

## Operationalization and Detection Verification of a Narrative-Based UX Evaluation Framework: A Multiple-Case Analysis of a Public Digital Service Task

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]

In public digital services, task completion alone cannot sufficiently explain the changes in users' states and perceived control (Dominance) that occur during the process of use. The purpose of this study is to verify the detection function of the operationalized analysis procedure of the narrative-based UX evaluation framework proposed in prior research by applying it to an actual public digital service task. Five participants performed the same task on a prototype that reproduced the resident registration card reissuance process of Gov24, and PAD self-reports anchored to 27 micro-events, behavioral data, and retrospective think-aloud (RTA) accounts were collected and compared. The stage assignments of two independent coders agreed on 26 events (Cohen's  $\kappa = .948$ ), free narratives were classified into two cases of construction success, two of partial construction, and one of construction failure, and 16 convergent, 3 partially convergent, and 1 divergent case appeared across 20 analysis units. These results show that the objective task structure and the subjective experience structure constructed by users may not coincide, and that behavioral completion of a task does not guarantee the recovery of perceived situational control (Dominance). This study is meaningful in that it operationalized narrative into analyzable UX data on a shared micro-event timeline. These results constitute an initial verification of the framework's applicability in five cases, and its general validity will be examined with larger samples in a planned follow-up study.

▶ **Key words:** Narrative-based UX evaluation, Freytag's Pyramid, PAD model, Micro-event, 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. 09. 09, Revised: 2026. 09. 22, Accepted: 2026. 10. 08.

## [요 약]

공공 디지털 서비스에서는 과업의 완료 여부만으로 이용자가 이용 과정에서 경험하는 상태 변화와 통제감(Dominance)의 변화를 충분히 설명하기 어렵다. 본 연구의 목적은 선행 연구에서 제안한 내러티브 기반 UX 평가 프레임워크를 실제 공공 디지털 서비스 과업에 적용하여 조작화된 분석 절차의 탐지 기능을 검증하는 것이다. 참여자 5인이 정부24 주민등록증 재발급 과정을 모사한 프로토타입에서 동일한 과업을 수행하였고, 27개 마이크로 이벤트에 정박한 PAD 자기보고와 행동 자료, 회고적 발화(RTA)를 수집하여 대조하였다. 분석 결과 두 독립 코더의 Freytag 단계 판정은 26개 이벤트에서 일치하였고(Cohen's  $\kappa = .948$ ), 참여자 5인의 자유 서사는 구성 성공 2사례, 부분 구성 2사례, 구성 실패 1사례로 구분되었으며, 20개 분석 단위에서 수렴 16건, 부분 수렴 3건, 발산 1건이 나타났다. 이는 객관적 과업 구조와 이용자가 구성한 주관적 경험 구조가 일치하지 않을 수 있고, 과업의 행동적 완료가 행위적 통제감(perceived situational control, Dominance)의 회복을 보장하지 않음을 보여준다. 본 연구는 내러티브를 마이크로 이벤트 시간축 위에서 분석 가능한 UX 자료로 조작화하였다는 의의가 있다. 이 결과는 5인 사례에서 프레임워크의 실제 적용 가능성을 확인한 초기 검증이며, 프레임워크의 일반적 타당성은 후속 연구에서 대규모 표본으로 검증할 예정이다.

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

## I. Introduction

인간은 익숙한 상황에서 사건이 발생할 때 이를 하나씩 개별적으로 처리하기보다 과거 경험으로 형성된 지식 구조를 활용하여 상황을 이해한다. 이러한 지식 구조 중 하나가 인지적 스크립트(cognitive script)이다. Schank와 Abelson[1]은 스크립트를 특정 상황에서 일반적으로 예상되는 사건의 순서와 관계를 나타내는 구조로 설명하였다. 인간은 이러한 스크립트를 바탕으로 명시되지 않은 사건을 추론하고 이후의 전개를 예상할 수 있다[1]. 따라서 특정 과업에 대한 경험이 충분히 축적되지 않은 경우에는 활용할 수 있는 스크립트가 제한되며, 과업이 어떤 순서로 진행되고 다음에 어떤 사건이 발생할지를 예측하기 어려워지는 경향이 있다[1].

공공서비스 이용에서는 과업 이해의 어려움과 함께 정부와 상호작용하는 과정에서 발생하는 행정 부담(administrative burden)도 고려할 필요가 있다. 행정 부담은 시민이 공공서비스를 이용하면서 경험하는 학습 비용(learning costs), 순응 비용(compliance costs), 심리적 비용(psychological costs)으로 구성된다[2]. 이 부담은 단순한 불편에 그치지 않으며, 정책의 설계와 집행 방식에 따라 서비스나 권리에 접근하는 과정 자체를 가로막을 수 있다[2][3].

Bruner[4]는 인간이 시간에 따라 발생하는 사건을 단순히 순서대로 인식하는 것이 아니라, 사건 사이의 관계와

의미를 구성함으로써 경험을 하나의 내러티브로 이해한다고 보았다. 특히 그는 사건의 전개가 사람들이 일반적으로 예상하는 전형적인 흐름(canonical script)에서 벗어날 때 내러티브적 해석이 촉발된다고 지적하였다[4]. 즉 예상하던 사건의 흐름이 깨질 때 인간은 그 사건이 왜 발생했는지, 전체 경험에서 어떤 의미인지를 다시 구성한다[4].

McCarthy와 Wright[5]는 Bruner의 관점을 기술 사용 경험으로 확장하였다. 이들은 기술 경험의 구성적 가닥(compositional thread)을 경험의 개별 사건과 전체 경험 사이의 관계를 이해하는 과정으로 개념화하였다. 이용자는 기술을 사용하는 동안 무엇이 일어났는지, 현재 어디에 있는지, 다음에 무엇이 일어날지를 지속적으로 해석한다[5]. 또한 전개 중인 경험을 이해하기 위해 지금까지 발생한 사건과 앞으로 발생할 가능성이 있는 사건을 연결하며, 그 전개를 예측하기 어려운 경우에는 불안이 발생할 수 있다고 보았다[5].

내러티브 구성이라는 관점에서 공공 디지털 서비스 이용 경험은 개별 화면이나 특정 시점의 반응만으로 충분히 설명하기 어렵다. 이용자는 과업을 수행하는 동안 현재 위치와 이전에 발생한 사건, 이후 예상되는 사건을 지속적으로 연결하면서 전체 경험의 의미를 구성한다. 따라서 내러티브는 서비스 자체에 고정된 속성이라기보다 이용자가 과업의 전개를 해석하면서 구성하는 경험적 구조로 볼 수

있다. 같은 사건의 순서를 경험하더라도 그것을 하나의 완결된 이야기로 구성하는 방식이나 어떤 사건을 경험의 중심으로 인식하는지는 이용자마다 달라질 수 있다[4][5].

인지적 스크립트와 내러티브는 모두 시간에 따라 전개되는 사건의 관계를 조직한다는 점에서 구조가 비슷하다. 스크립트가 과거 경험을 바탕으로 예상되는 사건의 전개를 가능하게 한다면, 내러티브는 이용자가 실제로 발생한 사건을 서로 연결하여 현재의 경험을 이해하도록 돕는다. 따라서 특정 과업에 대한 스크립트가 충분히 형성되지 않은 이용자에게는 내러티브 구성, 곧 서비스 이용 과정에서 발생하는 사건을 일관된 흐름으로 연결하여 이해하는 과정이 중요하게 작용할 여지가 있다. 본 연구진의 선행 연구는 이러한 내러티브 구성 과정이 이용자가 지각하는 통제감(Dominance)의 유지와 붕괴로 연결될 수 있다는 이론적 전제를 제시하였다[6].

이러한 전제를 바탕으로 선행 연구에서는 Freytag의 5단계 구조(Freytag's Pyramid: 발단-상승-절정-하강-결말, 이하 Freytag 구조)를 공공 디지털 서비스의 사용자 여정에 적용하고, 각 단계에서 PAD(Pleasure-Arousal-Dominance) 자기보고와 히트맵 기반 행동 데이터를 함께 수집하여 삼각검증(triangulation)하는 내러티브 기반 UX 평가 프레임워크를 제안하였다[6]. 전문가 검증에서는 프레임워크의 개념적 타당성을 확인하였으며, 특히 의무적 공공서비스 맥락에서 통제감(Dominance)이 중요한 경험 지표가 될 가능성이 제기되었다[6]. 다만 선행 연구는 개념 수준에서 가능성을 검토한 탐색적 연구였기 때문에 실제 적용 과정에서 확인해야 할 몇 가지 방법론적 과제가 남아 있다. Freytag 구조의 단계 경계를 실제 과업에서 어떻게 조작적으로 정의할 것인지, 자기보고와 행동 데이터가 서로 다른 결과를 나타낼 때 이를 어떻게 해석할 것인지, 과업 중 반복 측정이 이용 경험 자체에 영향을 미치는 반응성을 어떻게 통제할 것인지에 대한 검증이 필요하다[6].

이에 본 연구는 선행 연구에서 제안한 내러티브 기반 UX 평가 프레임워크를 실제 이용자 과업에 적용하여 조작화된 분석 절차의 탐지 기능을 검증(verification)하는 다중 사례 연구로 수행하였다. 참여자 5인이 동일한 공공 디지털 서비스 과업을 수행하는 과정에서 프레임워크가 이용자의 내러티브 구성 방식과 개인차를 포착할 수 있는지, 그리고 서로 다른 자료 간의 일치와 불일치가 어떠한 의미인지를 확인하였다. 본 연구는 소규모 사례를 이용한 프레임워크 검증을 목적으로 하며, 대규모 표본을 통한 통계적 타당화(validation)는 후속 연구의 범위로 두었다. 이를 위해 정부24 이용 과정을 모사한 프로토타입 과업을 구성하고, 연구자가 사전에 핵심 교란 사건으로 가정한 지점에

사진 규격 오류와 입력 정보 소실을 배치하였다.

본 연구의 이론적 전제와 연구질문은 다음과 같이 대응한다. 인지적 스크립트가 충분히 형성되지 않은 과업에서 이용자는 발생한 사건을 경험 내러티브로 해석하며[1][4][5], 이러한 인지적·감정적 부담이 누적되는 공공 디지털 서비스에서는 기대한 내러티브의 구축과 그 방해가 통제감(Dominance)의 유지 혹은 붕괴로 이어질 수 있다[6]. Freytag 구조는 이러한 사건의 전개를 사례 간 비교가 가능한 공통 시간 구조로 조작화하는 틀이고(RQ1), PAD 모델의 Dominance 차원은 사건별 행위적 통제감을 측정하는 도구이며(RQ2), 이용자가 사건을 선택하고 연결하는 방식의 차이는 회고적 발화로 확인한다(RQ3). 다음의 세 연구질문과 연구 방법을 통해 단계 경계의 판정 가능성, 통제감(Dominance) 레이블의 해석, 동일 과업에서 나타나는 내러티브 구성의 개인차를 검토하였다.

RQ1. 마이크로 이벤트 단위로 조작화한 Freytag의 단계 경계는 독립 판정자 간에 일관되게 판정되는가?

RQ2. 행위적 통제감(perceived situational control)을 측정하기 위해 제시한 Dominance 레이블은 이용자에게 의도된 의미로 해석되는가?

RQ3. 동일한 과업과 동일한 결과를 경험하는 조건에서 이용자의 내러티브 구성은 어떻게 다르게 나타나는가?

