

# 모바일 어플리케이션의 사용성 평가 방법 비교 분석

## Comparison of Usability Evaluation Methods for Mobile Application

박주환<sup>1</sup>, 한성호<sup>1\*</sup>, 박재현<sup>1</sup>, 박원규<sup>1</sup>, 김현경<sup>1</sup>, 홍상우<sup>2</sup>

<sup>1</sup>포항공과대학교 산업경영공학과, 포항, 790-784

<sup>2</sup>SK플래닛 성장전략실, 서울, 100-999

### ABSTRACT

최근 스마트폰, 태블릿PC 등의 모바일 기기가 널리 보급됨에 따라 모바일 어플리케이션의 사용성을 평가하고 개선하기 위한 다양한 시도가 이루어지고 있다. 어플리케이션은 기획 및 개발 기간이 다른 제품이나 서비스에 비해 짧은 만큼 쉽고 빠르게 사용성 문제점을 도출해야 할 필요가 있다. 본 연구에서는 모바일 어플리케이션 개발 단계에서 활용 가능한 사용성 평가 방법에 대한 비교 분석을 수행하였다. 이를 위해 실제 개발 중인 모바일 어플리케이션을 대상으로 4인의 전문가가 참여하는 휴리스틱 평가 방법과 30인의 사용자가 직접 참여하는 대표작업 평가 방법을 수행하였다. 사용성 평가 결과 휴리스틱 평가와 대표작업 평가에서 각각 140개와 66개의 중요 사용성 문제점이 도출되었고, 이를 13가지의 사용성 원칙과 13가지의 화면 구성요소를 기준으로 분류하고 분석하였다. 분석 결과 전문가는 정보제공성과 관련된 문제점을, 그리고 사용자는 효율성 및 친숙성과 관련된 문제점을 주로 기술한 것으로 나타났다. 본 연구의 결과는 추후 모바일 어플리케이션의 사용성 평가에 참고자료로 사용될 수 있을 것으로 기대된다.

Keywords: Mobile application, Usability Testing, Heuristic Evaluation, Benchmarking test

## 1. 서론

사용성은 대상 사용자 그룹이 주어진 환경에서 특정 작업을 효율적이고, 효과적이며, 주관적으로 만족스럽게 수행할 수 있는 정도를 나타낸다(ISO, 1993; Shackel, 1991). 사용성은 사용자 경험을 구성하는 주요한 구성 요소로 알려져 있으며(Park et al., 2011), 제품 경쟁력을 높여준다(Kwahk and Han, 2002). 이미 사용성 평가를 위한 다양한 방법이 개발되어 있으며(Kim and Han, 2008), 전문가에 의한 휴리스틱 평가 방법(Heuristic Evaluation), 인지적 시찰법(Cognitive Walkthrough), 사용자에 의한 대표작업 평가(Benchmarking Test), 설문(Questionnaire), 인터뷰(Interview) 등의 방법이 주로 사용된다. 사용성 평가는 제품 개발의 초기와 마무리 단계에서 반복적으로 이루어지며(Hix and Hartson, 1993), 제품 개선 작업과 밀접하게 연관을 맺는다.

최근 모바일 기기의 보급으로 관심을 받고 있는 어플리케이션의 경우에도 물론 다른 제품, 서비스와 같이 개발된 사용성 평가 방법을 적용하면 다양한 문제점을 도출할 수 있다. 하지만 다른 제품, 서비스에 비해 기획 및 개발 기간이 짧은 만큼 효율적이고 효과적인 사용성 평가 방법에 대한 요구가 절실하다. 대부분의 어플리케이션은

사용성 평가 절차 없이 개발되는 것으로 알려져 있다. 그동안 사용성 평가 방법 자체를 비교하는 연구가 없었던 것은 아니지만 평가 환경을 비교하거나 새로운 방식의 검증에 주안점을 둔 한계가 있다(Kaikkonen et al., 2005).

본 연구에서는 어플리케이션 개발 시 사용성 문제점을 도출하는 방법 중 각각 전문가와 사용자가 참여하는 휴리스틱 평가와 대표작업 평가 방법을 비교하였다. 대표작업 평가의 경우 비디오 회상법(Retrospective Videotaping)을 사용하여 주어진 작업 수행에 방해가 되지 않도록 했다. 본 연구의 결과는 추후 사용성 문제점 도출 방식 선택 시 참고 자료로 활용될 수 있을 것으로 기대된다.

