



영어 문답 연습 컴퓨터 프로그램이 중학교 영어 학습자의 성취도 및 오류에 미치는 영향 연구*

Kyoung Rang Lee (Sejong Univ.) Soonil Kwon (Sejong Univ.)



This is an open-access article distributed under the terms of the Creative Commons License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: Jan. 21, 2021
Revised: Feb. 14, 2021
Accepted: Feb. 25, 2021

Kyoung Rang Lee (1st author)
Professor, Sejong Univ.
Tel: (02) 3408-3118
kranglee@sdjong.ac.kr

Soonil Kwon (co-author)
Professor, Sejong Univ.
Tel: (02) 3408-3847
skwon@sejong.edu

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2017S1A5A2A01027326).

ABSTRACT

Lee, Kyoung Rang and Soonil Kwon. 2021. Effects of a computer program to practice answering English questions on Korean middle school English learners' performance scores and errors. *Korean Journal of English Language and Linguistics* 21, 110-131.

This study was planned to explore the effects of a computer program to practice answering English questions on Korean middle school English learners' performance scores and errors. Including negative questions that Korean students consider most difficult (1,182 questions, more than 8,000 possible answers, and various feedback), five levels of yes-no questions with different verbs were set with 17 stages in total. With a computer expert's collaboration, the questioning program was carefully developed. The English performance scores of 108 middle school English learners were compared as well as the types and numbers of errors produced while practicing with the program. First, practicing answering English questions using the program seemed to be most effective for the intermediate students; the beginners were also improved while the advanced were not. No gender differences in scores were found. Also, the program was effective for diagnosing the students' errors. The beginners committed significantly more types of errors more often than the others. A few gender differences in terms of errors were observed. Using a pre-planned, well-designed questioning program with immediate feedback, this study hopes to provide a firm basis for future studies about helping English learners interact with a more advanced computer program. This will also help teachers diagnose their students' abilities to answer in English effectively and suggest a new way to research on adapting ourselves to new environments caused by the COVID-19.

KEYWORDS

English questions and answers, computer program, English performance scores, types of errors, immediate feedback

1. 서론

코로나로 인해 학교에 가지 못하고 컴퓨터나 모바일 기기를 활용하여 집에서 학습을 해야 하는 초유의 사태가 우리나라를 포함한 전 세계의 교육 환경에 큰 변화를 초래하였으며, 이에 대한 현명한 대처가 필요한 상황이다. 이에 영어교육 분야에서도 이러한 상황에 도움이 될 수 있도록 효과적인 영어 학습 도구를 더욱 발굴하고 개발하여 교사와 학생, 학부모들에게 소개할 필요가 있다.

컴퓨터와 인터넷 등 테크놀로지가 발달하면서 영어교육 분야에서 효과적으로 영어를 학습할 수 있도록 돕기 위한 여러 장치들이 개발되었으며, 4차 산업혁명 이후 테크놀로지를 활용한 교육의 효과를 알아보기 위한 다양한 노력이 시도되었다. 1912년부터 2012년까지 100년 동안 영어 학습에 어떤 테크놀로지가 활용되었는지 확인해보니, 오디오 미디어 198건, 활동 사진 미디어 311건, 컴퓨터 미디어 198건 등 일부 저널에 출간된 논문(*English Journal*)만으로도 엄청난 양을 찾아볼 수 있어 4차 산업혁명 전으로도 매우 많은 영어 교사들이 현장에서 새로운 기술을 활용하여 영어 학습을 촉진하기 위한 노력을 지속해오고 있다는 것을 알 수 있었다(McCorkle and Palmeri 2016). 그러나 바둑(알파고)이나 의료(왓슨) 등 인간 고유의 영역이었던 부분에서도 급격하게 발달한 새로운 테크놀로지를 활용하고 있는 것에 비하면 교육 분야에서에서의 기술 적용은 상대적으로 느린 편이라 볼 수 있겠다.

