

초·중학생 대상 웹기반 진단평가관리시스템 사용성 개선을 위한 휴리스틱스 개발 및 검증

차현진* · 황윤자**

온교육디자인연구소* · 단국대학교 공학교육혁신센터**

요 약

4차 산업혁명 시대를 맞이하여 증거기반의 교육정책을 위해, 전국단위의 진단평가 데이터를 수집하기 위한 시스템을 구축·활용하고 있다. 본 연구는 이러한 전국 단위 초·중생이 활용하고 있는 진단평가관리시스템의 사용성 개선을 도울 수 있는 휴리스틱스 개발을 목적으로 하였다. 이를 위해 Nielsen의 기본 휴리스틱스를 비롯하여 교육적 목적에 종속적으로 개발된 휴리스틱스에 대한 선행연구를 기반으로 유사시스템인 진단·보정학습시스템의 사용성 문제를 분석함으로써 휴리스틱스 초안을 개발하였다. 이는 델파이 방법을 통해 15명의 전문가에 의해 2회의 타당화 과정을 수행함으로써 수정·보완되었고 최종안이 도출되었다. 본 연구는 교육 맥락에서 사용성 평가 도구를 제공함으로써 전국단위의 관계자(교사 등)와 학생들이 활용하는 시스템의 사용성 개선을 위한 기반을 마련하였다는데 의의가 있다.

키워드 : 휴리스틱스, 사용성, 교육진단평가, 사용성 평가 도구, 웹기반평가시스템

Development and validation of heuristics for usability improvements on web-based diagnostic assessment management system for primary-middle schoolers

Hyunjin Cha*, YunJa Hwang**

ON Education Design Research Institute*,

Center for Innovative Engineering Education, Dankook University**

ABSTRACT

Preparing the 4th industrial revolution age, a diagnostic assessment management system which can collect nationwide data has been implemented and utilized for the development of the evidence-based educational policy. This study aims to develop heuristics to help to improve usability of such systems, which are utilized by primary and middle school community including teachers and students in the nationwide. To achieve the objective, previous studies related to the development of heuristics in the educational contexts as well as basic heuristics from Nielsen were reviewed, and usability problems of the web-based diagnostic-supplement system being currently

교신저자 : 황윤자(단국대학교)

논문투고 : 2017-11-06

논문심사 : 2017-11-14

심사완료 : 2017-12-01

utilized, as a sample system, were analyzed. Then, the first version of heuristics was developed. The developed heuristics were validated and revised through 2 rounds of the delphi methods with 15 experts. This study has an implication with preparing foundations on improving usability of such diagnostic assessment management system utilized by school community at the nationwide in educational contexts.

Keywords : Heuristics, usability, educational diagnostic assessment, usability evaluation tools, web-based assessment system

1. 서론

ICT의 발달은 교육 분야에서 패러다임 변화를 가져왔으며 그 중 하나가 전국 단위의 학생들의 정보를 전산화하여 관리함으로써 증거기반의 데이터를 활용한 효율적인 교육정책을 시행하고자 하는 노력들이 있어왔다[17]. 국내에서도 유치원부터 초·중·고까지 학교에서 수행되는 학생들의 데이터를 관리하고 수합하는 NEIS 시스템을 비롯하여 다양한 학생 정보 및 교육 정보 데이터를 수집·활용함으로써 4차 산업혁명 시대에 대응하기 위한 다양한 교육 시스템을 국가단위로 개발·활용하는 노력을 지속해 오고 있다[16].

2011년부터 우리나라는 스마트 교육(SMART Education) 정책을 선도하면서 스마트 기술을 활용한 온라인 평가 체제 구축 및 활용을 세부 추진과제로 제시하였다. 2012년부터 전국 단위의 기초학력 진단 평가 방식을 도입하여 학습 부진을 조기에 예방할 수 있는 기초학력 지원 서비스를 위한 웹기반 평가 시스템을 구축하였고 2013년부터 5개 시·도를 시작으로 전국에서 기초학력 진단·보정 시스템을 활용하고 있다[24]. 일반적으로 웹기반 진단·평가 관리시스템의 특징[7][34][11]은 대규모의 학생들을 대상으로 일관적인 시험 수행, 평가문항에서 다양한 멀티미디어 자료 활용, 적응적 시험, 학생들의 평가 결과에 대한 추이 분석, 평가 후 즉각적인 결과와 피드백 제공, 평가 결과에 따른 맞춤형의 학습 자료 또는 관련 추천 정보 제공 등이 있다. 특히, 본 연구의 목적과 연계된 예제 시스템으로 기초학력 진단·보정 시스템은 매년 전국 학생들의 기초 학력 및 수준을 측정하고 학생들의 학력 수준 파악하여 개별적인 학생들의 기초학력에 대한 변화 추이를 관리하고, 학생들의 수준별 맞춤형 보정학습 자료를 제공한다. 이러한 기초학력 진단·보정 시스템은 전국 학교들의 기초학력 수준을 파악하고 기초학력 지원을

위한 우수사례 및 모델 도출을 유도할 수 있다는 장점을 가지고 있다[24].

하지만, 이와 같이 국가단위로 활용되는 시스템은 전국의 교사 및 다양한 학생들이 활용한다는 점에서 사용성은 중요한 문제 중 하나이다[24]. 학생들에게는 시험이라는 가장 주요한 과업(task)을 고려할 때, 시스템을 활용하면서 부수적 과업(task) 발생으로 인해 주요한 과업(task)이 방해를 받는다면, 위에서 언급된 웹기반 평가 시스템이 가지는 많은 장점에도 불구하고 시스템이 가지는 가장 본연의 과업을 수행하지 못해 오히려 사용자에게 심각한 문제를 일으키는 원인이 될 수 있다[26].

사용성은 사용자가 가지는 목적을 좀 더 효과적이고 효율적이며 만족적인 방법으로 성취할 수 있도록 주어진 맥락에서 제품이나 시스템을 활용하는 정도를 말한다[26][6]. 즉, 사용성이 전제되지 않는다면 기존 오프라인 상에서 수행되는 과업을 온라인으로 시스템화함으로써 가지는 장점을 극대화시키기 어려우며, 특정한 시스템의 경우 심각한 문제의 원인이 될 수 있다. Nielsen[21]은 거시적 개념에서의 사용성을 점검할 수 있는 원칙("broad rules of thumb")으로 휴리스틱스(heuristics)를 제시하였다. Nielsen에 의해 제시된 10가지 휴리스틱스는 특정 영역에 독립적으로 일반적인 사용자 인터페이스(User Interface) 설계를 위해 제시된 만큼 초기에 기본적인 컴퓨팅 시스템의 사용성을 개선하기 위해 많이 활용되었다. 하지만, 이러한 거시적 개념의 사용성 원칙을 모바일이나 특정 영역 종속적인 시스템(domain specific system)에 적용하는데 한계를 가지게 되면서 시스템이 가지는 본연의 목적을 수행하는데 필요한 특정 영역에 종속적인 휴리스틱스를 개발하는 연구들이 다양한 플랫폼을 기반으로 수행되었다[30]. 그 중에서 교육 영역에 종속적으로 개발된 사용성 평가도구 또는 휴리스틱스는 교육용 소프트웨어를 평가하기 위한 휴리스틱스 개발[32]을 비롯하여 웹사

이트 기반 학습 사이트에 대한 사용성 평가 요소[27], 이러닝 [8]과 디지털교과서 플랫폼의 사용성 평가 준거[18], LMS(Learning Management System)를 위한 휴리스틱스 [20] 등이 있다. 이는 각각의 독립적인 교육적 맥락과 플랫폼을 중심으로 사용성 평가 요소와 기준 개발을 위해 수행된 연구로 교육 시스템의 활용 목적 및 특징에 따라 사용성의 준거가 달라질 수 있음을 보여주고 있다. 특히, [28]의 연구에서는 교육적 맥락이라는 영역적 틀에서 규명되는 사용성을 pedagogical usability로 정의하고, [20]의 연구에서는 교육 분야에서 수행된 교육적 사용성 요소들을 종합하여 교육적 효과를 위한 휴리스틱스(didactic effectiveness heuristics)라는 용어로 논의하기도 하였다.