## II. Theoretical Background

### 2.1 Framework and Methodological Challenges of the Prior Study

서비스 이용이 끝난 뒤 실시하는 단일 만족도 측정은 이용자에게 전 과정의 경험을 하나의 평가값으로 요약하도록 요구한다. 그러나 경험에 대한 회고 평가는 경험의 모든 순간을 동일하게 반영하지 않을 수 있다. Redelmeier와 Kahneman[7]은 의료 처치 과정에서 실시간으로 기록한 통증과 사후 평가를 비교하였다. 전체 경험에 대한 회고적 판단은 가장 강한 통증이 나타난 시점과 마지막 구간의 통증에 크게 영향을 받았으며, 이는 피크-엔드 법칙(Peak-End Rule)으로 알려져 있다. 이 결과는 의료 경험을 대상으로 한 것이지만, 기술 이용 경험 역시 시간에 따라 전개되는 사건의 연쇄로 구성되며[4][5], 공공 디지털 서비스의 이용 여정을 이러한 서사 구조로 분석할 수 있음은 선행 연구에서 전문가 검증을 거쳤대[6]. 따라서 회고 평가가 경험의 특정 시점에 편향된다는 점은 여러 단계를

거치는 서비스 이용 경험의 사후 평가에도 적용될 수 있다. 경험 종료 후 이루어지는 단일 평가만으로는 이용 과정에서 발생한 개별적인 어려움과 상태 변화를 충분히 포착하기 어려울 가능성이 있다.

본 연구진의 선행 연구[6]는 이러한 한계에 주목하여 측정을 경험 종료 후 실시하는 대신 과업이 전개되는 주요 단계에 배치하는 방식을 제안하였다. 구체적으로 이용 여정을 Freytag의 5단계 구조[8]에 따라 발단-상승-절정-하강-결말로 구성하고, 과업 진행 과정에서 상태 변화를 포착할 수 있도록 주요 지점에 측정을 배치하였다. 본 연구에서는 단계 판정과 측정 위치가 순환적으로 결정되지 않도록 단계명을 직접 측정 조건으로 삼지 않고 사전에 정의한 주요 사건에 측정을 정박하였다. 이후 독립 코더가 각 사건의 Freytag 단계를 별도로 판정하였다.

단계별 경험을 측정하는 정서 차원으로는 PAD 모델을 채택하였다. Mehrabian과 Russell[9]은 환경에 대한 인간의 정서 반응을 즐거움(Pleasure), 각성(Arousal), 통제감(Dominance)의 세 차원으로 설명하였다. 선행 연구[6]는 공공 디지털 서비스에서의 통제감을 구조적 통제감(structural perceived control)과 행위적 통제감(perceived situational control)의 두 층위로 구분하였다. 구조적 통제감은 이탈 불가능성과 대안 부재에서 비롯되어 서비스 이용 전반을 제약하는 배경 조건을 의미한다. 행위적 통제감은 실제 과업 수행 과정에서 단계별 진행 상황에 따라 변화하는 통제 가능성을 이용자가 어떻게 지각하는가를 가리킨다[6]. 본 프레임워크의 PAD 측정은 이 가운데 행위적 통제감을 측정 대상으로 하며, 특정 사건에서 이용자가 과업의 진행을 어느 정도 통제할 수 있다고 지각하는지를 통제감(Dominance) 차원을 통해 포착한다[6].

선행 연구에서 제기된 세 가지 운용 과제는 본 연구의 설계와 다음과 같이 대응한다. Freytag 구조의 단계 경계에 대한 조작적 정의는 과업을 마이크로 이벤트 단위로 분해하고 독립 코더 간 배정의 일치도를 산출하는 방식으로 검증하였다. 서로 다른 측정 자료 간의 차이는 자기보고, 행동 자료 및 회고 발화를 동일한 사건 좌표에서 상호 대조하는 수렴-발산 분석으로 해석하였다. 과업 도중 이루어지는 측정의 반응성은 측정 인터페이스를 저개입 방식으로 재설계하고 과업 종료 후 참여자가 지각한 측정 방해 정도를 확인하였다.

## 2.2 Operationalizing Stage Boundaries: Micro-event Decomposition

Freytag 구조를 경험 측정의 좌표로 활용하기 위해서는 실제 서비스 과업에서 각 단계의 경계를 판정할 수 있어야 한다. 고객 여정 연구는 일반적으로 고객 경험을 시간에 따라 전개되는 여러 단계와 다양한 접점(touchpoint)의 연속으로 파악한다[10]. 그러나 이러한 단계나 접점의 구분만으로는 본 연구에서 관심을 두는 경험의 내러티브적 전환을 충분히 설명하기 어렵다.

Freytag 구조는 발단에서 상승을 거쳐 절정에 이른 후 다시 하강하여 종결되는 사건의 전개를 기본 형태로 한다[8]. 본 연구는 이러한 구조를 공공 디지털 서비스 과업에 적용하기 위해 과업 진행 과정에서 사건의 기능이 변화하는 지점을 단계 경계로 조작화하였다. 하나의 화면이나 동일한 접점 안에서도 사건의 기능이 달라질 수 있으므로 화면이나 접점보다 세분화된 사건 단위가 필요하다.

본 연구에서는 이용자의 행동이나 시스템의 반응으로 식별할 수 있는 최소 수준의 관측 가능한 사건을 ‘마이크로 이벤트(micro-event)’로 정의하고 전체 과업을 이 단위로 분해하였다. 이후 각 마이크로 이벤트가 Freytag 구조의 어느 단계에 해당하는지 독립 코더가 배정하고, 판정 결과의 일치 정도를 단계 경계 조작화의 검증 기준으로 삼았다. 실제 디지털 서비스에서는 오류 발생과 재시도가 반복되면서 사건의 진행이 순환하는 경우도 존재한다. 이러한 순환적 진행을 선형적인 Freytag 구조에서 어떻게 처리할 것인지는 선행 연구의 전문가 검증에서도 주요 쟁점으로 제기되었다[6]. 본 연구는 각 단계를 화면 위치 대신 사건의 기능과 진행 상태를 기준으로 정의하고 반복되는 구간에는 회차 정보를 부여하였다. 동일한 과정을 여러 번 수행하더라도 각 회차의 사건을 같은 판정 기준으로 분석할 수 있도록 하였다. 마이크로 이벤트의 분해 기준, 단계별 조작적 정의 및 반복 회차의 표기 규칙은 3.3절에서 구체적으로 제시한다.

## 2.3 Retrospective Verbal Reports and Narrative Structure Analysis

이용자가 과업을 수행하면서 구성한 내러티브는 행동 로그나 척도 응답만으로 직접 확인하기 어렵다. 동일한 과업을 경험한 이용자들이 그 경험을 서로 어떻게 다르게 조직하고 의미화하는지를 확인하려면 과업 종료 후 이용자가 자신의 경험을 어떻게 이야기로 구성하는지를 별도로 수집하고 분석할 필요가 있다. 본 연구는 이를 위해 회고적 발화 분석(Retrospective Think-Aloud, 이하 RTA)을

활용하였다. 언어보고는 과업과 동시에 이루어질 수도 있고 수행 이후 회고적으로 수집될 수도 있다[11]. 사용성 평가 연구에서 동시 발화는 회고적 발화보다 과업 수행 시간을 늘리는 경향이 있는 반면, 회고적 발화는 더 많은 설명과 문제 진술을 산출하지만 수행 당시 사고를 완전하게 회상하지 못할 가능성이 있다[12]. 실제 과업 수행에 말로 개입하는 정도를 최소화하려고 수행 이후에 회고적 발화를 수집하였다. 다만 회고 자료를 과업 당시 경험의 객관적 재생으로 간주하지 않고 이용자가 수행 경험을 사후에 어떻게 기억하고 의미화하는지를 보여주는 자료로 활용하였다[12][13].

수집된 발화의 서사 구조를 분석하기 위해 Labov와 Waletzky[14]가 제안하고 Labov[15]가 발전시킨 개인 경험 내러티브 분석 틀을 적용하였다. 이 접근에서는 개인 경험의 내러티브에서 요약(Abstract), 상황 설정(Orientation), 문제 사건(Complicating Action), 평가(Evaluation), 해소(Resolution), 종결(Coda)과 같은 기능적 요소를 구분한다[14][15]. 요약과 종결은 모든 내러티브에서 반드시 출현하지는 않는다. 평가는 하나의 독립된 구간으로 나타날 수도 있고 서사의 여러 부분에 분산되어 나타날 수도 있다[14][15]. 특히 평가에는 화자가 어떤 사건을 중요하게 인식했는지와 그 경험이 자신에게 어떠한 의미를 갖는지 드러난다.

본 연구에서는 이 이론을 바탕으로 상황 설정(Orientation), 문제 사건(Complicating Action), 평가(Evaluation), 해소(Resolution)의 네 기능을 이용자가 경험을 하나의 서사로 조직하였는지를 판정하기 위한 핵심 분석 요소로 조작화하였다. 네 요소의 출현 자체를 내러티브의 보편적 완결 조건으로 가정하지 않았으며, 각 요소가 어떠한 사건을 중심으로 구성되고 서로 어떠한 관계를 형성하는지를 함께 분석하였다. 분석의 초점은 이용자가 어떤 사건을 경험의 중심 사건으로 선택하였는지, 그 사건을 어떻게 평가하였는지, 이후의 사건을 어떠한 결과나 해소로 연결하였는지에 두었다. 또한 이용자의 자유 서사가 실제 수행 과정의 사건 순서를 그대로 재현하지 않거나 일부 사건을 축소·생략하더라도 이를 단순한 보고 오류로 처리하지 않았다. 내러티브란 화자가 사건을 선택하고 연결하여 의미를 구성하는 과정이지 경험의 복제가 아니기 때문이다[4][14]. 본 연구는 이러한 서사적 선택과 재구성을 실제 행동 로그 및 사건별 자기보고 자료와 대조하였다. 이 대조로 동일한 과업을 경험한 이용자 사이에서 나타나는 내러티브 구성의 개인차를 분석하였다.

### III. Research Method

#### 3.1 Study Design

본 연구는 선행 연구[6]에서 제안한 내러티브 기반 UX 평가 프레임워크를 실제 이용자 과업에 적용한 단일 집단 다중 사례 연구이다. 조작화된 측정 및 분석 절차가 의도한 방식으로 작동하는지를 확인하는 데 목적을 두었다. 참여자 5인(P01~P05)이 동일한 공공 디지털 서비스 프로토타입에서 동일한 과업을 수행하였다. 각 참여자의 전체 과업 수행 경험을 하나의 사례로 설정하고 사례 내 분석(within-case analysis)과 사례 간 분석(cross-case analysis)을 실시하였다[16].

본 연구에서 프레임워크의 작동 여부는 세 가지 연구 질문에 나누어 확인하였다. RQ1에서는 마이크로 이벤트 단위로 조작화한 Freytag 구조의 단계가 서로 독립된 코더에게 일관되게 판독되는지를 검토하였다. RQ2에서는 PAD의 Dominance를 측정하기 위해 재설계한 한국어 레이블이 이용자에게 연구자가 의도한 행위적 통제감의 의미로 이해되는지를 살펴보았다. RQ3에서는 동일한 과업과 동일한 최종 결과를 경험하더라도 이용자가 과업 중 발생한 사건을 어떻게 선택하고 연결하여 서로 다른 내러티브로 구성하는지를 분석하였다.

연구 질문에 답하기 위해 세 종류의 자료를 수집하였다. 주요 사건에 정박하여 수집한 PAD 자기보고, 과업 전 과정에서 연속적으로 기록한 히트맵과 이벤트 로그 등의 행동 자료, 과업 종료 후 수집한 회고적 발화(RTA)이다. 세 자료는 27개의 마이크로 이벤트(ME01~ME27)를 공통 시간축으로 하여 상호 대응시켰다. UX 평가에서 서로 다른 측정방법을 결합하는 삼각검증은 단일 자료원만으로 경험을 해석할 때 발생할 수 있는 한계를 보완할 수 있다[17].