그럼에도 불구하고 교사들은 학생들을 위한 수업 도구로 컴퓨터가 유용하다고 생각하여, 1980년대 초반부터 교실에서 컴퓨터를 활용한 수업을 지속적으로 시도하고 있으며, 특히 쓰기 등 산출 수업에 효과적이라(Ulusoy 2006) 여겨 쓰기 활동에 많이 활용하고 있다(151건, 76%). 또한 컴퓨터를 활용했을 때 학생들 개개인에 맞추어(individualized), 적응형(adaptive) 학습 도구를 개발하고자 하는 노력을 지속적으로 하고 있으며(Hsu, Hwang and Chang 2013), 특히 어휘 학습 분야에서는 *Hot Potatoes*와 *WordChamp* 같은 프로그램, *Lexical Tutor*라는 웹사이트, 즉각적인 사진 기능을 활용한 경우 등 테크놀로지를 이용하여 학습 효과를 증대시키기 위한 노력들을 하였다(Abraham 2008, Li 2010, Yoshii 2006, Yun 2011).

우리나라나 일본처럼 원어민과의 접촉 빈도가 낮은 환경에서는 컴퓨터나 모바일 기술을 활용한 학습 도구를 개발하여 활용한 후 긍정적인 효과를 보고하는 경우가 더 많이 있다(이경량 2018, Lee and Choi 2010, Lee and Kwon 2014, Yip and Kwan 2006). 이처럼 기술을 활용하면 교실에서도 일대일 학습 환경을 제공할 수 있어 학생들의 학습 참여도 및 학습 성취도가 향상된다고 한다(Chou, Block and Jesness 2012, Clarke and Abbott 2013, Falloon and Khoo 2014). 컴퓨터나 모바일 기술을 활용하면, 일대일 학습 환경을 제공한다는 장점 외에도 피드백을 즉각적으로 제공할 수 있어(Yekta and Dafeian 2016), 틀리거나 불필요한 행동을 교정하거나(Zahorik 1987), 학습 행동에 문제가 없을 경우에도 긍정적 행동을 지속시킬 수 있어 효과가 있다는(Eggen and Kauchak 2004) 즉각적 피드백의 장점을 최대한 살릴 수 있다. 또한 피드백을 제공할 때 컴퓨터 프로그램을 활용하면, 영어와 같은 외국어를 학습할 때 발음이나 문법이 틀릴까봐 걱정(foreign language anxiety)하는 등 외국어 학습 과정이나 평가 결과에 부정적인 영향을 미치는 정도(Hewitt and Stephenson 2011, MacIntyre and Gardner 1991)를 완화시킬 수 있다.

그러나 컴퓨터를 활용하여 원어민과 상호작용을 할 수 있는 기술력은 되지만, 교육 분야에서 기술 적용이 상대적으로 느리게 되고 있는 이유는 테크놀로지의 부적절한 활용이 오히려 문제를 일으킬 수도 있기 때문이다. 예를 들어, 요즘 상용화되고 있는 가상현실(virtual reality) 프로그램을 선불리 도입할 경우, 차원이 다른 또 다른 수준의 걱정, 예를 들면, 정서적인 거부감(resistance)

등을 느낄 수도 있다(Childs 2010). 수업 중 대학생들에게 *Second Life*와 같은 가상현실 프로그램을 탐색해보고 활용하는 과제를 부여했을 때 새로운 기술에 대한 기대감보다는 기피하다며 거부감을 표현하거나, 현재 논란이 되고 있는 *이루다*와 같은 상업적 인공지능 대화 프로그램에서 볼 수 있듯이, 기술적으로는 이미 가상현실 속의 사람과 대화가 가능한 수준이더라도, 학습 상황에서는 아직은 더 익숙한 방식의 프로그램을 활용하여 개선 방안을 도출한 후 교육 현장에 적합하게 개발하는 것이 더 적절하다고 보았다(이장호, 김혜영, 신동광, 양혜진 2019). 특히 중학생처럼 어린 학생들이 정서적으로 편하게 영어를 학습하고 연습할 수 있는, 향후에 가상현실 속의 원어민과 영어로 말하거나 채팅 연습과 같은 실시간 상호 작용까지 나아가기 위해서는 간단한 문답 형태의 글쓰기 상호작용을 통해 정서적 완충 역할을 할 수 있도록 컴퓨터를 활용할 방법을 찾을 필요가 있다고 사료되었다.