교육적 맥락 중 본 연구에서 다루고 있는 평가 시스템과 관련해서는 2009년에 컴퓨터보조평가(Computer Assisted assessment, CAA)를 위한 휴리스틱스 개발 연구[30]가 수행된 바 있다. [30]의 연구에서는 컴퓨터에서 수행되는 가장 기본적인 ‘시험(assessment)’이라는 과업을 중심으로 사용성을 평가하기 위한 기준들을 제시하였다. 하지만, [30]에 의해 개발된 휴리스틱스는 기존 LMS 시스템 상에서 ‘시험’을 수행하기 위한 기초적인 과업들을 중심으로 진행함으로써 본 연구에서 다루고자 하는 진단과 진단에 따른 관리 시스템이 가지는 특징을 갖춘 웹기반 진단·관리시스템을 다루는 데에는 한계를 가진다. 특히, 소규모 학급 단위의 평가가 아니라 전국대상 초·중(K-4에서부터 K-9)의 학생들에게 일괄적으로 ‘시험’이 적용되고, ‘시험’의 목적이 형성·총괄평가가 아니라 진단 평가라는 점, 전국 단위의 학생 데이터와 통계 관리가 필요하다는 점, 즉 학생 뿐 아니라 교사들의 관리 기능과 학생들의 맞춤형 피드백에 대한 기능들이 주요한 시스템의 과업이라는 점에서 CAA 휴리스틱스는 이러한 시스템의 사용성을 개선하기 위한 기준들로 활용하는데 한계점을 가진다. 그러므로 본 연구는 초·중학생을 대상으로 전국적으로 시행되는 진단평가 등의 과업을 위한 시스템을 개발할 때, 설계 지침과 사용성 개선을 위한 평가도구로 활용될 수 있는 휴리스틱스를 개발하고자 한다.

2. 휴리스틱스 관련 선행연구 분석

사용성에 대한 일반적인 가이드라인을 제공한 Nielsen[21]은 사용성의 요소로 5가지를 제시하였고, 사

용자와 상호작용하는 시스템 설계를 위한 10가지 원칙을 제시하였다. 또한, Nielsen은 사용성 문제를 찾아내는데 집중할 수 있는 원칙으로 휴리스틱스를 제안하였다[22]. 휴리스틱스는 국내 논문에서 평가측정 항목, 평가측정 도구 등의 용어로 풀이되어 활용되기도 하는데, 본 연구에서는 연구의 목적이 초·중등 대상 웹기반 진단 평가시스템 사용성 개선을 위해 Nielsen이 제시하고 있는 사용성을 점검할 수 있는 원칙들로 구성된 리스트를 개발하고자 하는 목적에 따라 원어의 우리말 표기인 ‘휴리스틱스’라는 용어로 사용하고자 한다.

전국적으로 시행되는 진단평가관리시스템의 초·중학생을 대상으로 휴리스틱스를 개발하기 위해 우선 기존 휴리스틱스를 분석하였다. 우선, 휴리스틱스에 관한 선행연구 중 가장 기본이 되고 사용성 평가에서 일반적으로 가장 널리 활용[19]되고 있는 Nielsen이 제시한 휴리스틱스를 분석하였다. 또한, 본 연구의 목적에 따라 교육 영역 종속적인 시스템을 평가하기 위해 개발된 휴리스틱스를 분석하였다.

2.1 Nielsen의 휴리스틱스

Nielsen[21]은 사용성이 학습용이성(learnability), 효율성(efficiency), 기억의 용이성(memorability), 오류(error), 그리고 만족감(satisfaction) 등으로 구성된다고 하였다. 특히, 사용자 인터페이스(UI) 사용성을 위한 구체적인 원칙으로 다음의 10가지 휴리스틱스를 제시하였다[22]. 첫째, 알기 쉬운 시스템 상태(사용자에게 지속적으로 시스템 상태를 알려줄 수 있는 상태 정보 제공), 둘째, 실세계와 적합한 시스템(사용자가 이해할 수 있는 언어와 개념을 사용하여 현실세계에서처럼 자연스럽게 논리적으로 정보 제공), 셋째, 사용자 주도권(사용자가 실수한 경우에도 쉽게 빠져나올 수 있고 언제든지 원하지 않는 상태를 빠져나오거나 회복할 수 있는 기능 제공), 넷째, 일관성과 표준화(사용자가 알고 있는 단어, 상태, 행동에 대한 일관성과 표준성 유지), 다섯째, 오류 방지(오류가 일어날 것을 방지하고, 오류발생시 오류 메시지를 명확히 제공), 여섯째, 기억보다는 쉽게 배울 수 있는 디자인(사용자는 사용방법을 기억함으로써가 아닌 직관적으로 사용할 수 있도록 디자인), 일곱째, 유연성과 효율성(초보자와 전문가 모두를 만족시킬 수 있는 유연하고 효율적인 디자인), 여덟째, 심미적 디자인, 아홉째, 오류를 인지하

고 회복할 수 있는 도움말, 마지막으로, 도움말 설명서 제공이다. 이 10가지 휴리스틱스는 특정 영역에 종속적인 휴리스틱스를 개발할 때 기본이 되는 가이드라인으로 활용되고 있다.

2.2 웹기반 진단·처방시스템 사용성 평가

국내에서는 2005년부터 전국적으로 정규 수업 후에 보충 수업을 위한 웹기반 사이버가정학습 시스템을 개발하여 보급하고 있다[16]. 2008년에는 사이버가정학습 시스템의 보조적 기능으로 학생들이 스스로 자신의 학습 수준을 진단하고 그에 따라서 처방적 콘텐츠를 학습할 수 있도록 도와주는 진단·처방학습관리시스템이 개발되어 보급된 바 있다[16]. [8]의 연구에서는 사이버가정학습 진단·처방학습관리시스템에 대하여 사용성 평가를 실시하기 위해 전문가 평가와 사용자를 대상으로 사용성 테스트를 수행하였다. 전문가 평가 도구와 사용성 평가 분석을 위한 사용성 원칙은 웹기반 학습 사이트 사용성 평가와 개선을 도와주기 위해 개발된 휴리스틱스[27]를 기반으로 하였다.

웹기반 학습 사이트를 위한 사용성 평가 도구 개발은 [27]의 연구에서 수행되었다. [27]에서는 선행연구 분석을 통해 평가 영역과 항목을 도출한 후, 델파이 조사와 표적집단면접법(FGI)을 실시하여 합의된 초안을 개발하였다. 다음으로 AHP(Analytic Hierarchy Process)를 활용하여 각 항목의 중요도에 따른 가중치를 도출하였다. 도출된 휴리스틱스는 크게 학습내용 구성, 학습진행 방법, 화면표현, 오류 및 도움처리의 4가지 영역에서 12~16개의 체크리스트로 구성되어 있다.

2.3 디지털교과서 사용성 평가

2015년 개정 교육과정에 따르면, 초·중등학교의 일부 교과(사회, 과학, 영어 등)에 디지털교과서가 개발·보급될 예정이다. 우리나라 정부는 2007년부터 디지털교과서에 대한 연구·개발을 지속적으로 수행해 왔으며 변화하는 플랫폼과 신기술·트렌드를 적용하여 다양한 형태의 콘텐츠 및 플랫폼을 디지털교과서에 적용·시범 운영해 오고 있다. 디지털 교과서는 미래형 교과서로 다양한 학생들에 의해 전국적으로 사용될 수 있다는 점에서 초기 디지털교과서가 개발되면서 사용성과 사용자 편의성을 높이기 위한 평가

준거 및 도구 등에 대한 연구들이 진행되었다[18][14][31]. 2008년에 정영식[14]에 의한 디지털교과서 사용성 평가 준거 개발 연구를 시작으로, 2009년에는 디지털교과서 플랫폼에 대한 사용성 평가 도구 개발 및 적용 연구[18], 어포던스 기반 디지털교과서 사용자 편의성에 미치는 요인에 대한 탐색 연구 등이 수행되었다[31].

가장 초기에 수행된 [14]의 연구에서는 문헌 조사 결과를 바탕으로 델파이 조사와 현장 적합성 평가를 통해 교과 내용, 교수 설계, 플랫폼, 사용자 등 4가지 영역을 나누고 영역별 평가 요소를 도출한 후 평가 지표를 개발하였다. [18]의 연구에서는 디지털교과서의 사용성 평가도구를 개발하기 위해, 선행연구 및 디지털교과서 사용 환경을 분석하고 전문가 자문을 통하여 도출된 평가 항목들을 델파이와 현장 사전 검토를 통하여 타당성 검증 실시하였다. 최종적으로 도출된 사용성 평가도구는 교사와 학생들을 대상으로 설문 형식의 도구를 개발하였으며, 인터페이스, 교수·학습 지원, 기술적 안정성의 3가지 영역에서 사용자 기대, 사용자 통제, 학습(교수) 지원, 상호작용, 사용자 지원 및 보편적 접근성, 기술적 안정성의 6가지 하위 요소들을 도출한 후 각각 요소별 세부 항목들을 사용자별로 개발하였다. [31]의 연구에서는 어포던스 관점에서 어떠한 요인이 디지털 교과서의 사용자 편의성에 영향을 미치는지 알아보기 위한 탐색적 연구가 시행되었는데, ‘인터페이스 디자인’, ‘상호작용적 멀티미디어’, ‘인지적 조작’, ‘물리적 조작’ 요인의 순서대로 영향력을 미치는 것으로 나타났다.