본 연구는 5인의 사례로 프레임워크의 작동과 탐지 가능성을 확인하는 데 목적을 두며 모집단에 대한 통계적 일반화는 목적으로 하지 않는다. PAD와 행동 자료의 수치는 사례별 변화 양상을 기술하고 다른 자료와 대조하는 데 활용하였으며, 대규모 표본을 이용한 통계적 타당화(validation)는 후속 연구의 범위로 두었다.

#### 3.2 Participants

참여자는 목적적 표집(purposive sampling) 중 동질적 표집(homogeneous sampling) 전략을 적용하여 선정하였다[18]. 이는 참여자 간 연령과 디지털 활용 경험의 차이가 본 연구에서 관찰하려는 과업 경험의 차이와 지나치게 혼재되는 것을 줄이기 위한 것이다.

연령은 만 30세 이상 49세 이하로 제한하였다. 한국지능정보사회진흥원의 「2025 디지털정보격차 실태조사」에서는 연령에 따라 디지털 역량과 활용 수준에 차이가 나타난다고 보고한다[19]. 본 연구에서는 연령 범위를 제한하여 사례 간 비교 조건을 가능한 한 유사하게 유지하고자 하였다. 이 연령 범위는 연구 조건을 통제하기 위한 표집 기준일 뿐 연구 결과를 다른 연령대로 일반화하는 근거는 아니다.

포함 기준은 최근 2년 이내 정부24, 홈택스, 고용24 등 공공 디지털 서비스를 1회 이상 이용한 경험이 있고 PC 기반 인터넷 서비스를 일상적으로 사용하는 사람으로 설정하였다. 본 연구의 과업과 동일한 주민등록증 온라인 재발급 경험이 있는 사람은 과업 특정 경험이 결과에 영향을 미칠 가능성을 줄이기 위해 제외하였다. UX·HCI·UI/UX 디자인 또는 IT 개발 직군 종사자도 제외하였다. 전문적인 사용성 평가 경험이나 시스템 구조에 대한 지식이 일반 이용자의 과업 수행과 다른 방식으로 영향을 미칠 가능성을 줄이려는 조치이다.

최종 분석 사례는 5인으로 구성하였다. 참여자 수는 통계적 검정력으로 산정하지 않았다. Malterud 등[20]은 질적 연구에서 필요한 표본 규모가 고정된 숫자로 정하는 대신 연구 목적의 범위, 표본의 구체성, 기존 이론의 활용 정도, 자료의 질과 분석의 깊이에 따른 정보력(information power)을 고려하여 판단할 수 있다고 설명하였다. 본 연구는 이 네 조건에 다음과 같이 대응하였다. 연구 목적을 하나의 공공서비스 과업에서 프레임워크의 작동 여부를 확인하는 것으로 좁게 한정하였고, 동질적 표집으로 참여자의 연령과 이용 경험 조건(만 30~49세, 동일 과업 경험이 없는 일반 이용자)을 특정하였으며, 사전에 정의된 프레임워크와 판정 규칙을 분석에 적용하였고, 각 사례에서 PAD, 행동 로그, 화면 기록, RTA를 함께 수집하여 심층적으로 분석하였다. 이러한 조건에서는 적은 참여자 및 사례 수로도 연구 목적에 필요한 정보력을 확보할 수 있다고 보았다. 따라서 5인의 결과는 대표 통계치로 일반화하기 위한 것이 아니라 제안된 프레임워크의 작동을 확인하기 위한 것이다.

### 3.3 Task and Prototype Design with Micro-event Operationalization

실험 과업은 정부24의 주민등록증 재발급 과정을 모사한 프로토타입에서 수행하는 단일 과업으로 구성하였으며, 서비스 검색부터 민원신청 확인증 확인까지가 그 범위였다. 참여자에게는 기존 주민등록증의 사진을 최근 사진으로 변경하기 위해 온라인으로 재발급을 신청한다는 동일한 과업 맥락을 제시하였다. 인증 방식과 입력할 개인정보는 참여자 간 조건을 동일하게 유지하기 위해 연구자가 지

정한 가상 정보를 제공하였다.

실험은 실제 정부24 웹사이트 대신 연구 목적에 맞게 제작한 코드 기반 프로토타입에서 수행하였다. 실제 서비스에는 연구용 PAD 측정 모듈과 행동 추적 기능을 삽입할 수 없기 때문에 별도의 프로토타입을 구축하였으며, 모든 참여자가 동일한 화면 구조와 시스템 반응을 경험하도록 조건을 통제하였다. 프로토타입에는 디지털정부 서비스 UI/UX 가이드라인(KRDS)의 디자인 체계가 반영되어 있으며, 실제 서비스 화면의 정보 구조와 주요 상호작용도 재현하였다[21].

과업 중에는 사진 규격 확인과 관련하여 이용자가 예상한 진행 흐름이 중단될 수 있는 조건을 배치하였다. 참여자는 제공된 사진을 이용하여 신청을 진행하였으며, 규격을 충족하지 않는 사진을 제출하면 사진 규격 오류가 표시되고 기존 입력 정보가 소실되도록 구현하였다. 이후 참여자는 사진 규격 안내를 확인하고 사진을 조정된 뒤 신청 과정을 다시 수행하도록 하였다.

이와 같이 사용 과정의 사건과 정서 상태를 시간적으로 대응시키는 접근은 이용자의 조작 사건과 PAD 변화를 연결하여 분석한 선행 연구에서도 활용된 바 있다[22]. 본 연구에서도 일정한 시간 간격 대신 과업의 진행에 의미 있는 변화가 발생하는 사건을 측정과 분석의 기준으로 삼았다.

Table 1. Operational definitions of Freytag's five stages

| Stage          | Operational definition                                                                                                                 | Observable indicators                                                                           |
|----------------|----------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|
| Exposition     | Entry and exploration segment for reaching the task goal target                                                                        | Search, item identification, navigation before reaching the target function                     |
| Rising action  | Segment of fulfilling the requirements imposed by the system to achieve the goal                                                       | Chains of requirements and fulfillment such as authentication, information input, and selection |
| Climax         | Segment in which progress is blocked or previously fulfilled requirements are invalidated, and no valid resolving action has yet begun | Error display, input invalidation, exploration of causes and resolution paths                   |
| Falling action | Segment in which valid resolving actions for the crisis begin and the task proceeds again                                              | Problem-solving actions, re-fulfillment of requirements                                         |
| Dénouement     | Segment of confirming the outcome after the task completion signal                                                                     | Completion feedback, confirmation of result details                                             |

Table 2. Summary of micro-events in the experimental task

| ME        | Event description                                                                    |
|-----------|--------------------------------------------------------------------------------------|
| ME01      | Entry into the Gov24 main page                                                       |
| ME02-ME03 | Entering search terms and identifying the resident registration card reissuance item |
| ME04-ME05 | Selecting 'Apply' and selecting the individual non-member application                |
| ME06-ME09 | Selecting simple authentication, entering information, and completing authentication |
| ME10      | Reappearance of the search list and re-entry into the application screen             |
| ME11-ME12 | Confirming applicant information and entering reissuance information                 |
| ME13      | Searching for and selecting the receiving office                                     |
| ME14      | Entry into the photo upload section                                                  |
| ME15      | Uploading the photo                                                                  |
| ME16      | Submitting the application form                                                      |
| ME17      | Photo specification error and initialization of the entered information              |
| ME18      | Return to the initialized application screen and exploration of resolution paths     |
| ME19      | Adjusting the photo using an external editing tool                                   |
| ME20      | Return to the task screen and re-entering the information                            |
| ME21      | Re-uploading the adjusted photo                                                      |
| ME22      | Acceptance of the reapplication and transition to the fee payment screen             |
| ME23      | Mock payment                                                                         |
| ME24-ME26 | Exploring the application history and confirming the details                         |
| ME27      | Confirming the civil application certificate and completing the task                 |

Note. ME = micro-event.

전체 과업은 27개의 마이크로 이벤트(ME01~ME27)로 분해하였다. Grimaldi[23]는 제품 사용 경험을 시간의 흐름에 따라 이어지는 사건 또는 마이크로 이벤트의 연쇄로 파악하고 이를 세부 사건 단위로 분석하였다. 본 연구는 이러한 사건 분해의 원리를 디지털 서비스 과업에 적용하였다. 마이크로 이벤트는 이용자의 행동이나 시스템 반응으로 관측할 수 있는 최소 수준의 사건으로 정의하였다. 화면 전환, 입력 요구, 오류 발생, 시스템 반응, 성공 피드백, 화면 이탈·복귀 및 목표 대상 식별 등을 주요 분해 기준으로 활용하였다.

이 분해 기준은 Grimaldi가 직접 제시한 분류 체계가 아니라 그의 사건 분해 원리를 디지털 서비스 과업에 적용하기 위해 본 연구에서 조작적으로 정의하였다. 각 마이크로 이벤트의 Freytag 구조 단계는 참여자의 감정을 추정하여 배정하지 않고 과업 진행 과정에서 해당 사건이 수행하는 기능을 기준으로 판정하도록 하였다. 단계별 조작적 정의와 관찰 지표는 Table 1과 같다.

Table 2는 ME01~ME27의 전체 과업 구조를 요약한 것이다. 연속적으로 동일한 기능을 수행하는 이벤트는 지면을 고려하여 하나의 구간으로 묶어 제시하였으나 실제 단계 판정에서는 27개 이벤트를 각각 독립된 단위로 사용하였다. 오류와 재시도가 반복되는 경우에는 동일한 이벤트 번호에 회차를 추가하여 ME17.1, ME17.2와 같이 기록하였다. 예를 들어 PAD2가 측정된 ME17.1은 첫 번째 사진 규격 오류를 가리킨다. 같은 오류가 반복되더라도 새로운 PAD 측정을 실시하지 않고 회차만 추가하여 기록하였다.

### 3.4 PAD Self-report Measurement and Behavioral Data Collection

PAD 세 차원은 0~100의 연속형 시각 아날로그 척도(Visual Analogue Scale, 이하 VAS)로 측정하였다. Reips와 Funkel[24]는 웹 기반 연구에서 VAS가 연속적인 응답값을 제공하며 등간수준에 근접한 측정에 활용될 수 있음을 보고하였다. 본 연구에서는 사건 사이에서 나타나는 비교적 작은 상태 변화까지 기록할 필요가 있으므로 연속형 VAS를 사용하였다. PAD의 대표적인 비언어적 측정 도구인 SAM(Self-Assessment Manikin)을 그대로 사용하지 않고 한국어 레이블을 포함한 인터페이스로 재설계하였다. Bradley와 Lang[25]은 SAM과 의미미분법을 비교한 결과 Pleasure와 Arousal에서는 높은 일치치가 나타난 반면, Dominance에서는 상대적으로 낮은 일치를 보고하였다. 디지털 기반 정서 측정도구인 Affective Slider는 Pleasure와 Arousal 두 차원을 측정하지만 Dominance는 포함하지 않는다[26].

선행 연구[6]에서 정의한 행위적 통제감을 이용자가 직접 이해할 수 있도록 Dominance 레이블을 재구성하였다. Fig. 1은 프로토타입에서 PAD 측정 인터페이스 화면을 캡처한 것으로 최종 레이블은 Pleasure의 경우 '불쾌하다-유쾌하다', Arousal은 '차분하다-긴장된다', Dominance는 질문 '원하는 대로 되고 있나요?'와 양 끝점 '안 되고 있다-잘 되고 있다'로 구성하였다. 여기서 Dominance는 특정 시점에서 이용자가 과업의 진행을 어느 정도 자신이 원하는 방향으로 통제할 수 있다고 느끼는지를 의미한다. 측정 인터페이스는 과업 화면 하단에서 올라오는 드로어(drawer) 방식으로 구현하였다. 별도의 설문 화면으로 전환하지 않고 현재 과업 화면을 유지한 상태에서 응답할 수 있도록 하였으며, 응답 완료 후 과업을 계속 수행하도록 설계하였다.