## 2. 사용성 문제점 도출 방법

본 연구에서는 모바일 어플리케이션 개발 과정에서 사용될 수 있는 두 가지 사용성 문제점 도출 방법을 비교하였다. 그 대상으로 사용성 문제점 도출에 일반적으로 많이 사용되는 전문가 휴리스틱 평가와 사용자 대표작업 평가를 선정하였다.

휴리스틱 평가 방법은 사용성 평가 전문가가 제품과 관련된 사용성 원칙을 기준으로 사용성 문제점을 도출하는 방법이다(Bowman et al., 2001; Nielsen, 1992; Nielsen and Molich, 1990). 상대적으로 적은 숫자의 평가자로 다수의 중요한 사용성 문제를 도출할 수 있는 것으로 알려져 있으며(Yen et al., 2009), 빠르고 효율적으로 수행할 수 있다는 장점이 있다(Te'eni et al., 2007). 평가 시 기준이 되는 원칙에 좌우되는 측면이 있지만 많은 학자들이 사용성 원칙에 대해 연구하여 왔고 어느 정도 합의가 이루어졌기 때문에(Nielsen, 1992), 사용성 원칙보다는 전문가의 역량이 중요한 변수로 작용한다.

대표작업 평가 방법은 실제 사용자가 제품, 혹은 어플리케이션을 사용할 때 접하게 될 것으로 예상되는 대표적인 작업을 분류하고 선정 후, 이를 수행해보도록 함으로써 문제점을 파악하는 기법이다(Hix and Hartson, 1993). 이 기법은 전문가의 의견이 아니라 실제 사용자의 요구사항을 수집할 수 있다는 데에 의의가 있으며, 평가의 목적과 시점에 따라 수많은 변형된 방법이 존재한다. 참여자가 자신의 생각을 그대로 말로 읊기는 발성 사고법(Think Aloud)은 이 중 하나로 볼 수 있다(Haak et al., 2003; Some ren et al., 1994). 사용성 문제점을 직접 도출할 때에는 비디오 회상법 또한 자주 사용된다. 비디오 회상법은, 작업 수행 후 실험 참여자가 녹화된 동영상을 통해 작업 화면 및 자신의 행동을 회상하고, 느꼈던 사용성 문제점과 이에 대한 생각을 직접 말하는 방식으로 진행된다.

### 3. 사례 연구

모바일 어플리케이션의 실제 개발 과정에 휴리스틱 평가와 대표작업 평가 방법을 적용해 보았다. 사례 연구를 통해 두 방법의 우열을 비교하는 데에는 무리가 있지만 각 방법의 효용과 특색에 대한 단서를 얻을 수 있었다.

#### 3.1. 평가 대상 어플리케이션

대상 어플리케이션은 스마트폰에 설치되어 동작하며, 정보의 수집 및 작성을 핵심 기능으로 한다. 이를 이용하여 사용자가 원하는 단어를 등록하면 관련된 인터넷 정보를 정기적으로 다운로드 받고 상세 내용을 확인할 수 있다. 사용자의 조작이 많이 들어가고 기능과 인터페이스가 복잡하기 때문에 사용성 측면에서의 검토가 필요하였다. 개발 초기 단계로서 단일 어플리케이션으로 구현되지는 않았지만, 버튼의 배치, 크기, 모양 등의 화면 구성요소에 대한 정의와 화면 흐름 및 전환에 대한 정의가 완료된 상태였다.

#### 3.2. 평가 방법

평가는 크게 전문가 평가와 사용자 평가로 나뉘어 수행되었다. 앞에서 언급한 바와 같이 전문가 평가의 경우 휴리스틱 평가 방법을 사용하였고 사용자 평가의 경우 대표작업 평가를 사용하였다. 대표작업 평가 방법에는, 발성 사고법과 비디오 회상법이 사용되었다.

먼저 전문가 평가의 경우, HCI를 전공한 전문가 4인이 13가지 사용성 원칙(표 1 참조)을 이용하여 미리 정의된 각 화면 별로 13가지의 화면 구성요소(표 2 참조)를 평가하였다. 사용성 문제점을 도출하고 개선안 작성 시 우선순위 선정에 참고할 수 있도록 문제점의 심각도를 0-4점 척도로 평가하였다.