2.4 교육적 효과성 휴리스틱스

Mtebe & Kissaka[19]는 LMS(Learning Management System) 평가 휴리스틱스 개발을 위해, 그동안 교육 분야에서 수행된 사용성에 관한 연구를 집약하여 교육적 효과성(didactic effectiveness)이라는 용어를 정의하고, 교육적 효과성을 평가하기 위한 휴리스틱스를 제안하였다. Mtebe & Kissaka[19]는 교육용 소프트웨어[32], 교육용 멀티미디어[3], 이러닝[4], 디지털 자료에서의 교육적 사용성(pedagogical usability)에 대한 연구[23] 등에서 교육적 맥락에서 사용성 평가를 위해 도출된 준거 및 평가 항목, 기준 등을 집약하여 교육적 효과성 휴리스틱스를 제시하였다.

2.5 컴퓨터기반평가시스템 휴리스틱스

본 연구의 목적과 관련된 교육 맥락에서 ‘시험’이라는 과업과 관련하여 Sim et al.[30]는 컴퓨터기반 평가 시스템(computer assisted assessment, CAA) 휴리스틱스를 개발하였다. [30]에서는 학생 설문조사, 휴리스틱스 평가, 선행연구 분석 등의 방법을 통해 휴리스틱스를 개발하였다. CAA 휴리스틱스는 11가지로 구성되었는데, 첫째, 시험 성과에 영향을 줄 수 있는 예외 방지 및 실수 회복 가능성 제시, 둘째, 시험이 실세계 경험과 유사하도록 사용자 자율성 제공, 셋째, 적절한 도움말과 피드백 제공, 넷째, 어플리케이션에서 자유자재로 항해하고 시험을 직관적으로 끝낼 수 있도록 제공, 다섯째, 표준과 일치하고 사용자 과업을 지원할 수 있는 인터페이스 디자인, 여섯째, 질문 스타일과 답하는 프로세스가 명확하도록 인터페이스 요소와 일치한 디자인, 일곱째, 입력한 답안이 문제가 되지 않도록 방지, 여덟째, 학생들이 시험을 접근하는데 어려움이 없도록 제공, 아홉째, 문제의 텍스트가 문법적 오류가 없고 체점의 명확한 알고리즘과 문제에 대한 피드백 제공, 열 번째 형평성 없는 신뢰를 가진 평가 제공, 열한번째 플랫폼이 안정성이 있고 학습자에게 영향을 주는 외부 요소가 최소화되어야 함을 강조하였다.

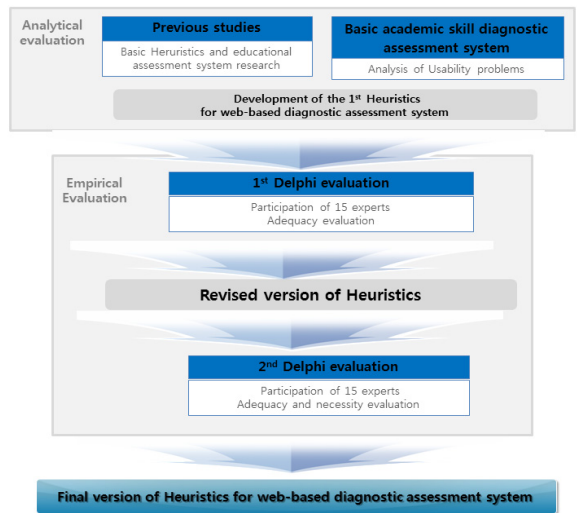
최근 들어, Sim & Read[29]는 위의 휴리스틱스를 활용하여 초보 사용성 평가자들 32명을 대상으로 LMS 시스템인 블랙보드에서 제공하는 평가도구에 대한 사용성 평가를 실시하고 분석해 봄으로써, 컴퓨터기반평가시스템(CAA)의 사용성 문제를 찾아내고 효율적으로 평가를 수행하는데 도움을 준다는 것을 확인하였다.

3. 연구 방법

3.1 연구방법 및 절차

본 연구에서는 웹기반 진단평가관리시스템에 대한 사용성 평가를 위한 휴리스틱스 개발을 위해 우선 문헌 분석을 통해 기본적인 사용성 평가 영역과 항목 등을

분석하여 휴리스틱스 초안을 개발하였고, 유사 시스템의 사용성 문제를 분석하여 기존 선행연구를 통해 얻기 어렵고 진단평가관리시스템 종속적으로 요구하는 사용성 평가 항목 및 요인들을 보완적으로 도출하였다. 이를 통해 도출된 휴리스틱스 평가 영역, 항목 및 기준 등에 대하여 전문가 델파이 방법을 활용하여 2회 조사를 실시함으로써 타당성을 측정하고 수정·개선할 사항들을 도출하였다.



(Fig. 1) Research methods and process

3.2 전문가 프로파일

선행연구와 기존 유사시스템 분석을 통해 1차 개발된 ‘웹기반 진단평가관리시스템을 위한 휴리스틱스’에 대하여 타당성 검증 및 개선을 위해 총 15명의 전문가가 델파이 조사에 참여하였다. 전문가의 선정 기준은 Grant와 Davis(1992)[12], Ericsson과 Charness(1994)[10]의 전문가 선정 기준에 따라 관련한 분야의 석·박사 학위소지자 또는 교육현장 등 현장 전문가의 경우에는 10년 이상 경력을 가지고 해당 주제와 유사한 프로젝트의 경험이 있는 전문가로 정하였다. 최종적으로 참여한 전문가 프로파일은 다음 <Table 1>과 같다.

<Table 1> Expert profile for Delphi method

No	Occupation	Expertise and career	years
1	Teacher	primary school, diagnostic assessment usability evaluation experience	20
2	UX Manager	Master in interaction design, HCI expert	10
3	Teacher	PhD candidate in Educational Tech, HCI expert	19
4	Teacher	middle school, diagnostic assessment usability evaluation experience	32
5	Professor	PhD in Educational Tech, HCI expert	5
6	Professor	PhD in Educational Tech, HCI expert	6
7	Researcher	PhD candidate in Educational Tech, HCI expert	11
8	Teacher	primary school, PhD candidate in Educational Tech, HCI expert	10
9	Professor	PhD in Educational Tech, HCI expert	10
10	Professor	PhD in Educational Tech, Web system expert	9
11	Professor	PhD in Educational Tech, Web system expert	10
12	Professor	PhD in Educational Tech, HCI expert	11
13	Researcher	PhD in Educational Tech, Web system expert	10
14	Researcher	Master in Educational Tech., Web system expert	10
15	Professor	PhD in Educational Tech, Web system expert	15

전문가는 기초학력 진단·보정시스템 사용성 평가 경험을 가진 현직 교사, UX 전문가, HCI(Human Computer Interaction) 학문과 관련 학위를 가지고 있는 전문가, 웹기반 교육시스템 전문가 등으로 구성함으로써 교육적 측면, 사용자 인터페이스(UI)적인 측면, 교육맥락에서 평가 및 관리라는 과업 실행적인 측면 등 다양한 관점에서 휴리스틱스를 평가할 수 있도록 하였다. 또한, 본 연구에 참여한 전문가들의 경우 유사시스템인 기초학력 진단·보정시스템에 대한 사용성 개선 연구에 자문단으로 참여한 경력이 있는 전문가를 선정함으로써 추상적인 관점에서의 항목 평가보다는 실질적인 시스템 활용과 개선 과정에서 논의된 항목들을 중심으로 타당도와 신뢰도 평가가 이루어질 수 있는 기반을 마련하였다.