PAD 측정에는 기저선과 네 개의 사건 기반 측정 시점을 설정하였다. 기저선은 과업 진입 직후 실제 조작을 시

작하기 전에 측정하였다. 수령기관 선택 완료 시점(ME13)에 PAD1을, 사진 규격 오류와 입력 정보 소실이 처음 발생한 시점(ME17.1)에 PAD2를 측정하였다. 재신청이 정상적으로 수용된 시점(ME22)에는 PAD3을, 민원신청 확인증을 통해 과업 종료로 확인한 시점(ME27)에는 PAD4를 측정하였다. 측정 위치는 Freytag 단계명을 직접 조건으로 삼지 않고 사전에 정의된 사건 좌표에 정박하였다. 이후 독립 판정자가 해당 사건들을 어느 Freytag 단계로 판정하는지를 별도로 분석하여 측정 위치와 단계 판정이 순환적으로 결정되지 않도록 하였다.

(a) Pleasure

지금 이 순간의 느낌을 알려 주세요

각 항목에서 지금 상태에 가장 가까운 위치를 클릭하거나 마우스를 드래그 하여 움직여 주세요. 표시한 뒤 아래 버튼을 눌러 다음으로 넘어갑니다.

현재 느낌은 어떤가요?

불쾌하다                      중립                      유쾌하다

클릭 또는 마우스 드래그

다음 →

(b) Arousal

지금 이 순간의 느낌을 알려 주세요

각 항목에서 지금 상태에 가장 가까운 위치를 클릭하거나 마우스를 드래그 하여 움직여 주세요. 표시한 뒤 아래 버튼을 눌러 다음으로 넘어갑니다.

지금 긴장 상태는 어떤가요?

차분하다                      중립                      긴장된다

클릭 또는 마우스 드래그

다음 →

(c) Dominance

지금 이 순간의 느낌을 알려 주세요

각 항목에서 지금 상태에 가장 가까운 위치를 클릭하거나 마우스를 드래그 하여 움직여 주세요. 표시한 뒤 아래 버튼을 눌러 다음으로 넘어갑니다.

원하는 대로 되고 있나요?

안 되고 있다                      중립                      잘 되고 있다

클릭 또는 마우스 드래그

완료

Fig. 1. Event-anchored PAD self-report interface

행동 자료는 과업 전 과정에서 연속적으로 수집하였다. 히트맵 분석으로 Click, Move, H2C(Hover-to-Click), Path, Dwell, Scroll 지표를 확보하였고, 자체 이벤트 로그를 통해 클릭, 화면 전환, 오류 발생 및 PAD 응답 시각을 기록하였다. 화면 녹화 자료는 수행 과정의 사건 순서를 확인하고 서로 다른 자료원의 시간 정보를 정렬하는 기준으로 사용하였다. 행동 자료는 이용자의 수행 위치, 시간, 오류를 보여준다. 다만 객관적 수행지표와 주관적 평가는 대체로 낮은 상관을 보이므로 행동만으로 경험의 심리적 의미를 확정하기는 어렵다[27]. 이 때문에 동일한 마이크로 이벤트에 대응하는 PAD 자기보고와 RTA 발화를 함께 대조하였다[17].

### 3.5 Experimental Procedure and Retrospective Think-Aloud Collection

실험은 참여자별 1:1 대면 방식으로 진행하였다. 먼저 연구 목적과 절차를 설명하고 동의를 받은 뒤 과업 화면 진입 시 PAD 기저선을 측정하였다. 이후 참여자가 주민등록증 재발급 과업을 수행하는 동안 행동 로그와 화면 기록을 연속적으로 수집하고 사전에 설정한 사건 발생 시점에서 PAD를 측정하였다.

과업 수행 중 연구자는 원칙적으로 참여자의 조작에 개입하지 않았다. 다만 참여자가 사진 규격 오류를 두 차례 연속 경험한 경우에는 1회 안내하였다. 두 번째 오류를 확인하고 초기화된 신청 화면으로 복귀한 시점에 모든 참여자에게 동일한 문구로 사진 규격을 확인한 뒤 다시 신청하도록 하였다. 이때 규격 안내의 위치나 사진 조정 방법을 직접 알려주지 않고 참여자가 화면 내 정보를 찾아 이후의 해결 과정을 수행하도록 하였다.

과업 완료 후에는 사후 설문을 실시하고 실험에서 제공한 사진 조건과 오류·입력 정보 소실 조건을 설명하는 디브리핑을 진행하였다. 이후 과업 경험에 대한 RTA 인터뷰를 실시하였다. 사용성 평가에서 과업 중의 동시 발화는 일부 수행 조건에 영향을 미칠 가능성이 보고되어 있다[12]. 본 연구에서는 과업 수행 자체에 말로 개입하지 않고 수행 이후에 발화를 수집하였다. 자극 회상 방식의 RTA(stimulated RTA)가 과업 당시의 주의와 경험을 회고하는 방법으로 활용될 수 있다는 선행 연구도 참고하였다[13].

RTA는 제시되는 정보가 이용자의 자유로운 경험 구성을 유도하지 않도록 단계적으로 진행하였다. U1a에서는 아무런 자료를 제시하지 않은 상태에서 과업을 처음부터 끝까지 하나의 이야기로 설명하도록 요청하였다. 참여자가 사건을 생략하거나 실제 수행 순서와 다르게 이야기하더라도 이를 수정하지 않았다. U1b에서는 자유 서사가 종료된 후 긴장이나 분위기가 달라졌다고 느낀 지점을 질문하였다. U2에서는 과업에서 상황이 자신의 의도대로 진행되거나 진행되지 않았다고 느낀 지점을 물었다. U1a, U1b, U2까지는 PAD 값이나 히트맵을 제시하지 않았다. U3부터 각 측정 시점의 화면과 자료를 순서대로 제시하면서 해당 시점의 정서를 회고하도록 하였다. U4에서는 사진 규격 오류와 입력 정보 소실 당시의 사고와 대응을 들었다. U5에서는 오류의 원인 귀속과 이후의 해결 과정을 질문하였다.

Dominance 레이블의 해석은 유도 효과를 줄이기 위해 단계를 나누어 점검하였다. 먼저 참여자가 과업 경험을 자

신의 언어로 설명하는 과정에서 ‘마음대로 되지 않았다’, ‘진행이 안 됐다’와 같은 통제 관련 표현이 자발적으로 나타나는지를 살폈다. 자발적인 표현만으로 의미를 판단하기 어려운 경우에만 실제 Dominance 레이블을 제시하고 응답 당시 해당 문구를 어떤 의미로 이해했는지를 본인의 말로 설명하도록 하였다. 디브리핑 이후 실시되는 인터뷰라는 점을 고려하여 오류 원인이나 당시의 판단을 질문할 때에는 ‘그 사실을 알기 전 그 순간에는 어떻게 생각했는가’와 같이 응답 시점을 과업 당시로 다시 고정하였다.

### 3.6 Data Analysis

자료 분석은 RQ1의 단계 경계 판정, RQ2의 Dominance 레이블 해석, RQ3의 내러티브 구성 분석으로 구분하였다. 27개 마이크로 이벤트 각각에 대해 두 명의 독립 코더가 발단, 상승, 절정, 하강, 결말 중 하나의 단계를 독립적으로 배정하였다. 코더에게는 연구자의 단계 가설이나 PAD 측정 위치를 제시하지 않고 마이크로 이벤트 설명과 단계별 조작적 정의 및 관찰 지표만 제공하였다. 코더 간 일치도는 관측 일치율과 Cohen's  $\kappa$ 로 산출하였다[28]. 서로 다른 단계가 배정된 이벤트가 단계 경계 부근에 집중되는지도 별도로 검토하였다. 여기서 전체적인 단계 판정의 재현성뿐 아니라 단계 전환의 조작적 정의가 모호하게 작동하는 위치를 함께 파악할 수 있었다.

Dominance 레이블의 해석은 RTA에서 수집된 발화를 의미 단위로 나누어 분석하였다[29]. 참여자가 ‘안 되고 있다-잘 되고 있다’를 과업 진행의 통제 가능성이나 자신이 원하는 방향으로 진행되는 정도로 해석한 경우에는 의도한 의미와 일치하는 것으로 보았다. 반면 단순한 성공/실패, 자신의 능력 평가 또는 Pleasure-Arousal과의 혼동이 나타난 경우에는 그 차이를 별도로 기록하였다. 이 분석은 Dominance 척도의 통계적 구성타당도 검증을 목표로 하지 않았다. 선행 연구[6]에서 제안한 행위적 통제감 개념을 실제 이용자가 의도한 의미로 해석하는지를 확인하기 위한 조작적 검증이다.

내러티브 구성은 세 단계로 분석하였다. U1a 자유 서사를 절(clause) 단위로 분절한 뒤 Labov 계열의 요약(Abstract), 상황 설정(Orientation), 문제 사건(Complicating Action), 평가(Evaluation), 해소(Resolution), 종결(Coda) 요소로 분류하였다[14][15]. 이 가운데 상황 설정, 문제 사건, 평가, 해소의 네 기능이 어떠한 관계로 연결되는지를 중심으로 분석하였다. 네 요소가 시간의 순서와 인과로 연결되어 하나의 경험 구조를 이루는 경우 구성 성공(NC-S), 일부 요소가 부족하지만 경

험의 진행 방향을 확인할 수 있는 경우 부분 구성(NC-P), 사건이 파편적으로 나열되어 문제 사건과 해소 사이의 연결을 확인하기 어려운 경우 구성 실패(NC-F)로 분류하였다. 이 세 유형은 Labov가 제시한 보편적 내러티브 유형이 아니다. 사례 간 비교를 위해 본 연구에서 설정한 조작적 분류이다.

각 서사 요소가 지칭하는 사건을 ME01~ME27과 대응시켰다. 실제 사건의 발생 순서는 이벤트 로그와 화면 녹화로 특정하였다. 그러나 참여자가 회고 과정에서 사건의 순서를 다르게 배열하거나 특정 사건을 생략한 경우에는 발화의 순서를 실제 로그에 맞게 수정하지 않았다. 이러한 차이는 실제 수행과 이용자가 구성한 내러티브 사이의 차이로 기록하였다. 이용자가 이야기 속에서 지각한 통제감(Dominance)의 상태를 N1~N4로 분류하였다. N1은 통제감 유지, N2는 통제감 동요, N3은 통제감 붕괴, N4는 통제감 회복을 의미한다. 이 상태는 PAD 값으로 직접 결정하지 않고 RTA 발화를 중심으로 판정하였다.

마지막으로 RTA, PAD 및 행동 자료를 동일한 마이크로 이벤트 시간축 위에서 대조하였다. 세 자료가 같은 방향의 경험 해석을 지지하면 수렴(C), 일부 자료가 추가적인 정보를 제공하지만 서로 모순되지 않는 경우 부분 수렴(P), 자료 간 해석 방향이 서로 충돌하는 경우 발산(D)으로 기록하였다. 발산한 자료를 어느 하나의 자료에 맞추어 수정하지 않고 측정 방식 간 차이를 보여주는 사례로 별도 분석하였다.

### 3.7 Ensuring the Trustworthiness of the Analysis

본 연구자는 검증 대상 프레임워크의 설계자이므로 단계 판정과 사례 해석에 연구자의 사전 가설이 개입할 가능성을 고려하였다. 마이크로 이벤트의 Freytag 단계는 독립 코더 2인이 배정하였다. 이들에게는 연구자의 단계 가설과 PAD 측정 위치를 제공하지 않았다. 코더 간 일치도는 최초 독립 코딩 결과를 이용하여 Cohen's  $\kappa$ 와 관측 일치율을 산출하였다[28].

