R. Schmidt, and J. R. Winkler, "Tracking Emotional Shifts During Story Reception: The Relationship Between Narrative Structure and Affective Responses," *Scientific Study of Literature*, Vol. 12, No. 1, pp. 17-39, 2023. DOI: 10.61645/ssol.177
- [30] P. Ekman, "An Argument for Basic Emotions," *Cognition and Emotion*, Vol. 6, No. 3-4, pp. 169-200, 1992. DOI: 10.1080/02699939208411068
- [31] R. Plutchik, "Emotion: A Psychoevolutionary Synthesis," Harper & Row, pp. 1-440, 1980.
- [32] L. F. Barrett, "Are Emotions Natural Kinds?," *Perspectives on Psychological Science*, Vol. 1, No. 1, pp. 28-58, 2006. DOI: 10.1111/j.1745-6916.2006.00003.x
- [33] J. A. Russell, "A Circumplex Model of Affect," *Journal of Personality and Social Psychology*, Vol. 39, No. 6, pp. 1161-1178, 1980. DOI: 10.1037/h0077714
- [34] M. M. Bradley, and P. J. Lang, "Measuring Emotion: The Self-Assessment Manikin and the Semantic Differential," *Journal of Behavior Therapy and Experimental Psychiatry*, Vol. 25, No. 1, pp. 49-59, 1994. DOI: 10.1016/0005-7916(94)90063-9
- [35] U.-D. Reips, and F. Funke, "Interval-level Measurement with Visual Analogue Scales in Internet-based Research: VAS Generator," *Behavior Research Methods*, Vol. 40, No. 3, pp. 699-704, 2008. DOI: 10.3758/BRM.40.3.699
- [36] K. Hornbæk, "Current Practice in Measuring Usability: Challenges to Usability Studies and Research," *International Journal of Human-Computer Studies*, Vol. 64, No. 2, pp. 79-102, 2006. DOI: 10.1016/j.ijhcs.2005.06.002
- [37] I. Pettersson, F. Lachner, A. K. Frison, A. Riener, and A. Butz, "A Bermuda Triangle? A Review of Method Application and Triangulation in User Experience Evaluation," *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '18)*, Paper No. 461, pp. 1-16, Montreal, Canada, April 2018. DOI: 10.1145/3173574.3174035
- [38] M. Csikszentmihalyi, and R. Larson, "Validity and Reliability of the Experience-Sampling Method," *Journal of Nervous and Mental Disease*, Vol. 175, No. 9, pp. 526-536, September 1987. DOI: 10.1097/00005053-198709000-00004
- [39] J. Nielsen, "Heuristic Evaluation," in *Usability Inspection Methods*, John Wiley & Sons, pp. 25-62, 1994.
- [40] G. Guest, A. Bunce, and L. Johnson, "How Many Interviews Are Enough? An Experiment with Data Saturation and Variability," *Field Methods*, Vol. 18, No. 1, pp. 59-82, February 2006. DOI: 10.1177/1525822X05279903
- [41] Y. S. Lincoln, and E. G. Guba, "Naturalistic Inquiry," Sage Publications, pp. 1-416, 1985.

Authors



Sunyoung Jang received her Master of Business Administration(MBA) in e-Business from the Graduate School of Public Policy and Industry at Sookmyung Women's University in 2013.

She is currently a Ph.D. candidate in the Department of Technology and Innovation Management at Hanyang University. She serves as the CEO of SNC Lab Co., Ltd., a consulting firm specializing in digital accessibility and UX/UI. She contributed to the development of the Digital Government Service UI/UX Guidelines for the Ministry of the Interior and Safety and serves as a member of the Korean Standards (KS) committee on user accessibility. Her research interests include data-driven UX design evaluation and inclusive design for the elderly and persons with disabilities.



Hokyoung Ryu received his Master's degree in Industrial Engineering from the Korea Advanced Institute of Science and Technology (KAIST) in 1999 and his Ph.D. in Psychology from the University of York,

UK, in 2014. He is currently a Professor in the Department of Technology and Innovation Management at Hanyang University. His research interests include cognitive information design, new product development, new service development, learning technology, game-based learning, and medical HCI.