3.3 유사 시스템 분석: 기초학력 진단-보정시스템

본 연구에서는 웹기반 진단평가관리시스템에 대한 휴리스틱스를 개발하기 위해 유사 시스템을 분석하였다. 유사 시스템은 현재 국내에서 전국의 초·중학생을 대상으로 기초학력에 대한 진단을 통해 학교 현장에 좀 더 체계적이고 지속적인 기초학력 지원을 위한 프로그램을 제공하려는 목적을 가지고 개발된 기초학력 진단·보정시스템이다. 기초학력 진단·보정시스템은 특정 시기에 전국의 학생들에게 진단 평가를 수행하도록 하는 진단 평가 기능을 비롯하여, 보정학습 및 학습 이력 등을 제공하고, 교사들의 경우 시행된 진단평가를 관리하고 학생들의 현황 및 학교에 학습 부진아에 대한 관리 기능을 수행하고 있다[24].

기초학력 진단·보정시스템은 전국 단위의 학생 평가 데이터를 관리할 목적으로 사용자(교사 및 학생)들이 특정 기간 동안 일관적인 프로세스를 가지고 진단 평가 및 학생 관리를 수행하도록 정책이 수행되고 있다. 하지만, 시험 유형 및 관리 절차 등이 복잡하고 이러한 과업 프로세스를 명확하게 활용하도록 시스템이 설계되어 있지 않아 시스템 활용에 대한 문제점이 제기되면서 많은 교사가 사용성에 대한 불만을 제기해왔다[13]. 특히, 시스템의 주요한 사용자 그룹 중 학생들의 경우 진단평가와 학생 이력 관리 및 보정학습을 온라인으로 수행해야 함에도 불구하고 시스템 활용에 어려움을 겪으면서 오프라인으로 시험을 보고 이를 교사가 다시 온라인으로 입력하는 등 시스템의 활용도가 낮다. 교사의 경우에는 이렇게 이중적으로 학생 관리, 시험 답안 입력 등으로 인하여 시스템에 대한 부정적인 시각을 가져 왔다. 그러므로 본 시스템에 대한 사용성 개선에 대한 연구가 시행되어야 한다. 사용성 연구를 통해 드러난 문제점 중 시험에 종속적인 항목으로는 시험 문제 제시, 시험 절차와 방법, 시험 결과 처리, 시험 통계 처리, 시험 응답 관리 및 교수학습 지원 등에 관한 것으로 드러났다.

3.4 자료의 해석 및 신뢰성 확보

본 연구에서는 초안으로 개발된 웹기반 진단평가관리 시스템을 위한 휴리스틱스에서 도출된 영역과 항목에 대한 필요성을 5점 Likert 척도로 평가하고, 각 항목에

대하여 검토 및 개선 의견을 제시할 수 있는 오픈 문항으로 구성된 설문지를 배포하고 이를 분석하였다. 오픈 문항의 경우, HCI 연구를 바탕으로 교육공학 박사학위가 있는 본 연구의 연구자 2인 및 외부 검토자 1인의 검토를 통해 본 연구의 목적과 문맥에 적합한지 여부를 논의하고 최종적으로 합의된 항목을 중심으로 개선된 휴리스틱스를 도출하였다. 본 연구에서는 전문가들이 전문 영역은 다르지만 평가 항목에 대한 일관된 시각으로 가질 수 있도록 유사 시스템인 기초학력 진단·보정 시스템의 사용성 평가에 참여한 경험을 가진 전문가가 참여함으로써 평가 항목에 대한 필요성과 중요도를 일관된 시각에서 평가할 수 있도록 하였다.

4. 웹기반 진단평가관리시스템 휴리스틱스 초안

지금까지 선행연구를 통해 분석한 휴리스틱스를 바탕으로 본 연구의 주요 영역인 웹기반 진단평가관리시스템을 위한 휴리스틱스 초안을 개발하였다. <Table 2>과 같이 초안으로 개발된 휴리스틱스의 대영역은 Nielsen의 휴리스틱스를 기본으로 하여 국내 교육 분야에서 논의된 교육맥락에 특화된 사용성 기준을 추가하고, 본 연구에 다루고 있는 ‘평가’라는 과업에 초점을 둔 ‘평가방법 및 환경’이라는 영역을 추가하여 총 7개의 영역으로 구성하였다.

첫 번째 영역은 ‘학습용이성’으로 Nielsen(1993) [21], 서영석(2007)[27], 임철일 외(2009)[18], 차현진과 안미리(2009)[8]에서 논의된 것처럼 사용자들이 친숙한 기존 국내에서 활용하고 있는 시스템 환경 및 인터페이스와 ‘일관성(1.1 Consistency)’과 사용자가 생각한 과업 및 기능에 대한 직관성을 따라 갈 수 있는 ‘알기 쉬운 시스템 상태(1.2 Visibility)’의 두 개 항목으로 구성되었다. 둘째, 효율성의 관점으로 조작이 편리하도록 사용자 통제를 부여하고(2.1 User control), 인지적 부담 없이

원하는 과업을 수행할 수 있으며(2.2 Minimize cognitive burdens), 사용자가 원하는 대로 자유롭게 항해와 과업 수행이 가능하며(2.3 Usable to navigate and do tasks), 자주 활용하는 메뉴를 쉽게 찾을 수 있는 사용자 효율성(2.4 Efficiency of use)을 포함하고 있다. 효율성의 측면은 Nielsen(1993)[21]의 휴리스틱에서 강조되고 있는 요소이면서, 임철일 외(2009)[18], 차현진과 안미리(2009)[8]등의 교육용 사이트/플랫폼을 얼마나 편리하고 효율적으로 활용할 수 있도록 인터페이스가 디자인 되어 있는지에 대한 논의가 사용성 평가 측정에서 있어왔다는 점에서 본 연구를 위한 휴리스틱스로 포함되었다.

셋째, 효과성의 측면은 교육에서 가장 중요한 종속변인으로 활용되는 만큼 본 연구에서도 중요한 휴리스틱스 요인이다. 진단·평가에서의 효과성이라는 부분은 학생들이 시험을 치를 때, 오류 없이 진행되어야 한다는 점이 가장 중요하기 때문에 효과성 영역에서 가장 주요한 요인으로 오류 방지(3.1 Error prevention)를 제시하였다. 또한, Mtebe & Kissaka(2015) [19] 및 차현진과 안미리(2009)[8] 등에서처럼 적합한 교수/학습 자료 및 자원들(3.2 Teaching materials and tools) 제공을 통해 내용적인 측면에서의 효과성 증진을 고려할 수 있도록 그 원칙을 하위요인으로 포함하였다. 넷째, 안정성 및 접근성 측면은 본 연구의 샘플 시스템으로 전국적으로 시행되는 기초학력 진단·평가시스템의 경우에서도 시스템의 인프라 및 활용 환경(4.1 Environments)이 자주 문제로 언급된 바 있고, 정영식(2008)[14]의 연구에서도 안정성은 중요한 교육시스템의 요인으로 다루어진 점에서 반영하였다. 접근성은 시스템을 다양한 학습자가 활용되어야 한다는 점(4.2 Accessibility)에서 서영석(2007)[27], 정영식(2008)[14], 차현진과 안미리(2009)[8], Mtebe & Kissaka(2015)[19] 등 다양한 교육시스템과 관련하여 언급된 바 있어 본 연구를 위한 휴리스틱스에도 포함하였다.