RTA 분석에서는 발화의 분절, Labov 요소 배정, 마이크로 이벤트 정렬 및 통제감 상태 판정에 관한 규칙을 분석 전에 문서화하였다. 발화와 로그의 순서가 상충하는 경우, 모호한 요소 배정이 발생한 경우, 반복 오류의 회차를 특정하기 어려운 경우에는 판단 과정과 근거를 감사추적(audit trail)에 기록하였다[30].

자료 수집 및 회고 인터뷰 과정에서 절차를 벗어난 사례가 나온 경우에도 사전에 정한 원칙에 따라 분석 범위를 제한하였다. P01의 PAD2는 최초 오류 직후의 사건 정박

값으로 직접 해석하지 않았다. 대신 행동 기록과 회고적 발화를 함께 참고하였다. 처리 과정은 모두 감사추적에 기록하였다.

## IV. Results and Analysis

### 4.1 Overview of Task Performance and Measurement Results

참여자 5인 전원이 주민등록증 재발급 신청 과업을 완료하였다. 실험 설계상 사진 규격을 사전에 확인하여 오류를 회피하는 경로도 가능하였으나 실제 참여자는 모두 사진 규격 오류를 경험하는 경로로 진행하였다. 사진 규격 오류는 P02가 4회로 가장 많이 경험하였으며 나머지 4인은 각각 2회의 오류를 경험하였다.

최초 제출 이전에 사진 규격 안내를 열람한 참여자는 P02와 P05의 2인이었다. 두 참여자 모두 안내를 확인하였으나 요구 규격과 제공된 사진을 정확하게 대응시키지 못해 이후 오류를 경험하였다. 나머지 참여자들은 최초 제출 이전에 규격 안내를 열람하지 않았다. 오류와 입력 정보 초기화 이후에는 모든 참여자가 신청 정보를 다시 입력하거나 사진을 선택·조정하는 과정을 반복하였다. 두 번째 오류 이후에는 모든 참여자가 동일한 연구자 안내를 받았으며, PAD3과 PAD4는 이 안내 이후 구간에서 측정된 값이다. 순수 과업 시간은 참여자에 따라 4.7분에서 24.0분까지 큰 차이를 보였다. 동일한 과업을 수행하고 모두 최종 완료에 도달하였지만 문제를 발견하고 해결하는 과정과 이에 소요된 시간은 사례별로 같았다.

Table 3은 기저선과 네 개의 사건 정박 시점에서 측정된 Pleasure, Arousal, Dominance 값을 참여자별로 제시한 것이다. 세 차원의 절대값과 변화 양상은 참여자마다 서로 달랐고 기저선에서도 P·A·D 수준이 같았다. 이후 분

석에서는 참여자 간 절대값을 직접 비교하는 대신 각 사례 내에서 사건 전후의 변화 방향을 중심으로 살펴보았다. 전반적으로 오류 발생 이후 Pleasure가 낮아지고 Arousal이 높아지면서 Dominance가 낮아지는 사례가 있었다. 변화 폭이 작거나 과업 성공 이후에도 Dominance가 즉시 회복되지 않는 사례도 함께 나타났다. 본 연구에서 행위적 통제감으로 조작화한 Dominance는 연구질문과 직접 연결되는 핵심 분석 축으로 사용하였다. Pleasure와 Arousal은 Dominance 변화의 정서적 맥락을 해석하는 보조 자료로 활용하였다. Fig. 2는 이 가운데 본 연구의 핵심 분석 축인 통제감(Dominance)의 참여자별 변화 궤적을 별도로 시각화한 것이다.

### 4.2 RQ1: Inter-coder Agreement on Freytag Stage Assignment

두 명의 독립 코더가 ME01~ME27의 27개 마이크로 이벤트에 Freytag 단계를 배정한 결과 27개 중 26개 이벤트에서 동일한 단계가 배정되었다. 관측 일치율은 96.3%였으며, Cohen's  $\kappa$ 는 .948이었다[28]. 코더 간 불일치는 ME18에서 1건 나타났다. ME18은 사진 규격 오류와 입력 정보 소실 이후 초기화된 신청서 화면으로 돌아와 재입력 또는 해결 경로 탐색을 시작하는 사건이다. 코더 1은 재입력이 시작되었다는 점에서 이를 하강으로 판정한 반면, 코더 2는 사진 규격 문제가 아직 해결되지 않은 상태라는 점에서 절정으로 판정하였다.

두 판정은 서로 다른 기준을 적용한 단순한 오타라기보다 ME17에서 동시에 발생한 두 문제 가운데 무엇을 위기의 기준으로 보는가에 따라 달라진 것으로 해석하였다. 입력 정보의 소실을 기준으로 보면 재입력의 시작은 해소 행동으로 볼 수 있으나 사진 규격 오류를 기준으로 보면 문제의 해결 방법을 아직 찾지 못한 상태이므로 절정이 지속되는 것으로 볼 수 있다. 따라서 ME18은 절정과

Table 3. PAD scores across event-anchored measurement points

| Participant | Baseline        | PAD1 (ME13)   | PAD2 (ME17.1)             | PAD3 (ME22)   | PAD4 (ME27)  |
|-------------|-----------------|---------------|---------------------------|---------------|--------------|
| P01         | 59 / 67 / 100   | 26 / 74 / 15  | 15 / 50 / 10 <sup>a</sup> | 78 / 25 / 92  | 96 / 15 / 62 |
| P02         | 100 / 100 / 100 | 49 / 100 / 51 | 54 / 100 / 50             | 47 / 100 / 50 | 53 / 50 / 86 |
| P03         | 98 / 9 / 70     | 81 / 21 / 79  | 60 / 29 / 52              | 54 / 41 / 64  | 62 / 37 / 64 |
| P04         | 60 / 59 / 59    | 60 / 58 / 62  | 59 / 63 / 61              | 71 / 68 / 68  | 71 / 56 / 66 |
| P05         | 49 / 71 / 91    | 13 / 67 / 36  | 5 / 82 / 20               | 6 / 87 / 13   | 33 / 51 / 62 |

Note. Values in each cell are presented in the order of Pleasure / Arousal / Dominance. <sup>a</sup>P01 did not respond to the PAD2 interface presented immediately after the first photo specification error, closed it, and continued the task; the interface was reopened and answered approximately 66.7 seconds later while the information was being re-entered. Therefore, this value was not directly compared as an event-anchored measurement immediately after ME17.1.

하강의 경계를 판정할 때 ‘유효한 해결 행동의 시작’을 어떤 문제를 기준으로 정의할 것인지 더 명료하게 규정해야 할 사건으로 남았다.

연구자의 사전 가설과 두 코더의 결과가 함께 달랐던 사건은 ME04였다. 연구자는 ‘신청하기’ 버튼을 선택하는 ME04부터 상승이 시작되는 것으로 가정하였으나 두 코더는 모두 ME04를 발단으로 판정하였다. 코더들은 이 시점에서는 아직 시스템이 인증이나 정보 입력과 같은 구체적인 요구를 제시하지 않았으며 실제 요구가 시작되는 ME05 이후부터 상승으로 보는 것이 조작적 정의에 부합한다고 판단하였다. 사전에 단계 경계 후보로 설정한 사건은 ME04, ME15·ME16, ME18, ME23·ME24의 6개였다. 코더 간 불일치 1건(ME18)과 연구자 사전 가설과의 이탈 1건(ME04)은 모두 이 경계 후보에 포함되었으며 나머지 21개 이벤트에서는 두 코더와 연구자의 사전 가설이 모두 일치하였다.

이상의 결과에서 마이크로 이벤트에 적용한 단계 조작적 정의는 대부분의 사건에서 높은 판정 일치도를 보였다. 판정 차이는 실제 단계가 전환되는 경계 사건에 몰려 있었다. RQ1에서 전체적인 단계 판독은 일관되게 이루어졌으나 일부 경계에서는 더 세부적인 판정 규칙이 과제로 남았다. 이에 따라 ME18과 ME04에서 확인된 판정 차이를 반영하여 후속 적용을 위한 경계 규칙을 다음과 같이 보완하였다. 하나의 사건에서 복수의 문제가 동시에 발생한 경우에는 과업의 진행을 직접 차단하는 문제를 기준 문제로 지정한다. 절정에서 하강으로의 전환은 기준 문제에 대한 유효한 해결 행동이 처음 관측되는 사건으로 정의하며, 유효

한 해결 행동은 기준 문제의 원인을 제거하는 행동을 말한다. 소실된 입력 정보의 재입력처럼 기준 문제를 해결하지 않은 채 이전 상태를 복구하는 행동은 유효한 해결 행동에 포함하지 않는다. 이 규칙을 적용하면 ME17의 기준 문제는 사진 규격 오류이고, 재입력이 시작되는 ME18은 절정으로 유지되며, 사진 조정이 시작되는 ME19가 하강의 시작에 해당한다. 발단에서 상승으로의 전환은 이용자의 행동 의도가 아니라 시스템이 인증이나 정보 입력과 같은 구체적인 요구를 처음 제시하는 사건으로 정의하며, 이 기준에 따르면 ‘신청하기’ 선택(ME04)은 발단, 신청 유형 선택 요구가 제시되는 ME05가 상승의 시작이다. 이 보완 규칙은 Table 1의 조작적 정의를 대체하지 않고 경계 사건의 판정 절차를 구체화한 것이며, 본 연구의 코더 판정 결과를 사후에 변경하지 않았다. 규칙의 재현성은 후속 연구에서 새로운 코더와 과업을 대상으로 확인해야 한다.

#### 4.3 RQ2: User Interpretation of the Dominance Label

본 연구에서 Dominance는 선행 연구[6]에서 정의한 행위적 통제감을 측정하기 위해 ‘원하는 대로 되고 있나요?’라는 질문과 ‘안 되고 있다-잘 되고 있다’의 양극 레이블로 제시하였다. 회고 인터뷰에서 이 레이블은 대체로 자신의 능력 자체보다는 현재 과업이 원하는 방향으로 진행되고 있는지를 판단하는 의미로 사용되었다. P01은 상황이 잘 진행되는 경우 ‘되고 있다’ 쪽에 응답하는 의미로 이해했다고 설명하여 연구자가 의도한 과업 진행의 통제 가능

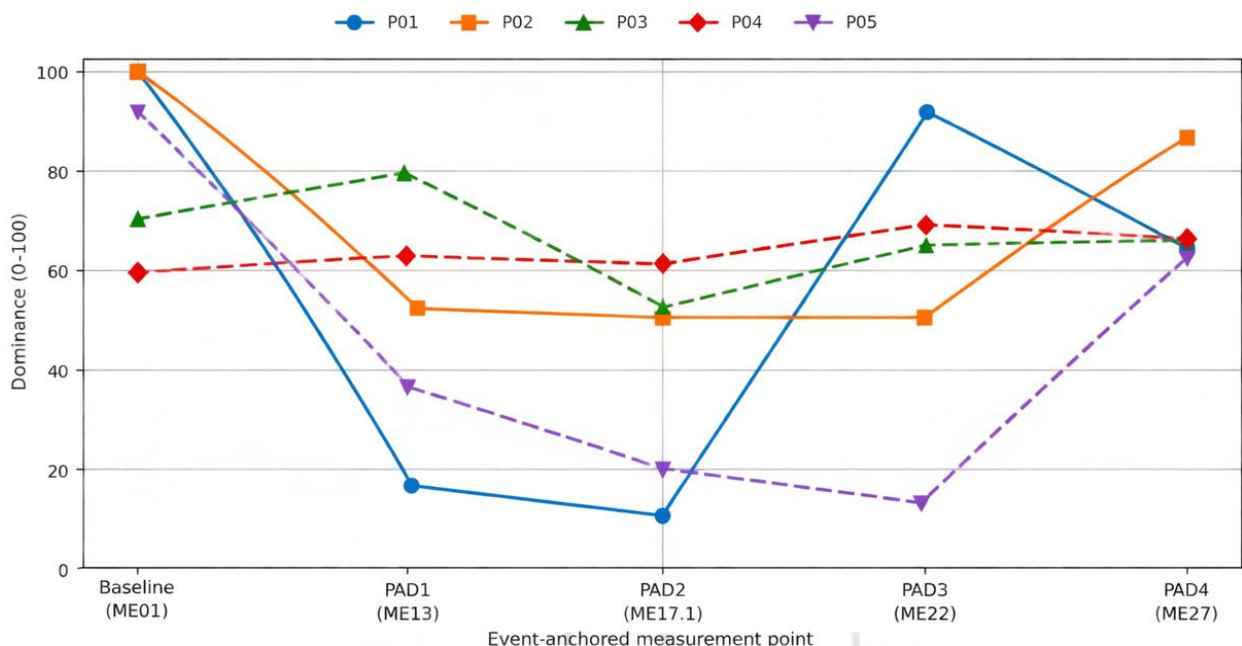


Fig. 2. Individual Dominance trajectories across event-anchored measurement points

성과 직접 대응하는 해석을 보였다.