실제 사용자 평가의 경우 어플리케이션의 주 사용자층으로 예상되는 젊은 사용자 30명을 대상으로 수행되었다. 이들의 평균 연령은 24.3세(표준편차: 4.2세)로 다양한 지역에 거주하는 참여자를 섭외하였으며 남자는 19명, 여자는 11명이 참여하였다. 참여자는 어플리케이션의 기획자, 개발자 및 사용성 평가 전문가의 논의로 도출된 6개의 대표작업을 수행하고, 각 작업이 끝난 후에 촬영된 영상을 보며 사용성 문제점을 지적하였다. 마찬가지로 도출된 문제점 별로 0-4점의 심각도 평가를 수행하였다.

#### 3.3. 평가 결과 및 분석

전문가와 사용자가 참여한 휴리스틱 평가와 대표작업 평가 결과 각 213개와 372개의 잠재적 문제점이 도출되었다. 그리고 유사하거나 동일하다고 판단되는 문제점을 통합하고, 평균 심각도 수준이 3점 이상인 문제의 수는 휴리스틱 평가와 대표작업 평가가 각 140개와 66개였다.

휴리스틱 평가의 경우에는 접근성(Accessibility), 정보제공성(Informativeness), 단순성(Simplicity), 가시성(Visibility) 등의 원칙과 관련이 있는 것으로 나타났고, 대표작업 평가의 경우에는 효율성(Efficiency), 친밀성(Familiarity), 예측성(Predictability), 가시성 등의 원칙과 관련이 있는 것으로 나타났다(표 1 참조).

예를 들어 “특정 버튼의 크기가 작아 선택 하기 어려움(접근성)”, “색을 통해 콘텐츠를 구분할 수 있는데 의미와 색 사이 연관성이 적음(정보제공성)”, “특정 화면에 표시되는 정보의 종류가 많고 복잡(단순성)”, “특정 버튼이 배경 이미지와 겹쳐서 가시성이 떨어짐(가시성)”, “특정 아이콘/심볼의 모양이 익숙하지 않음(친밀성)” 등의 문제점이 도출되었다.

관련된 화면 구성요소 별로 분석해보면(표 2 참조), 기능(Function), 레이아웃(Layout), 화면전환(Navigation) 측면에서의 문제점이 공통적으로 많이 도출된 것을 알 수 있

다. 휴리스틱 평가에서는 정보의 계층적 구조(Structure), 대표작업 평가에서는 아이콘(Icon)과 관련된 문제가 부각되기도 했다. 특히 아이콘에 대한 사용성 문제점으로 “클립 모양의 아이콘은 기존에 ‘첨부’ 기능에 많이 쓰이기 때문에 ‘가져오기’ 기능에 사용하는 것은 혼란스럽다”, “콘텐츠를 ‘스크랩’할 때 쓰이는 버튼이 다운로드 버튼처럼 보여서 혼란스럽다” 등과 같은 의견이 있었다.

참고로, 사용자는 사용성 원칙이나 화면 구성요소 목록에 대한 정보 없이, 문제점을 설명하도록 요구 받았으며, 표 1과 표 2는 사용성 문제점을 정리/통폐합한 후 원칙과 구성요소 측면에서 분류한 결과를 수록하였다.

표 1. 사용성 원칙 별 문제점 분포

사용성 원칙	사용성 문제점 개수 (비율, %)	
	휴리스틱 평가	대표작업 평가
접근성 (Accessibility)	19 (13.6)	4 (6.1)
일관성 (Consistency)	13 (9.3)	1 (1.5)
직접성 (Directness)	1 (0.7)	0 (0.0)
효율성 (Efficiency)	12 (8.6)	16 (24.2)
친밀성 (Familiarity)	4 (2.9)	15 (22.7)
피드백 (Feedback)	8 (5.7)	1 (1.5)
유연성 (Flexibility)	2 (1.4)	0 (0.0)
오류수용성 (Forgiveness)	3 (2.1)	1 (1.5)
정보제공성 (Informativeness)	32 (22.9)	6 (9.1)
예측성 (Predictability)	11 (7.9)	8 (12.1)
단순성 (Simplicity)	16 (11.4)	3 (4.6)
사용자 조작 (User control)	4 (2.9)	0 (0.0)
가시성 (Visibility)	15 (10.7)	11 (16.7)
합계	140 (100.0)	66 (100.0)