<Table 2> The first version of Heuristics for Web-based Assessment

Area	Heuristics	Descriptions	Nielsen (1993)	Nielsen (2009)	Seo (2007)	Jeong (2008)	Lim et al. (2009)	Cha & Ahn (2009)	Sim et al. (2009)	Mtebe & Kissaka (2015)
1. Learnability	1.1. Consistency	UI features and components are consistent?	Consistency and standards		Consistency on menu, functions		Consistency	Familiarity with UI		
	1.2. Visibility	Systems are visible enough to easily use according to user's expectation?	Visibility of system status	Memorability	Visibility of learning process		match with user's expectation	Learning process	User freedom	
2. Efficiency	2.1. Usable to navigate & do tasks	It is easy to navigate where a user wants to go and do tasks?	Flexibility and efficient to use	Efficiency		Durability	User control	Display of UI		
	2.2. Minimize cognitive burdens	The contents, menu and texts can be easily recognized?	Recognition rather than recall				Minimize cognitive burdens		Navigation	
	2.3. User freedom	Users can do what they intend?		Learnability	Precise learning process			Sitemap		
	2.4. efficiency of use	Users can control the functions and menus in a efficient way?	efficiency of use					Information on availability of buttons		
3. Effectiveness	3.1. Error prevention	Menus and buttons do not cause errors?	Error prevention	Error	Error and Help	Stability	Error prevention	Error prevention	Error prevention & recovery	
	3.2. Teaching materials and tools	Materials, tools, and functions are helpful and effective in doing teaching or learning activities?						Proper learning contents and multimedia		Teaching materials (QMRS,2012;Abion,1999,Squires&Prece,1996)
4. Stability & Accessibility	4.1. Environemtns	The system is stable at school environments?				Function stability				
	4.2. Accessibility	Diverse students are accessible to the system?			Accessibility and compatibility	Accessibility		UI for user with disability		Accessibility (Alsumait&Al-Osaimi,2010)
5. Helps and Manuals	5.1. User Manuals and documentation	The manuals provided are helpful to search the information? The manuals provided explain clear methods and process of how to use the n system?	Help	Documentation		User introvertness	Help			
							Help		Help and feedbacks	
6. Testing Methods and Environments	6.1. Display of questions	Questions on the test are displayed in a same condition as an offline test environment? (ex, time, manner, frequency, tools)							User Control and freedom	
	6.2. Testing methods and process	The testing methods and process are fair compared with offline testing environments?							Credibility and fairness	
	6.3. Results	Scoring and producing results are easy and efficient?								
	6.4. Statistics	Statistic student's records are easy to mange and are helpful to user?								
	6.5. Display of answers and explanations	Supplementary explanations with incorrect answers are provided?								
7. Satisfaction	7.1. Aesthetic design	Basic UI(font, color, layout etc.) is easy to use?	Aesthetic and miminalist design		UI with the aesthetic representation		Aesthetic design	Design		Motivation to learn (Alsumait & All-Osaimi, 2010)
	7.2. User control	It is easy to find the contents and materials?					User controls	Obvious status	User Control	Learner control (Squires&Prece,1996)
	7.3. Overall satisfaction	Overall usability and UI are satisfactory?		Satisfaction						Maintenance of motivation (Alsumait & All-Osaimi, 2010)

다섯째, 도움말과 설명서 측면(5.1 User manuals and documentation)은 기초학력 진단·평가시스템의 복잡한 시험절차와 관리 과업이 자주 언급되었다는 측면에서도 중요한 사용성 평가 원칙이 될 수 있고, Nielsen(1993)[21]의 기본 원칙에서도 항상 언급되는 요인이다.

여섯째, 본 연구에서 중점적으로 다루고 있는 평가라는 부분에 종속된 영역으로 평가에 종속적인 사용성 측정 도구를 제한한 Sim et al.(2009)[30]의 연구와 기초학력 진단평가 시스템에서 많이 제기된 문제점[24][13]들을 기반으로 평가 시스템이 가지는 고유한 과업에 해당하는 요인들로 구성하였다. 이를 통해 여섯 번째 영역에서 하위원칙들은 시험문제 제시(6.1 Display of questions), 시험 절차 및 방법(6.2 Testing methods and process), 결과처리(6.3 Results), 통계 관리(6.4 Statistics), 정답 제시 및 오답 설명(6.5 Display of answers and explanations)으로 구성되었다. 마지막으로 만족도 측면은 심미적 요소(7.1 Aesthetic design), 사용자가 스스로 통제할 수 있는 부분에서의 만족도(7.2 User Control), 전반적 만족도(7.3 Overall satisfaction)로 구성하여 감성적인 측면에서의 고려사항들을 휴리스틱스 마지막 영역에 포함하였다.

5. 휴리스틱스 검증 결과

전문가 델파이 조사를 2회 수행함으로써 문헌연구와 사례분석을 통해 최종적으로 개발된 진단평가관리시스템 사용성 개선을 위한 휴리스틱스의 각 영역과 항목의 필요성 및 항목에 대한 타당성을 평가하였다.

5.1 1차 델파이 결과

1차 델파이에서는 문헌연구와 사례분석을 통해 최종적으로 개발된 진단평가관리시스템의 사용성 개선을 위한 휴리스틱스 각 영역과 항목에 대한 타당성을 중심으로 평가하였는데, 전문가들은 타당성과 함께 항목 내용에 대한 세부적인 의견을 제시하였다. 1차 델파이를 수행한 결과, 대부분의 항목에서 평균 4.5이상으로 높게 나타났다.

<Table 3> The 1st Delphi Results

Area	Heuristics	Descriptions	Adequacy	
			M	SD
1. Learnability	1.1. Consistency	UI features and components are consistent?	4.67	0.62
	1.2. Visibility	Systems are visible enough to easily use according to user's expectation?	4.53	0.64
2. Efficiency	2.1. User operation and functions?	It is easy to find the feature?	4.60	0.63
	2.2. Minimize cognitive burdens	The contents, menu and texts can be easily recognized?	4.40	0.63
	2.3. User freedom	Users can do what they intend?	4.07	0.96
	2.4. efficiency of use and menus in a efficient way?	Users can control the functions in a efficient way?	3.73	0.70
3. Effectiveness	3.1. Error prevention	Menus and buttons do not cause errors?	4.67	0.72
	3.2. Teaching materials and tools	Materials, tools, and functions are helpful and effective in doing teaching or learning activities?	4.40	0.63
4. Stability & Accessibility	4.1. Environemtns	The system is stable at school environments?	4.53	0.64
	4.2. Accessibility	Diverse students are accessible to the system?	4.53	0.64
5. Helps and Manuals	5.1. User Manuals and documentation	The manuals provided are helpful to search the information? The manuals provided explain clear methods and process of how to use the system?	4.40	0.74
	6.1. Display ofquestions	Questions on the test are displayed in a same condition as an offline test environment? (ex, time, manner, frequency, tools)	3.33	1.05
6. Testing Methods and Environments	6.2. Testing methods and process	The testing methods and process are fair compared with offline testing environments?	3.80	1.21
	6.3. Results	Scoring and producing results are easy and efficient? Statistic student's records are easy to mange and are helpful to user?	4.80	0.41
	6.4. Statistics	Supplementary explanations with incorrect answers are provided?	4.67	0.49
7. Satisfaction	7.1. Aesthetic design	Basic UI(font, color, layout etc.) is easy to use?	4.40	0.74
	7.2. User control	It is easy to find the contents and materials?	4.27	0.59
	7.3. Overall satisfaction	Overall usability and UI are satisfactory?	3.87	1.06

평균이 4.0이하인 항목은 총 4개 항목이었는데, 전문가들의 검토 의견을 바탕으로 3개의 항목은 삭제하였고 1개의 항목은 평가 내용을 수정·보완하였다. 삭제된 항목은 '2.4 자주 활용하는 메뉴', '6.2 시험절차 및 방법', '7.3 전반적 만족도'의 항목이다. '6.2 시험절차 및 방법'의 경우 여섯 번째 평가 방법 및 환경이라는 항목에 별도로 두기 보다는 전문가의 의견에 따라 '2.3 작업의 흐름'과 '7.2 사용자 통제성'에서 시험 절차와 방법 등에 대한 명확성과 사용자의 통제에 관한 내용을 포괄하고 있고, '6.1 시험문제제시'에서 '6.2의 시험 절차의 공정한 환경'에 대한 내용을 포함하도록 수정하고 6.2의 항목을 삭제하였다.

전문가들이 제시한 검토의견을 종합해 보면, 우선 구체적으로 제시하지 않은 용어 등에 대하여 구체적인 예시 또는 가이드라인이 필요함을 논의하였다. 예를 들어, 1.1에서 화면의 구성요소가 정확히 무엇인지를 평가자가 파악할 수 있도록 예제를 통해 설명할 필요가 있음을 논의하였고, 2.2에서는 '적절히'라는 표현이 애매하다고 지적하여 '직관적으로'라는 좀 더 분명한 문구로 수정하였다. 둘째, '1.2.사용자의 기대와의 일치성'과 '2.2. 작업의 흐름'의 경우 1.2는 UI의 직관성 측면을 평가하고 2.2의 경우 과업 프로세스 관점에서 절차를 따라갈 수 있는 정보 구조와 메뉴를 구분하여 논의하였음에도 불구하고, 설명이 구체적이지 않아 구분이 모호하다고 지적하여 두 개의 원칙을 분명히 구분할 수 있도록 문구를 수정하였다. '3.2. 교수학습 자료' 원칙에서도 진단·평가시스템의 메인 기능이 아닌 만큼 전문가들이 어떤 문맥에서 평가해야 할지 모호하다는 점을 지적하여 시스템에서 진단·평가와 관련하여 제공하고 있는 교수학습 자료, 도구, 기능 등이 본 시스템의 목적 수행에 부합하는지를 구체적으로 명기하여 본 시스템과의 연계성 있는 항목으로 개선하였다.