P05에서는 과업의 성공 여부와 Dominance가 반드시 동일하게 움직이지 않는 사례가 나타났다. 최초 오류 직후인 PAD2에서는 P=5, A=82, D=20으로 Pleasure와 Dominance가 낮고 Arousal이 높았다. 재신청이 수용된 PAD3에서도 P=6, A=87, D=13으로 Dominance가 회복되지 않았다. 참여자는 이 시점에 대해 문제는 해결되었지만 연구자의 안내 후 해결한 것이기 때문에 자신이 과업을 잘 진행하고 있다고 느끼지는 않았다고 설명하였다. 적어도 P05 사례에서 '잘 되고 있다'라는 응답은 단순한 과업 성공 여부보다 과업의 진행을 자신이 주도하고 있다는 인식과 연결되어 쓰였다.

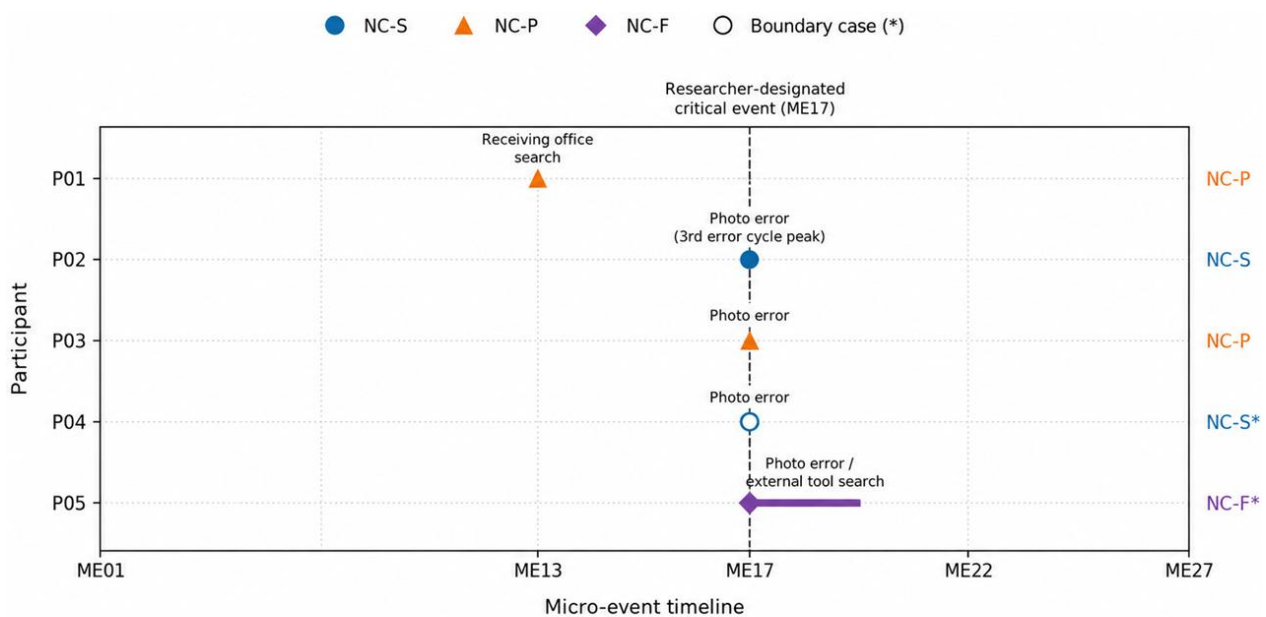
PAD의 각 차원이 참여자 경험에서 완전히 분리되어 해석된 것은 아니었다. P04는 과업이 잘 진행되지 않을 때 번거로움이나 귀찮음, 짜증과 같은 불쾌감도 함께 발생한다고 설명하여 Dominance와 Pleasure 사이의 경험적 연관성을 언급하였다. 이는 Dominance 레이블 자체가 이해되지 않았다는 의미라기보다 실제 서비스 경험에서는 과업 진행의 통제감과 정서적 쾌·불쾌가 서로 연동되어 경험될 수 있음을 보여준다.

측정 인터페이스와 관련된 운용 특성도 확인되었다. P02는 VAS가 연속적으로 값을 선택하는 방식임을 충분히 이해하지 못하고 슬라이더를 끝까지 이동시킨 경험을 보고하였다. P02의 기저선에서 P·A·D가 모두 100이었고 이후 Arousal이 반복적으로 100에 머문 현상은 이러한 조작 특

성의 영향을 받았을 가능성이 있다. 반면 P04는 기저선에서 P=60, A=59, D=59였고 오류가 발생한 PAD2에서도 P=59, A=63, D=61로 큰 변화를 보이지 않았으며 Dominance는 전 측정 시점에서 59~68의 좁은 범위에 머물렀다. 참여자는 인터뷰에서 실제 경험이 '보통' 수준이었으며 중양 부근의 응답이 당시 상태를 반영한 것이라고 설명하였다. 따라서 P04의 저진폭 궤적은 P02에서 확인된 VAS 조작 특성과 구분되는 실제 중양 응답 사례로 해석하였다.

측정 드로어의 방해 정도 역시 참여자마다 달랐다. 일부 참여자는 크게 방해되지 않았다고 보고한 반면 P05는 과업을 빠르게 계속하고 싶은 상황에서 드로어가 나타나는 데 대해 약 60% 수준의 방해를 느꼈다고 보고하였다. P05는 Arousal의 '차분하다'라는 표현도 공공서비스 이용 맥락에서는 다소 어색하게 느껴진다고 설명하였다.

이상의 사례는 Dominance 응답값의 차이가 어디에서 비롯되는지에 따라 세 유형으로 구분할 수 있다. P02는 VAS 조작 방식을 충분히 이해하지 못한 상태에서 응답하였다고 보고하였으므로 측정 인터페이스의 영향이 응답값에 개입하였을 가능성이 있는 사례이다. P04는 저진폭 응답이 실제 경험 수준을 반영한 것이라고 설명하여 측정 방식이 아니라 실제 상태를 나타낸 사례이다. P05는 과업 성공 이후에도 낮은 Dominance를 보고하고 그 이유를 과업 진행의 주도 여부로 설명하여 레이블 해석이 의도한 의미와 일치한 사례이다. 따라서 P02의 결과는 행위적 통제감 자체의 변화로 해석하기보다 측정 방식의 영향 가능성을



Note. The dashed vertical line indicates ME17, the researcher-designated critical event. \* indicates a boundary case.

Fig. 3. Participant-specific narrative peak locations and narrative construction types across the micro-event timeline

함께 고려해야 하며, 본 연구의 자료만으로는 실제 상태 변화와 측정 인터페이스의 영향을 완전히 분리하지 못하였다.

RQ2에서 Dominance 레이블이 행위적 통제감이라는 의도한 의미로 해석될 가능성은 확인되었다. 다만 이 결과는 레이블의 조작적 해석 가능성에 관한 것이며, 척도의 신뢰도와 구성타당도를 포함한 심리측정학적 속성은 본 연구의 검증 범위에 포함되지 않는다. PAD 차원 간 경험적 연동, VAS 조작 방식 및 측정 인터페이스의 반응성은 후속 설계에서 별도로 검토할 과제로 남았다.

#### 4.4 RQ3: Cross-case Differences in Narrative Construction

본 연구의 조작적 판정 기준에 따라 자유 서사를 분석한 결과, 구성 성공(NC-S)은 P02와 P04의 2인, 부분 구성(NC-P)은 P01과 P03의 2인, 구성 실패(NC-F)는 P05의 1인이었다. P04의 NC-S와 P05의 NC-F는 인접 유형으로도 해석할 수 있는 경계 사례로 함께 기록하였다. 사례별 판정은 핵심 서사 요소가 출현했는지만 보지 않고 문제 사건과 평가, 해소가 하나의 경험 구조로 연결되는지를 기준으로 삼았다.

참여자 자신이 경험에서 가장 긴장이 높거나 진행이 어렵다고 지목한 사건도 사례마다 달랐다. 이를 공통 마이크로 이벤트 시간축에 정렬한 결과를 Fig. 3에 제시하였다. P02, P03과 P04는 사진 규격 오류가 발생한 ME17을 주관적 절정으로 지목하였다. P02는 사건 수준에서는 ME17을 절정으로 구성하였으나 반복된 오류 가운데 세 번째 오류에서 긴장이 가장 높았다고 회고하였다. P05는 사진 규격 오류에서 외부 도구를 이용한 해결 탐색으로 이

어지는 ME17~ME19를 하나의 높은 긴장 구간으로 구성하였다. 반면 P01은 연구자가 사전에 핵심 교란 사건으로 설정한 ME17 대신 수령기관 검색이 이루어진 ME13을 자신의 경험에서 가장 어려운 사건으로 구성하였다. 연구자가 동일하게 배치한 핵심 사건이 모든 참여자의 주관적 경험에서 같은 위치의 절정으로 구성되지는 않았다.

P01이 ME13을 주관적 절정으로 구성한 결과는 같은 사건에 정박한 자기보고와 행동 자료에서도 확인되었다. P01의 Dominance는 기저선 100에서 ME13의 PAD1에서 15로 낮아졌고 Pleasure는 59에서 26으로 감소한 반면 Arousal은 67에서 74로 상승하였다. 행동 자료에서도 같은 구간에서 수령기관 목록 검색이 74회 반복되었다. 이후 사진 규격 오류가 발생한 ME17에서는 Dominance가 10까지 낮아졌지만 P01의 자유 서사에서는 오류와 입력 정보 소실 자체가 경험의 중심 사건으로 구성되지 않았다. 이는 측정값의 최소점과 참여자가 회고적으로 구성한 경험의 중심 사건이 반드시 동일하지 않을 수 있음을 보여주는 사례이다. 다만 P01의 PAD2는 최초 오류 직후가 아니라 약 66.7초 뒤 재입력이 진행된 상태에서 측정된 값이므로 ME17 직후의 직접적인 반응값으로 해석하지 않았다.

동일한 사진 규격 오류를 자유 서사에서 처리하는 방식도 참여자마다 달랐다. P02는 반복된 오류와 재시도, 해결 과정을 주요 사건으로 상세하게 연결하였다. P03은 이를 ‘조금 오류’가 있었다는 정도로 축소하여 서술하였고 P04는 오류와 입력 정보 소실을 ‘재로딩’으로 다시 명명하였다. P01은 자유 서사에서 오류와 입력 정보 소실을 중심 사건으로 다루지 않았다. P05는 사진 문제를 로그인, 주소 검색, 결말 탐색에서의 마찰과 함께 연속적으로 나열하였다.