표 2. 화면 구성요소 별 사용성 문제점 분포

화면 구성요소	사용성 문제점 개수 (비율, %)	
	휴리스틱 평가	대표작업 평가
색 부호화 (Coding)	3 (2.1)	4 (6.1)
피드백 (Feedback)	12 (8.6)	2 (3.0)
문자 (Font)	1 (0.7)	2 (3.0)
기능 (Function)	32 (22.9)	10 (15.2)
아이콘 (Icon)	5 (3.6)	18 (27.3)
레이아웃 (Layout)	26 (18.6)	11 (16.7)
명명 (Naming)	11 (7.9)	1 (1.5)
화면전환 (Navigation)	24 (17.1)	7 (10.6)
검색창 (Search box)	1 (0.7)	1 (1.5)
바로가기 (Shortcut)	2 (1.4)	5 (7.6)
크기 (Size)	2 (1.4)	0 (0.0)
구조 (Structure)	18 (12.9)	5 (7.6)
유형 (Type)	3 (2.1)	0 (0.0)
합계	140 (100.0)	66 (100.0)

## 4. 토의

전문가와 사용자가 도출한 문제점은 대체로 유사했지만 서로 다른 측면을 중요시하는 것으로 나타났다. 전문가가는 한 화면에서 정보를 얼마나 짜임새 있고 구체적으로 정확하게 제시하는가와 관련된 문제점을 많이 도출한 반면, 사용자는 작업을 얼마나 효율적으로 수행할 수 있는가, 그리고 자신에게 얼마나 친숙한가를 기준으로 문제를 설명하였다. 이는 사용자가 기술한 문제의 약 27%가 아이콘 모양에 대한 것으로 대부분 익숙하지 않은 모양에 대한 불편함을 표현하는 것으로도 설명할 수 있다.

정량적인 결과를 바탕으로 판단하기는 어렵지만 위와 같이 전문가와 사용자 평가 결과 나타나는 문제점을 사용성 원칙과 화면 구성요소 별로 분류하면 서로 다른 양상으로 나타난다. 이로 미루어 소규모 전문가 평가가 사용자 대상의 평가를 완전히 대체하기 어렵다는 사실을 알 수 있다. 그 역도 성립하며, 결론적으로 두 가지 방법은 상호 보완적으로 수행될 수 있는 것으로 분석된다. 단, 두 가지 방법의 비용과 노력 측면에서 비교하였을 때에는 4인의 전문가가 참여한 평가가 사용자 평가에 비해 더 효율적임에는 의심할 여지가 없다.

일반적으로 대표작업 평가 자체에 필요한 인원수에 대한 명확한 지침은 없다. 목적이나 평가 결과의 통계 분석 여부에 따라 달라질 수 있지만, 사용자를 대상으로 한 평가의 경우 사용성 문제점을 포괄적으로 찾기 위해서는 약 10명의 평가자가 필요하다는 연구 결과가 있다(Hwang and Salvendy, 2010; Lewis, 1994; Virzi, 1992). 이 점을 고려하여 본 연구에서보다 많은 약 10명 정도의 전문가를 활용하게 되면 더 좋은 결과가 나올 수 있음을 추측해볼 수 있다.

또한 본 연구에서는 평균 심각도가 3이상인 사용성 문제점을 일괄적으로 분석하였는데, 전문가와 사용자 간 평가 기준 차이를 고려하지 못한 한계가 있음을 덧붙인다.

## 5. 결론

본 연구에서는 실제 모바일 어플리케이션 개발 과정에서 전문가가 참여하는 휴리스틱 평가 방법과 사용자가 참여하는 대표작업 평가 방법을 적용해보았으며 결과로 도출된 사용성 문제점을 바탕으로 두 가지 방법을 비교해보고자 하였다. 사용성 원칙과 화면 구성요소를 기준으로 도출된 문제점을 분류하면 서로 상이한 분포가 나타났다. 이로 미루어 두 가지 방법이 상호 보완적으로 수행되는 것이 이상적인 대안임을 알 수 있다. 또한, 전문가 휴

리스트릭 평가법이 중요 사용성 문제점을 도출하는 과정에서 더 효율적이라는 사실을 재확인하였다. 본 연구의 결과물은 향후 모바일 어플리케이션의 사용성을 평가에 참고 자료로 활용될 수 있을 것으로 기대된다.