이외에도 가장 많이 논의가 진행된 문항은 본 연구의 목적에 부합하는 여섯 번째 '평가' 관련 영역에서 하위 항목들에 대한 내용이었는데, '6.1 시험문제 제시' 항목에서는 본 연구의 유사 시스템인 기초학력 진단평가시스템의 경우 같은 시험 문제를 온라인 또는 오프라인으로 선택해서 수행할 수 있는 정책을 가지고 있다는 점에서 1차 델파이 문항에서는 오프라인과 동일한 환경의 시험 문제 제시하는 것에 초점에 맞춰있었는데, 전문가들은 어떠한 환경에서도

온라인과 오프라인 시험 환경이 동일한지 판단하기 어렵다는 의견을 많이 제시하였다. 이러한 의견에 따라 문항을 공정성에 맞추어 진행하였고, UI/UX 측면에서의 공정성 및 고려요인에 초점을 맞춘 시험 문제와 절차를 논의할 수 있는 문항으로 수정하면서 '6.2 시험 절차와 방법'에 관한 원칙을 통합하여 '시스템에서 제공하는 평가가 온·오프라인 방법이 선택적인 경우에는 온라인에 특화된 시험 방법, 절차 등이 시험 결과에 영향을 미칠 수 있는 UI/UX 요인이 동일한 조건에서 고려되고 있는가?'라는 문항으로 구성하였다.

마지막으로 연구자의 입장에서 평가시스템에서의 오류는 평가라는 주요 과업에서의 효과성을 저해한다는 측면에서 '오류방지'를 효과성 영역에 포함하여 구성하였는데, 많은 전문가는 '오류방지' 항목이 포함되는 것에 대한 이의를 제기함으로써, 5번의 대 영역을 '5. 오류방지'와 '도움말'로 수정하고 이 항목에 '오류방지'를 세부항목으로 옮겨서 수정하였다. 또한, '7.3. 전반적 만족도' 항목은 답변하기에 너무 주관적이고 정성적 항목으로 대체할 수 있다는 의견에 따라 휴리스틱스에서는 제외하는 것으로 결정하여 7.3의 항목은 제거하였다.

5.2 2차 델파이 결과

2차 델파이에서는 1차 델파이 결과를 바탕으로 수정·보완된 휴리스틱스의 항목과 내용까지를 최종 검토하기 위해, 각각 항목에 대한 필요도, 항목 내용의 타당성을 각각 별도로 평가할 수 있도록 설문지를 구성하였다. 2차 델파이 결과를 살펴보면, 필요도에 대한 평가에서 모두 4.4점 이상의 높은 평가를 받았고, 1개 항목(4.2. 접근성, M=4.13)을 제외하면 내용의 타당성에서도 모두 4.4점 이상의 높은 타당도를 보여주었다.

전문가들의 의견을 종합해 보면, 우선 HCI 관점에서 '학습용이성(Learnability)'이라는 용어가 교육 시스템에서 평가될 때 '학습(Learning)'과 혼용되어 교육전문가들은 '학습용이성'을 학습(Learning)의 관점에서 해석하려는 경향성을 보여주면서 학습용이성이 아니라 시스템 용이성으로 바뀌어야 한다는 의견을 제시하였다.

<Table 4> The 2nd Delphi Results

Area	Heuristics	Descriptions	Necessity		Adequacy	
			M	SD	M	SD
1. Learnability	1.1. Consistency	UI features and components are consistent?	4.87	0.35	4.87	0.35
	1.2. Corresponding to users	The menus and functions are corresponding to user's expectation even if the user is novice?	4.73	0.46	4.60	0.63
2. Efficiency	2.1. User operation	It is easy, correct, efficient to find the feature and functions?	5.00	0.00	4.93	0.26
	2.2. Minimize cognitive burdens	The contents, menus and texts can be easily recognized and found without recall?	4.80	0.41	4.67	0.49
	2.3. User freedom	Users can do what they intend?	4.80	0.41	4.67	0.62
3. Effectiveness	3.2. Teaching materials and tools	Materials, tools, and functions are helpful and effective in doing teaching or learning activities?	4.80	0.41	4.80	0.56
4. Stability & Accessibility	4.1. Environments	The system is stable at school environments?	4.87	0.35	4.73	0.59
	4.2. Accessibility	Diverse students are accessible to the system?	4.60	0.63	4.13	0.99
5. Help and Error prevention	5.1. Error prevention	Menus and buttons do not cause errors? The manuals provided are helpful to search the information?	4.87	0.52	4.73	0.59
	5.2. User Manuals and documentation	The manuals provided explain clear methods and process of how to use the system?	4.53	0.74	4.53	0.74
	6.1. Testing conditions	Tests are the same condition as an offline test environment? (ex, time, manner, frequency, tools) Scoring and producing results are easy and efficient?	4.80	0.41	4.80	0.56
6. Testing Methods and Environments	6.2. Results	Statistic student's records are easy to manage and are helpful to user?	4.80	0.41	4.80	0.56
	6.3. Statistics	Statistic student's records are easy to manage and are helpful to user?	4.87	0.35	4.80	0.41
	6.4. Display of answers and explanations	Supplementary explanations with incorrect answers	4.93	0.26	4.80	0.41

Area	Heuristics	Descriptions	Necessity		Adequacy	
			M	SD	M	SD
7. Satisfaction	7.1. Aesthetic design	are provided? Basic UI(font, color, layout etc.) is easy to use?	4.67	0.49	4.67	0.49
	7.2. User control	It is easy to find the contents and materials?	4.73	0.59	4.67	0.72

또한, 어색한 문장을 다듬어서 제시한 경우 그 의견을 최대한 반영하여 내용에 대한 설명이 좀 더 명확하고 이해하기 쉽도록 다듬었고, 항목간 중복된 내용을 포함한 경우 본 항목에서 배타적으로 측정하고자 강조한 문장으로 다듬어서 최종안을 마련하였다. 특히, 영역의 포괄성에 비해 내용은 일부만을 제시하고 있거나 내용은 포괄적인데 영역이 축소되어 있는 경우에서, 영역의 제목이나 내용을 수정하도록 하는 의견이 많이 제시되어, 영역과 내용이 일치할 수 있도록 수정·보완하였다. 예를 들어, 효과성의 영역의 ‘교수·학습 자료’에 대한 항목에서도 평가가 주요한 과업이고 교수·학습 자료는 부수적인 자료를 제공하는 것이 본 시스템의 목적이어서, 대영역의 제목을 효과성을 효용성으로, 소영역의 제목을 ‘교수·학습 자료’에서 ‘교수·학습 자료의 적합성’으로 수정하는 것이 논의되었다. 가장 이슈가 되었던 항목은 필요도와 타당도의 점수가 낮았던 ‘4.2. 접근성’ 항목이었는데, 우선 장애학생과 다문화권 학생 등 다양한 학습자를 모두 고려한 시스템을 개발하는 것이 어렵고, 현재 학교 시스템은 이러한 학생들이 제외된 상태에서 운영되는 경우가 많다는 점도 제시되었으며, 보조시스템으로 해결되어야 한다는 점 등 아직 접근성에 대한 교육 전문가들의 경험 부족으로 본 항목에 대하여 여러 가지 논의와 질문을 제시하였다. 하지만 최종적으로 ‘접근성’을 학습자의 다양성에 따라 하위 항목을 나누어 여러 개의 항목을 만들 경우 본 평가의 초점이 ‘접근성’으로 맞춰질 수 있어 한 개의 항목으로 포괄하는 것으로 결정하였다.

또한, 삭제할 경우 이 부분이 간과될 수 있지만 보편적 학습설계(Universal Design for Learning)의 관점을 제공하는 평가 항목이 한 개는 꼭 필요하다는 연구진의 합의에 따라 최종 평가도구에는 이를 포함하는 것으로 결정하였다. 마지막으로 도움말의 경우 하나의 내용으로 통합하는 것이 제안되었고, 오류방지의 경우 안정성 영역으로 이동할 것을 제안하였다.

5.3 최종 웹기반진단평가관리시스템 사용성 개선을 위한 휴리스틱스

다음 <Table 5>는 1차와 2차 델파이를 통해 전문가 의견을 반영하여 최종적으로 수정·보완한 웹기반 진단·평가관리시스템 사용성 개선을 위한 휴리스틱스 최종안을 보여주고 있다.