Table 4. Cross-case comparison of narrative construction

| Participant | Construction judgment | Subjective climax                              | Narrative treatment of the photo error                              | Main attribution                      | C/P/D    |
|-------------|-----------------------|------------------------------------------------|---------------------------------------------------------------------|---------------------------------------|----------|
| P01         | NC-P                  | ME13                                           | Did not treat the error and the information loss as a central event | Address: mixed, Photo: self           | C2·P2·D0 |
| P02         | NC-S                  | ME17 (subjective peak at the third occurrence) | Described the repeated errors in detail                             | Photo: mixed, Address: system         | C3·P0·D1 |
| P03         | NC-P                  | ME17                                           | Reduced to ‘a minor error’                                          | System                                | C4·P0·D0 |
| P04         | NC-S (borderline)     | ME17                                           | Renamed as ‘reloading’                                              | Photo: self-dominant, Address: system | C4·P0·D0 |
| P05         | NC-F (borderline)     | ME17-ME19                                      | Listed multiple friction events in fragments                        | System                                | C3·P1·D0 |

Note. ‘Main attribution’ indicates whether the participant attributed the cause of the photo specification error and the address/receiving-office problem to themselves (self), the system, or both (mixed).

동일한 사건도 참여자에 따라 상세화, 축소, 재명명, 주변화 또는 파편화되는 방식으로 서로 다르게 구성되었다.

이러한 차이는 실제 수행의 어려움이나 오류 횟수와 내러티브 구성 판정이 반드시 같은 방향으로 나타나지 않았다는 점에서도 드러났다. P02는 사진 오류를 네 차례 경험하여 참여자 가운데 가장 많은 반복 실패를 보였다. 그럼에도 자유 서사에서는 문제 사건과 재시도, 해소와 종결이 연결된 응집된 구조를 구성하여 NC-S로 판정되었다. 반면 P05는 과업을 최종적으로 완료하였으나 자유 서사에서는 여러 마찰 사건이 중심 사건으로 응집되지 않았고 명확한 Resolution도 나타나지 않아 NC-F의 경계 사례로 판정되었다. 과업을 행동으로 마쳤다는 사실이나 오류 횟수만으로 이용자가 자신의 경험을 하나의 서사로 어떻게 구성하는지를 판단하기는 어려웠다.

사례 간 결과는 Table 4에 종합하였다. 동일한 과업과 동일한 최종 완료 결과를 경험하였음에도 참여자들은 어떤 사건을 중심 사건으로 선택하는지, 동일한 사건을 어떻게 명명하고 평가하는지, 문제 사건과 해소를 하나의 경험 구조로 연결하는지에서 서로 다른 양상을 보였다. 따라서 RQ3에 대해 본 연구는 동일한 객관적 과업 구조가 동일한 주관적 내러티브 구조로 이어지지 않는다는 사례 간 차이를 확인하였다. 참여자는 과업 중 경험한 사건을 선택하고 재구성하여 서로 다른 내러티브를 형성하였다. 이 결과는 5인 사례에서 개인마다 다른 내러티브가 구성되는 것을 보여주는 것이며, 구성 유형의 분포나 발생 비율을 일반화하고자 하는 것은 아니다.

#### 4.5 Integrated Analysis: Convergence and Divergence across Three Data Sources

기저선을 제외한 네 개의 PAD 측정 시점에 대해 참여자 5인의 자료를 대조하여 총 20개의 분석 단위를 구성하였다. 각 시점에서 회고적 발화를 중심으로 PAD와 행동 자료의 방향을 비교한 결과, 수렴(C)은 16건, 부분 수렴(P)은 3건, 발산(D)은 1건이었다. 참여자별 판정 결과는 Table 5와 같다.

유일한 발산은 P02의 PAD3에서 나타났다. P02는 재신청이 수용된 이후 긴장이 낮아지고 통제감(Dominance)을 회복했다고 회고하였으나 PAD에서는 Dominance가 50으로 유지되었고 Arousal도 100에 머물렀다. P02가 인터뷰에서 VAS의 드래그 방식을 충분히 이해하지 못한 채 끝점까지 이동한 경험을 설명한 점을 고려하면, 측정 인터페이스의 조작 방식이 이러한 불일치에 영향을 주었을 가능성이 있다. 다만 본 연구의 자료만으로 이를 발산의 단일

원인으로 확정하지 않았다.

부분 수렴은 자료가 서로 직접 충돌하지는 않았으나 일부 축의 변화 방향이 다르거나 해당 시점을 설명하는 회고 근거가 충분하지 않은 경우에 나타났다. 예를 들어 P05의 PAD4에서는 Dominance가 13에서 62로 상승하여 과업 종료에 따른 회복 방향을 보였다. 그러나 자유 서사에는 종료 시점의 회복을 직접 설명하는 발화가 충분하지 않아 부분 수렴으로 판정하였다.

세 자료원을 함께 대조하면서 사건 정박 측정 사이의 변화를 보완할 수 있는 사례도 드러났다. P02는 PAD2가 정박된 최초 오류 이후에도 오류를 반복하였다. 회고 과정에서는 세 번째 오류(ME17.3)를 가장 긴장이 높았던 시점으로 지목하였다. 이 사건은 PAD2와 PAD3 사이의 무측정 구간에 위치하였다. P05에서도 PAD2 이후 두 번째 오류가 발생하였고 이후 745초 동안 제출을 보류하면서 외부 도구를 탐색한 행동이 기록되었다. 이 구간 역시 PAD2와 PAD3 사이에 위치하여 정박된 자기보고만으로는 통제감(Dominance) 저하가 지속된 시간을 직접 확인하기 어려웠다.

세 자료원의 대조는 PAD 값과 회고적 발화가 서로 일치하는지를 확인하는 데 그치지 않았다. 수렴 사례에서는 서로 다른 자료가 동일한 경험 변화를 지지하였고 부분 수렴과 발산 사례에서는 단일 자료만으로는 확인하기 어려운 측정 시점의 불확실성이나 응답 특성을 드러냈다. 행동 자료와 회고적 발화는 사전에 정한 사건 정박 측정 사이에서 발생한 경험 변화도 보완하였다. 여기서 본 연구의 다중 자료 대조가 경험 변화의 수렴 여부뿐 아니라 불일치가 발생한 위치와 측정 사이에서 누락된 변화를 탐지하는 데 활용될 수 있음을 확인하였다. 다만 높은 수렴 비율이 프레임워크의 일반적 타당성을 뒷받침하는지는 현재 계획 중인 대규모 실험을 통해 검증할 예정이다.

Table 5. Convergence and divergence across three data sources

| Participant | Convergence (C) | Partial convergence (P) | Divergence (D) |
|-------------|-----------------|-------------------------|----------------|
| P01         | 2               | 2                       | 0              |
| P02         | 3               | 0                       | 1              |
| P03         | 4               | 0                       | 0              |
| P04         | 4               | 0                       | 0              |
| P05         | 3               | 1                       | 0              |
| Total       | 16              | 3                       | 1              |

## V. Conclusions

본 연구는 선행 연구에서 제안한 내러티브 기반 UX 평가 프레임워크를 실제 공공 디지털 서비스 과업에 적용하여 조작화된 분석 절차가 이용자의 경험 변화를 탐지할 수 있는지를 확인하였다. 참여자 5인이 동일한 프로토타입에서 주민등록증 재발급 과업을 수행하는 동안 PAD 자기보고와 행동 자료를 수집하였다. 과업 종료 후에는 회고적 발화를 확보하여 마이크로 이벤트를 공통 시간축으로 대조하였다.

마이크로 이벤트 단위로 조작화한 Freytag 구조의 단계는 독립 코드 사이에서 전반적으로 높은 판정 일치도를 보였다. 두 코더는 27개 이벤트 가운데 26개를 동일하게 판정하였으며 관측 일치율은 96.3%, Cohen's  $\kappa$ 는 .948이었다. 다만 초기화된 신청서로 복귀하는 ME18에서는 절정과 하강의 판정이 갈렸고 두 코더는 ME04를 모두 발단으로 판정하여 연구자의 사전 가설과 다르게 판단하였다. 사건의 기능을 기준으로 Freytag 구조의 단계를 구분하는 방식은 실제 과업의 공통 시간 구조를 조작화하는 데 적용될 수 있었다. 단계가 전환되는 일부 사건에서 확인된 판정 차이를 반영하여 복수 문제 동시 발생 시 기준 문제의 지정과 유효한 해결 행동의 정의를 포함한 보완 규칙을 제시하였으며, 이 규칙의 재현성은 후속 연구에서 확인할 과제로 남았다.

행위적 통제감을 측정하기 위해 사용한 Dominance 레이블은 대체로 과업이 자신의 의도대로 진행되는지를 판단하는 의미로 해석되었다. 특히 재신청이 수용된 이후에도 낮은 통제감(Dominance)을 보고한 사례에서는 객관적인 과업 성공과 참여자가 상황을 주도한다고 느끼는 정도가 서로 다를 수 있음을 확인하였다. 반대로 변화 폭이 작았던 사례에서는 참여자가 실제 경험을 중간 수준으로 평가했다는 설명을 회고 과정에서 들었다. 다만 한 참여자에게서는 VAS 조작 방식의 영향 가능성이 나타났다. 따라서 Dominance는 본 연구에서 사건에 따른 행위적 통제감의 변화를 탐지하는 데 활용할 수 있었다. 그러나 응답값은 해당 값을 선택한 경험적 맥락과 함께 해석해야 하며 본 연구의 결과만으로 척도의 심리측정학적 타당성을 확정할 수는 없다.

동일한 과업을 수행하고 동일하게 신청을 완료한 참여자들의 내러티브 구성도 서로 달랐다. 두 사례는 구성 성공(NC-S), 두 사례는 부분 구성(NC-P), 한 사례는 구성 실패(NC-F)로 분류하였다. 일부 사례는 인접 범주와 맞닿은 경계 사례였다. 연구자가 핵심 교란 사건으로 설정한

사진 규격 오류는 다섯 명 가운데 네 명의 주관적 절정과 대응하였지만 한 참여자는 수령기관을 찾는 ME13을 자신의 경험에서 가장 중요한 문제로 구성하였다. 동일한 사진 오류도 상세한 사건으로 서술되거나 작은 오류로 축소되었고 '재로딩'으로 다시 명명되거나 핵심 이야기에서 생략되었으며 여러 마찰 가운데 하나로 파편화되기도 하였다. 서비스가 제공하는 사건의 객관적 순서만으로는 이용자가 경험한 이야기의 구조가 결정되지 않았다. 이용자는 같은 사건을 경험하더라도 자신의 기대와 문제 해결 과정, 귀인과 통제 경험에 따라 사건의 중요도를 다르게 구성하였다.

이러한 수행 양상은 서론에서 제시한 인지적 스크립트의 관점과도 연결하여 해석할 수 있다. 대부분의 참여자는 로그인과 본인 인증 등 반복적으로 경험한 공통 절차에서는 큰 어려움을 보이지 않았으나 사진 규격과 같이 해당 과업에 특수한 요건에서는 시행착오를 경험하였다. 일부 참여자는 규격 안내를 제출 전에 확인했음에도 이를 실제 수정 행동으로 연결하지 못하고 오류를 반복하였다. 다만 본 연구는 인지적 스크립트의 보유 여부를 직접 측정하지 않았다. 이 결과로 과업 특정 스크립트의 부재를 입증할 수는 없으며 관찰된 수행 양상을 설명하는 이론적 관점으로 한정해야 한다.