## 참고 문헌

- Bowman, D.A., Gabbard, J.L. and Hix, D., A survey of usability evaluation in virtual environments: classification and comparison of methods, *Presence: Teleoperators and Virtual Environments*, 11(4), 404-424, 2002.
- Haak, M.J.V.D., Jong, M.D.T., and Schellens, P. J., Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue, *Behaviour & Information Technology*, 22(5), 339-351, 2003.
- Hix, D. and Hartson, H.R., *Developing user interfaces: Ensuring usability through product and process*, John Wiley and Sons, 1993.
- Hwang, W. and Salvendy, G., Number of people required for usability evaluation: the 10±2 rule, *Communications of the ACM*, 53(5), 130-133, 2010.
- Kaikkonen, A., Kallio, T. and Kankainen, A., Usability Testing of mobile applications: a comparison between laboratory and field testing, *Journal of Usability Studies*, 1(1), 4-16, 2005.
- Kim, J. and Han, S.H., A methodology for developing a usability index of consumer electronic products, *International journal of industrial ergonomics*, 38(3-4), 333-345, 2008.
- Kwahk, J. and Han, S.H., A methodology for evaluating the usability of audiovisual consumer electronic products, *Applied Ergonomics*, 33(5), 2002.
- Lewis, J.R., Sample Sizes for Usability Studies: Additional Considerations, *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 36(2), 368-378, 1994
- Nielsen, J., "Finding usability problems through heuristic evaluation", *Proceedings of the SIGCHI conference on Human factors in computing*, Monterey, CA, USA, 1992.
- Nielsen, J. and Molich, R., "Heuristic evaluation of user interfaces", *Proceedings of the SIGCHI conference on Human factors in computing*, Seattle, WA, USA, 1990.
- Park, J., Han, S.H., Kim, H.K., Cho, Y. and Park, W., Developing elements of user experience for mobile phones and services: survey, interview, and observation approaches, *Human Factors and Ergonomics in Manufacturing*, in Press, 2011.
- Shackel, B., Usability – context, framework, design and evaluation. *Human Factors for Informatics Usability*. Cambridge University Press, 1991.
- Someren, M.W., Barnard, Y.F. and Sandberg, J.A.C., *The think aloud method: a practical guide to modelling cognitive processes*, Academic Press, 1994.

Te'eni, D., Carey, J. and Zhang, P., *Human computer interaction – developing effective organizational information systems*, John Wiley and Sons, 2007.

Virzi, R.A., Refining the test phase of usability evaluation: how many subjects is enough?, *Human Factors*, 34 (4), 1992.

Yen, P.Y. and Bakken, S., "A comparison of usability evaluation methods: heuristic evaluation versus end-user think-aloud protocol – an example from a web-based communication tool for nurse scheduling", *AMIA 2009 Symposium Proceedings*, 714-718, 2009.

## Author listings

**Joohwan Park:** pkjhwan@postech.ac.kr

**Highest degree:** BS, Dept. of Ind. & Mgmt. Eng., POSTECH  
**Position title:** PhD Candidate, Dept. of Ind. & Mgmt. Eng., POSTECH  
**Areas of interest:** User Interface, HCI, User experience, Intelligent UI

**Sung H. Han:** shan@postech.ac.kr

**Highest degree:** PhD, Ind. & Sys. Eng. Dept., Virginia Tech (VT)  
**Position title:** Professor, Dept. of Ind. & Mgmt. Eng., POSTECH  
**Areas of interest:** Human-Computer Interaction, Usability Engineering, Affective Product Design, Intelligent User Interfaces, User Experience

**Jaehyun Park:** parkdog3@postech.ac.kr

**Highest degree:** BS, Dept. of Ind. & Mgmt. Eng., POSTECH  
**Position title:** PhD Candidate, Dept. of Ind. & Mgmt. Eng., POSTECH  
**Areas of interest:** User Experience, User Value, Human-Computer Interaction, Affective Engineering

**Wonkyu Park:** p09plus1@postech.ac.kr

**Highest degree:** PhD, Dept. of Ind. & Mgmt. Eng., POSTECH  
**Position title:** PhD, Dept. of Ind. & Mgmt. Eng., POSTECH  
**Areas of interest:** User Interface, Mobile UI, User experience

**Hyunkyung Kim:** emokubi@postech.ac.kr

**Highest degree:** BS, Dept. of Ind. & Mgmt. Eng., POSTECH  
**Position title:** PhD Candidate, Dept. of Ind. & Mgmt. Eng., POSTECH  
**Areas of interest:** User Experience, HCI, Affective Engineering, Context awareness

**Sang Woo Hang:** sanghong@sk.com

**Highest degree:** PhD, Dept. of Ind. & Mgmt. Eng., POSTECH  
**Position title:** manager, 성장전략팀, SK플래닛  
**Areas of interest:** UX Design, Service Design, Design Ethnography, UX Quality Assurance, Biz Strategy