<Table 5> Heuristics for web-based diagnosis and assessment management system

Area	Heuristics	Description	Revised?
1. Learnability	1.1. Consistency	UI features and components are consistent?	O
	1.2. Corresponding to users	The menus and functions are corresponding to user's expectation even if the user is novice?	
2. Efficiency	2.1. User operation	It is easy, correct, efficient to find the feature and functions?	O
	2.2. Intuitive process	The contents, menus and texts can be intuitively recognized?	O
	2.3. User freedom	Users can do what they intend according to user's preference and usage patterns?	O
3. Effectiveness	3.2. Teaching materials and tools	Materials, tools, and functions are helpful and effective in doing teaching or learning activities appropriate to this system?	O
4. Stability & Accessibility	4.1. usage Environments	The system is stable at the environments where user are utilizing the systems?	O
	4.2. Accessibility	Diverse students are considered in accessibility design on the system?	O
	4.3. Error prevention	Menus and buttons do not cause errors?	
5. Help	5.1. User Manuals and documentation	The manuals provided explain clear methods and process of how to use the system?	O
6. Testing Methods and Environments	6.1. Testing condition	The equal UI and UX are considered on either online or offline testing environments? (ex, time, manner, frequency, tools)	O
	6.2. Results	Scoring and producing results are easy and efficient?	
	6.3. Statistics	Statistic student's records are easy to mange and are helpful to user?	
	6.4. Display of answers & explanations	Supplementary explanations with incorrect answers are provided?	
7.	7.1.	Basic UI(font, color, layout	

Area	Heuristics	Description	Revised?
Satisfaction	Aesthetic design	etc.) is easy to use?	
	7.2. User control	It is easy to proceed, use, exit the contents and materials except for the middle of testing?	O

<Table 5>에서 보여진 것처럼 1차와 2차 델파이를 통해 대영역과 하위영역에 대한 분류를 수정하였고, 일부 항목은 추가·삭제되었고, 문구가 보완적으로 기술되었다. Revised 항목은 1차와 2차를 거쳐 수정·보완된 항목을 표시해주고 있다.

6. 결론 및 제언

4차 산업혁명 시대를 맞이하여 교육 분야에서도 교수·학습을 지원하고 맞춤형 교육 환경을 제공할 목적으로, 전국 초·중학생의 데이터 수집, 관리 및 활용을 위한 시스템이 많아지고 있다[1][15]. 본 연구에서는 특히 전국적으로 시행되는 평가관리 시스템의 측면에서 현재 이미 국내에서 개발되어 시행되고 있는 진단평가 등의 과업을 수행하는 시스템의 사용성 개선을 도와 줄 수 있는 평가도구 개발을 목적으로 하였다. 이를 위해 본 연구에서는 문헌 분석과 기존 국내에서 활용되고 있는 유사시스템의 사용성 문제를 분석하여 1차 휴리스틱스 초안을 개발하였고 전문가 델파이를 2회로 시행함으로써 최종안을 마련하였다. 최종적으로 휴리스틱스는 총 7개의 대 영역에서 16개의 하위 항목으로 구성되었다.

본 연구를 통해 '시험'이라는 교육 영역에 종속적인 휴리스틱스를 개발하면서 가졌던 어려움을 논의하면서 향후 휴리스틱스 개발에 필요한 고려사항과 시사점에 대해 논의하고자 한다. 첫째, '시험'에 종속적인 시스템의 사용성을 평가하기 위해 전문가 선정의 문제가 있었다. 본 시스템은 교육에 특화된 과업을 수행하는 시스템인 만큼 과업의 절차와 내용이 교육적인 전문 지식 없이 평가되기에는 어려움을 가진다. 즉, 사용성과 UI에 대한 전문성만으로 평가할 수 없는 교육적 사용성(pedagogical usability)이라는 측면이 가지는 중요성이 있다[29]. 특히, 실제 시스템의 활용 경험을 가진 현장

전문가(예를 들어 교사 등)가 가지는 사용성의 어려움을 본질적으로 이해할 필요가 있었다. 그러므로 본 연구에서는 교육 현장 전문가의 의견이 필요했고, Ericsson과 Charness(1994)[10]이 제시하고 있는 현장 전문가의 선정 기준에 따라 교육 현장에서 10년 이상의 경력을 가진 자를 선정하였다. 이러한 현장 전문가는 UI 또는 사용성에 전문지식 없이 평가도구에 대한 타당성과 필요도를 논의하다보니, 논의의 대부분이 특정 과업에만 초점을 두거나 이미 활용해 본 유사시스템을 대상으로 평가하는 등의 한계를 가지고 있었다. 본 연구에서는 연구자가 교육 전문가이면서 HCI 전문가라는 점에서 이러한 전문가의 의견을 분별하고 조율하는 과정을 통해 최종안을 도출하였다. 하지만, 향후 비슷한 연구에서 특정 영역에 종속적인 평가도구를 개발하면서 과업의 특성으로 인해 전문가의 분야와 다양성이 넓을 경우, 연구자는 다양한 시각에서의 전문성을 보유함으로써 다양한 시각의 충돌과 특정 항목에 종속적인 내용들을 조율하고 연구의 목적에 집중하여 최종적인 의견을 도출할 수 있는 과정이 필요함을 시사하고 있다.

둘째, 사용자의 다양성을 고려하는 문제로, 본 시스템의 경우 사용자가 크게 교사(관리자 역할 포함)와 학생으로 분류되는데, 문제는 사용자에 따라 과업이 매우 다르다는 것이다. 즉, 교사는 관리적인 측면의 과업이고 학생은 실제 '시험'이라는 과업을 수행한다는 점에서 두 사용자의 작업 흐름도와 사용성의 문제가 전혀 다르다. 즉, 시스템 로그인시 사용자에 따라 UI와 메뉴는 많이 달라진다. 그럼에도 불구하고 본 연구에서는 진단·평가 시스템이라는 시스템의 관점에서 사용성을 개선할 수 있는 평가도구를 개발하는 연구 문제를 해결하다 보니 평가 영역과 항목이 이 두 사용자와 전혀 다른 과업을 포괄할 수 있도록 하는 부분이 큰 난관으로 작용하였다. 이것은 전문가들의 평가 의견을 통해서 확인할 수 있는데 포괄적인 문장이나 표현을 특정 사용자에게 맞춤형 표현으로 수정·보완하여 좀 더 명확한 평가 항목을 가져갈 수 있도록 제안하는 경우가 많았다. 하지만, 본 시스템은 사용자에게 관계없이 진단·평가시스템의 관점에서 사용성을 개선할 수 있는 포괄적인 용어와 내용으로 일반화된 휴리스틱스를 개발하고자 하였다. 결국, 향후 이러한 듀얼 사용자를 대상으로 한 시스템의 경우 각각의 사용자마다 필요한 과업을 중점적으로 평가할 수 있

는 별도의 휴리스틱스를 개발하는 것이 필요한지를 향후 연구로 제안하고자 한다.

셋째, 휴리스틱스에 '접근성'에 대한 평가 항목을 추가하는 문제에 대한 논의이다. 이미 W3C를 중심으로 접근성을 별도의 평가 항목[33]으로 평가하기 위한 노력들이 많이 진행되고 있고, 교육콘텐츠를 위한 접근성에 대한 연구도 시행된바 있다[2]. 이처럼, 접근성은 보편적 설계의 관점에서 교육 분야에서 중요한 항목임에도 불구하고 본 연구처럼 '접근성'을 세부 하위 항목으로 넣을 것인지를 판단하는 것에 전문가들의 의견이 있었다. 특히, 본 연구의 목적처럼 전국적으로 활용되는 교육 시스템의 경우 접근성 문제를 간과하기 어렵다. 그럼에도 불구하고 접근성의 영역이 넓고 교육 시스템에서 다양성의 개념도 포괄적이기 때문에 본 평가도구에서는 보편적 학습설계(Universal Design for Learning)의 이론이 제공하고 있는 원칙을 포괄할 수 있는 항목으로 본 시스템의 목적에 따라 세부 항목에 포함하였다[9]. 하지만 전문가들의 의견을 조율하는 과정에서 이를 사용자마다 여러 개의 평가 항목을 보다 구체적으로 추가할 것인가에 대한 판단의 어려움을 보여주고 있다. 결국 향후 연구를 통해 교육 맥락에서 특별한 과업을 수행하는 시스템의 경우(본 연구에서는 '시험'), '접근성'이라는 사용성 측면을 별도의 항목으로 고려할 것인지, 본 시스템에 종속적인 평가 항목으로 포함할 것인지 등에 대한 연구를 통해 교육 시스템에서의 접근성 측면에 대한 논의를 구체적으로 수행할 필요가 있다.