연구 초기의 조작화에서는 객관적인 사건 구조가 이용자 경험을 비교하기 위한 공통 시간축으로 기능할 것으로 보았다. 실제 결과는 이 전제를 부분적으로 지지하면서도 그 범위를 보다 명확하게 만들었다. 객관적인 사건 구조는 사례를 비교하기 위한 공통 시간축으로 기능했지만 이용자의 주관적 절정과 항상 일치하지는 않았다. 행동으로 과업을 마친 시점과 통제감(Dominance)을 회복한 시점도 반드시 일치하지 않았다. 과업의 객관적 완료와 이용자가 경험하는 통제감(Dominance)의 회복은 서로 구분하여 분석해야 한다.

PAD, 행동 자료와 회고 서사를 함께 비교한 결과에서도 단일 자료원으로는 확인하기 어려운 차이가 나타났다. 20개 분석 단위 가운데 16개는 수렴, 3개는 부분 수렴, 1개는 발산이었다. 이 수치는 다중 자료 대조가 자료 간 불일치의 위치를 탐지할 수 있음을 보여주며, 프레임워크의 일반적 타당성은 대규모 실험을 통해 검증할 예정이다. 수치만으로 설명하기 어려운 결과도 회고적 발화와 행동 자료를 함께 분석함으로써 의미를 구체화할 수 있었다. 클릭, 반복 검색, 이동 경로와 지연 같은 행동 자료는 문제가 발생한 위치를 파악하는 데 유용했지만 그 행동이 이용자에게 어떤 경험으로 받아들여졌는지를 단독으로 설명하지는 못하였다. 본 프레임워크는 하나의 지표로 이용 경험을 대

표하는 대신 사건별 자기보고, 실제 행동과 이용자의 의미 구성을 대조하여 자료 간 수렴과 발산을 탐지하는 방식으로 기능하였다.

본 연구의 학술적 기여는 세 가지로 요약할 수 있다. 이용 경험의 내러티브를 사후 서술 자료에 한정하지 않고 마이크로 이벤트라는 공통 시간축을 가진 분석 대상으로 조작화하였다. 과업의 행동적 완료와 통제감(Dominance)의 회복이 분리될 수 있음을 사건 단위에서 포착하였다. 사건 정박 자기보고, 행동 자료와 회고적 발화의 대조가 자료 간 수렴의 확인에 그치지 않고 불일치가 발생한 위치와 측정 사이에서 누락된 변화를 탐지할 수 있음을 사례 수준에서 보여주었다. 이 세 가지는 프레임워크의 일반적 타당성이 아니라 실제 과업에서의 조작 가능성과 탐지 기능에 관한 기여이다. 실무적으로는 디지털정부 서비스 UI/UX 가이드라인(KRDS)과 같은 가이드라인 기반 점검과 과업 완료 여부의 확인에 더하여 이용자가 실제 수행 과정에서 다음 행동을 예측하고 진행을 주도할 수 있었는지도 함께 평가해야 한다는 점을 시사한다. 본 연구의 방법은 가이드라인 기반 평가를 대체하지 않는다. 실제 이용자가 어느 사건에서 진행 가능성을 잃고 통제감(Dominance)이 흔들리는지, 어떤 사건을 자신의 경험에서 중요하게 구성하는지를 확인하는 보완적 평가 방법으로 활용할 수 있다. 사건에 정박한 Dominance와 회고적 발화를 행동 자료와 함께 분석하면 반복, 정체, 이탈과 완료의 의미를 이용자의 경험 관점에서 해석하는 보완적 근거도 얻을 수 있다.

본 연구는 제한된 사례를 대상으로 한 초기 검증이라는 점에서 한계가 있으며, 이는 연구 범위, 연구 방법, 측정 도구의 세 가지 측면으로 구분할 수 있다. 연구 범위 측면에서는 30~49세 참여자 5인, 정부24의 주민등록증 재발급이라는 단일 과업, 실제 서비스가 아닌 프로토타입 환경을 대상으로 하였으므로 결과를 다른 연령대, 다른 서비스와 과업, 실제 운영 환경으로 직접 확대하기 어렵다. 연구 방법 측면에서는 모든 참여자가 사전 규격 오류를 경험하여 실험 설계에서 가능했던 오류 비경험 경로와 직접 비교하지 못하였고, 두 번째 오류 이후 모든 참여자에게 동일한 연구자 안내를 한 차례 제공하였으므로 이후의 문제 해결과 통제감(Dominance) 변화를 이용자의 자발적인 해결 과정만으로 해석할 수 없다. 측정 도구 측면에서는 PAD를 사전에 정한 사건에만 측정하여 반복 오류 사이에서 발생한 세부적인 상태 변화를 직접 측정하지 못하였고, VAS 조작 방식과 반복 자기보고의 반응성이 응답값에 영향을 미쳤을 가능성이 있으며, Dominance 척도의 신뢰도와 타당도는 검증하지 않았다.

이러한 한계에 대응하여 후속 연구에서는 다양한 연령대와 복수의 공공 서비스 과업을 포함한 대규모 표본에서 본 연구에서 관찰된 패턴의 반복 가능성을 확인하고, 연구자 개입이 없는 조건과 오류 경험 조건, 오류 비경험 조건을 함께 구성하여 비교할 예정이다. 또한 Dominance 척도의 신뢰도와 타당도를 검증하고, 사건 정박 다중 측정과 과업 종료 후 단일 측정을 비교하여 두 방법이 경험 변화의 위치와 크기를 어느 정도 다르게 탐지하는지도 검증할 예정이다. 반복 오류와 같이 주관적 절정이 사전에 설정한 측정 지점에서 이동할 수 있는 상황에서는 사건의 반복 횟수나 행동 변화에 따라 측정 시점을 조정하는 방법도 함께 살펴볼 예정이다.

종합하면 본 연구는 공공 디지털 서비스의 이용 경험을 과업의 완료 결과나 하나의 사후 평가값으로 환원하지 않고 사건의 진행 구조, 행위적 통제감의 변화, 실제 행동과 이용자가 구성한 내러티브를 동일한 시간축에서 함께 분석할 수 있음을 확인하였다. 이는 프레임워크의 일반적 타당성을 확인한 결과가 아니라 실제 과업에서 조작 가능성과 탐지 기능을 확인한 초기 검증이다. 그럼에도 동일한 과업을 완료한 이용자들이 서로 다른 사건을 중요하게 구성하고 행동으로 과업을 마친 시점과 통제감(Dominance)을 회복한 시점이 분리될 수 있음을 포착했다는 점에서 향후 과정 중심의 공공 디지털 서비스 UX 평가를 정교화하기 위한 기초를 제공한다.

## REFERENCES

- [1] 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.
- [2] 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/muu009
- [3] P. Herd, and D. P. Moynihan, "Administrative Burden: Policymaking by Other Means," Russell Sage Foundation, pp. 1-360, 2018.
- [4] J. Bruner, "The Narrative Construction of Reality," *Critical Inquiry*, Vol. 18, No. 1, pp. 1-21, October 1991. DOI: 10.1086/448619
- [5] J. McCarthy, and P. Wright, "Technology as Experience," The MIT Press, pp. 1-211, 2004. DOI: 10.7551/mitpress/6687.001.0001
- [6] S. Jang, and H. Ryu, "Conceptual Validity Verification of a Narrative-Based User Experience Evaluation Framework: Expert

- Validation of Freytag's Pyramid and PAD-Heatmap Integration," *Journal of The Korea Society of Computer and Information*, Vol. 31, No. 5, pp. 179-195, May 2026. DOI: 10.9708/jksci.2026.31.05.179
- [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] G. Freytag, "Die Technik des Dramas," S. Hirzel, pp. 1-303, 1863.
- [9] A. Mehrabian, and J. A. Russell, "An Approach to Environmental Psychology," The MIT Press, pp. 1-266, 1974.
- [10] 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
- [11] K. A. Ericsson, and H. A. Simon, "Protocol Analysis: Verbal Reports as Data," Revised Ed., The MIT Press, pp. 1-443, 1993. DOI: 10.7551/mitpress/5657.001.0001
- [12] M. Hertzum, "Concurrent or Retrospective Thinking Aloud in Usability Tests: A Meta-Analytic Review," *ACM Transactions on Computer-Human Interaction*, Vol. 31, No. 3, Article 37, pp. 1-29, June 2024. DOI: 10.1145/3665327
- [13] Z. Guan, S. Lee, E. Cuddihy, and J. Ramey, "The Validity of the Stimulated Retrospective Think-Aloud Method as Measured by Eye Tracking," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*, pp. 1253-1262, Montreal, Canada, April 2006. DOI: 10.1145/1124772.1124961
- [14] W. Labov, and J. Waletzky, "Narrative Analysis: Oral Versions of Personal Experience," in J. Helm (Ed.), *Essays on the Verbal and Visual Arts*, University of Washington Press, pp. 12-44, 1967.
- [15] W. Labov, "The Transformation of Experience in Narrative Syntax," in *Language in the Inner City: Studies in the Black English Vernacular*, University of Pennsylvania Press, pp. 354-396, 1972.
- [16] R. K. Yin, "Case Study Research and Applications: Design and Methods," 6th Ed., SAGE Publications, pp. 1-319, 2018.
- [17] 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
- [18] M. Q. Patton, "Qualitative Research & Evaluation Methods: Integrating Theory and Practice," 4th Ed., SAGE Publications, pp. 1-806, 2015.
- [19] Ministry of Science and ICT, and National Information Society Agency, "2025 The Report on the Digital Divide," NIA VIII-RSE-C-25026, December 2025.
- [20] K. Malterud, V. D. Siersma, and A. D. Guassora, "Sample Size in Qualitative Interview Studies: Guided by Information Power," *Qualitative Health Research*, Vol. 26, No. 13, pp. 1753-1760, November 2016. DOI: 10.1177/1049732315617444
- [21] Digital Government Service UI/UX Guidelines and Design System(KRDS), <https://www.krds.go.kr>
- [22] X. Tian, W. Hou, and K. Yuan, "A Study on the Method of Satisfaction Measurement Based on Emotion Space," 2008 9th International Conference on Computer-Aided Industrial Design and Conceptual Design, pp. 39-43, Kunming, China, November 2008. DOI: 10.1109/CAIDCD.2008.4730515
- [23] S. Grimaldi, "Narrativity of Object Interaction Experiences: A Framework for Designing Products as Narrative Experiences," in P. Benz (Ed.), *Experience Design: Concepts and Case Studies*, Bloomsbury Academic, pp. 57-68, 2014.
- [24] 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, August 2008. DOI: 10.3758/BRM.40.3.699
- [25] 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, March 1994. DOI: 10.1016/0005-7916(94)90063-9
- [26] A. Betella, and P. F. M. J. Verschure, "The Affective Slider: A Digital Self-Assessment Scale for the Measurement of Human Emotions," *PLoS ONE*, Vol. 11, No. 2, e0148037, February 2016. DOI: 10.1371/journal.pone.0148037
- [27] K. Hornbæk, and E. L.-C. Law, "Meta-Analysis of Correlations Among Usability Measures," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*, pp. 617-626, San Jose, USA, April 2007. DOI: 10.1145/1240624.1240722
- [28] J. Cohen, "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement*, Vol. 20, No. 1, pp. 37-46, April 1960. DOI: 10.1177/001316446002000104
- [29] U. H. Graneheim, and B. Lundman, "Qualitative Content Analysis in Nursing Research: Concepts, Procedures and Measures to Achieve Trustworthiness," *Nurse Education Today*, Vol. 24, No. 2, pp. 105-112, February 2004. DOI: 10.1016/j.nedt.2003.10.001
- [30] 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 the B.S., M.S. and Ph.D. degrees in Industrial engineering and Psychology from Hanyang University, KAIST, and University of York in 1997, 2000 and 2003, respectively.

Professor Ryu founded the Imagine Lab at Hanyang University, Seoul, Korea, in 2011. He is currently a Professor in the Graduate school of technology & innovation management, Hanyang University. He is interested in Cognitive Information Design, New Product Development, New Service Development, Learning Technology, Game-based Learning, Medical HCI.