본 연구에서는 교육맥락에서 '시험'이라는 과업에 종속된 시스템의 사용성 개선을 위한 휴리스틱스를 개발하는 과정에서 다양한 어려움과 논의 사항을 보여주었다. 이러한 어려움에도 불구하고 그동안 '시험'이라는 교육에서의 중요한 과업에 목적을 둔 시스템의 사용성 개선을 위한 명확한 가이드라인이 부재하였지만 본 연구를 통해 개발된 휴리스틱스는 향후 전국 단위로 개발되는 평가 시스템 설계시 설계 가이드라인으로 활용될 수 있으며, 시스템의 사용성 개선을 위해 전문가가 평가 도구로 활용될 수 있다는 시사점을 가진다. 또한, 앞서 논의된 것처럼 '시험'이라는 교육에 특화된 과업에 목적을 가진 시스템의 휴리스틱스 개발 과정시 발생할 수 있는 다양한 어려움에 대한 시사점을 제공함으로써 향후 교육 맥락에서 활용될 수 있는 사용성 평가 도구 개발시 고려해야 할 사항들을 논의하고 있다.

그럼에도 불구하고 본 연구에서는 ‘시험’이라는 과업이 객관식의 선다형 문항에 초점을 맞춘 평가 시스템을 고려함으로써 현재 21세기 역량 측면[34]에서 수행형 기반의 역량 측정 도구를 개발하고 있는 트렌드를 반영하지 못했다는 한계점을 가지고 있다. 그러므로 향후에는 이러한 수행형 기반의 역량 측정 시스템의 사용성을 평가할 수 있는 휴리스틱스 개발에 대한 향후 연구를 제안하고자 한다.

참고문헌

- [1] Adams Becker, S., Cummins, M., Davis, A., Freeman, A., Hall Giesinger, C., and Ananthanarayanan, V. (2017). NMC Horizon Report: 2017 Higher Education Edition. Austin, Texas: The New Media Consortium.
- [2] Ahn, M.L., Noh, S.J., Kim, S.M. (2009). Research paper of Web accessibility.
- [3] Albion, P. R. (1999). Heuristic evaluation of educational multimedia: From theory to practice. In 16th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education, ASCILITE, 9-15. Brisbane, Australia.
- [4] Alsumait, A. a., & Al-Osaimi, A. (2010). Usability heuristics evaluation for child e-learning applications. *Journal of Software*, 5(6).
- [5] Baker, K., Greenberg, S., Gutwin, C. (2002). Empirical Development of a Heuristic Evaluation Methodology for Shared Workspace Groupware, CSCW'02, November 16 - 20, 2002, New Orleans, Louisiana, USA.
- [6] Blecken, A., Marx, W., (2010). Usability Evaluation of a Learning Management System, Proceedings of the 43rd Hawaii International Conference on System Science.
- [7] Bonham, S.W., Titus, A., Beichner, R.J., Martin, L. Education Research using Web-based Assessment Systems, *Journal of Research on Computing in Education*, 2000.
- [8] Cha, H.J., Ahn, M.L. (2009). Cyber Home Learning system diagnosis-prescription system usability evaluation and improvements on learning experience, HCI Conference 2009, 876-883.
- [9] David, H. R., Meyer, A., Strangman, N., & Rappolt, G., Teaching every student in the digital age: Universal Design for Learning, The Association for supervision and curriculum design(ASCD), from <http://www.ascd.org/publications/books/101042.aspx>
- [10] Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American psychologist*, 49(8), 725-747.
- [11] Fluck, A., Pullen, D. and Harper, C. (2009). Case study of a computer based examination system, *Australasian Journal of Educational Technology*, Vol. 25. No. 4. pp. 509-523
- [12] Grant, J. S., & Davis, L. L. (1997). Selection and use of content experts for instrument development. *Research in Nursing & Health*, 20(3), 269-274.
- [13] Jeon, S.J., Kim, H.S., (2016). A study on improvement of web-based diagnosis-supplement system for basic academic skills, *Journal of The Korean Association of Information Education*, 20(5), 487-498.
- [14] Jeong, Y.S., (2008). Digital textbook usability principle development, *Journal of Computer Science Education*, 11(30), 13-20.
- [15] Korea Education & Research Information Service (2017). The educational use of IT-Convergent technology in 4th industrial revolution era. KERIS Research Report RR 2016-7.
- [16] Korea Education & Research Information Service (KERIS). (2016). Whitepaper 2016.
- [17] Kwon, K.S., Park, J.S., Goo, C.D.(2014). Development of Agendas by using educational big data: Focused on analysis of SNS, KERIS, Research report KR 2014-10.
- [18] Lim, C.L., Song, H.D., Lee, Y.K., Lee. Y.T. (2009) A study on the developing and applying digital textbook platform usability evaluation tools, *Educational Technology Research*, 25(4), 125~155.
- [19] Mtebe, J. S., Kissaka, M. M. (2015), Heuristics

- for Evaluating Usability of Learning Management Systems in Africa, In IST-Africa Conference 2015, 1~13, IEEE
- [20] Mtebe, J. S. (2015). Heuristics for Evaluating usability of learning management systems in Africa, IIMC International Information Management Corporation
- [21] Nielsen, J. (1993). Usability Engineering, Academic Press Limited, London.
- [22] Nielsen, J. (1995). 10 Usability Heuristics for User Interface Design, Nielsen Norman Group. <https://www.nngroup.com/articles/ten-usability-heuristics/>
- [23] Nokelainen, P. (2006). An empirical assessment of pedagogical usability criteria for digital learning material with elementary school students. *Educational Technology & Society*, 9(2). 178-197.
- [24] Park, C.H., Kim, G.T., Lee, B.H., Kang, T.H., Jeong, G.H., (2014). A study on analysis of online basic academic skills improvement system status and development of policy plans, KERIS, Research report KR 2014-6
- [25] P21(Patnership for 21st century learning), Framework for 21st century learning, 2017, from <http://www.p21.org/our-work/p21-framework>
- [26] Preece, J., Rogers, Y., Sharp, H., (2015). *Interaction Design: beyond human-computer interaction*, 4th ed. John Wiley & Sons Ltd. West Sussex:UK.
- [27] Seo, Y.S. (2007). Development of usability evaluation tools for web-based learning site. PhD thesis.
- [28] Silius, K. & Tervakari, A.-M. (2003). The Usefulness of Web-based Learning Environments. The Evaluation tool into the Portal of Finnish Virtual University. in International Conference on Network Universities and E-Learning. Proceedings of mENU. Valencia, Spain, V. Peñarrocha.
- [29] Sim, G., and Read, J. C., (2015). Using computer-assisted assessment heuristics for usability evaluations, *British journal of educational technology*, 47(4).
- [30] Sim, G., Read, J.C., Cockton, G. (2009). Evidence based Design of Heuristics for Computer Assisted Assessment, IFIP Conference on Human-Computer Interaction, pp. 204~216.
- [31] Song, H.D., & Park, H.J. (2009). Exploring factors that affect the usability of digital textbook based on affordance perspectives, *Korean society for educational technology*, 25(3), 135-155.
- [32] Squires, D. and J. Preece, (1999). Predicting quality in educational software: Evaluating for learning, usability and the synergy between them. *Interacting with Computers*, Vol. 11. pp. 467-483.
- [33] W3C. Web Content Accessibility Guidelines, 2017 (WCAG) 2.0, from <https://www.w3.org/TR/WCAG20/>
- [34] Zhenming, Y., Liang, Z., Guohua, Z. (2003). A Novel Web-based Online Examinations System for Computer Science Education, 33rd ASEE/IEEE Frontiers in Education Conference, November 5-8.

저자소개

차 현 진



2012 한양대학교 교육공학(교육학 박사)

2015~ 2017 순천향대학교 교수학 습혁신센터 교수

2017~현재 온교육디자인 연구소 소장

관심분야 : 학습자 경험, UDL, 사용자 중심디자인, 미래교육환경

e-mail : lois6934@hanmail.net



황 윤 자

2003 한양대학교 컴퓨터교육(교육학 석사)

2013 한양대학교 교육공학과(교육학 박사)

2014~현재 단국대학교 공학교육 혁신센터 연구교수

관심분야: HCI, UDL, 공학교육, 스마트교육

e-mail : yjhwang@dankook.ac.kr