이를 위해 컴퓨터와 간단한 질문과 적절한 대답을 할 수 있는 연습 프로그램을 만들 필요가 있다고 생각하였고, 우리나라 학습자들이 가장 어려워하는 형태의 영어 질문과 대답이 무엇일지를 포커스 그룹 인터뷰를 통하여 알아보았다. 포커스 그룹 인터뷰에 참가한 모든 학생들은 영어와 달리 우리말에서는 의문문을 만들 때 주어와 서술어의 도치 없이 억양을 올려 질문임을 나타내기 때문에 영어로 의문문을 만드는 것이 어렵고, 영어로 대답하는 것은 상대적으로 어렵지 않지만, 부정의문문에 대해서 영어처럼 긍정과 부정이 일치하지 않는 경우 적절하게 대답하는 것이 가장 어렵다고 하였다. 예를 들면, 우리말로 ‘책을 읽었어’는 평서문의 억양으로 말할 때(하향: I read the book.)와 의문문의 억양(상향: Did you read the book?)으로 말할 때 억양의 차이만 있을 뿐 어순 등은 차이가 없이 똑같지만, ‘책을 안읽었어(Didn't you read the book)?’라는 부정의문문에 대한 대답으로는 ‘응, 안 읽었어(Yes, I didn't)’ 또는 ‘아니, 읽었어(No, I did)’처럼 긍정과 부정이 일치하지 않아 영어로 대화 시 가장 많은 오류를 볼 수 있었다.

궁극적으로는 인공지능이나 음성인식 기술 등을 활용하여 원어민과 실제 대화를 하기 어려운 상황에서 영어 학습자가 실제와 유사한 대화를 하는 것이 목표이겠지만, 우선은 텍스트로 질문에 대한 대답을 적절히 하는 문답 연습 프로그램으로 편하게 접근을 시도하여 학생들의 오류 현황 및 연습이 더 필요한 문답의 종류를 파악하는 것이 선행되어야 하겠다.

이 프로그램을 활용하여 영어 문답을 연습한 후에 일반적인 학교 영어 시험에서의 수행 정도가 향상되었는지, 영어 수준(상중하)에 따른 수행 정도는 차이가 있는지를 알아볼 필요가 있겠다. 또한 영어와 같은 외국어 학습에서는 여학생이 남학생보다 우수하고(Oxford, Young, Ito and Sumrall 1993, Zoghi, Kazemi and Kalani 2013), 영어 철자나 맞춤법 등에서의 오류도 적다는 결과들이 많으므로(Chen 1996, Ellis 2012), 영어 수준과 함께 성별에 따른 차이도 알아볼 필요가 있다고 사료되었다.

따라서 본 연구는 영어 원어민과 대화할 기회가 상대적으로 많지 않은 우리나라 중학교 영어 학습자들을 대상으로, 향후 이용 가능한 실시간 상호작용 프로그램에 적응하기 위한 중간 단계로서, 특히 우리나라 사람들이 어려워하는 부정의문문을 포함한 영어 대답 연습을 돕기 위한 프로그램을 개발하여, 영어 수준과 성별에 따라 영어 성취도 점수와 오류 현황에 어떤 차이가 있는지 알아보기 위하여 계획되었다.

2. 이론적 배경 및 선행 연구

2.1 영어 학습을 위한 테크놀로지 활용

구성주의(constructivism)에 기반한 모바일 학습 이론(theory of mobile learning: Sharples, Arnedillo-Sanchez, Milrad and Vavoula 2009)에 따르면 학습은 교실뿐만 아니라 교실 밖에서도 지속적으로 이루어져야 하며, 컴퓨터를 이용하여 이러한 학습 환경을 조성했을 때 더욱 효과적이라고 한다(Xu 2018). 특히 상호작용 가설(interaction hypothesis)에서 강조하듯이 언어 학습의 효과를 위해서는 상호작용이 필수적인데, 모바일 기술의 발전으로 학습자들이 실제적이고 구체적인 내용으로 즉각적인 상호작용을 할 수 있는 환경도 조성되어 있으므로(Zimmerman and Howard 2013), 언제, 어디서나 학습에 참여할 수 있어(Hess and Gunter 2013), 학습자뿐만 아니라 교사의 역할에도 긍정적인 변화를 초래하게 되었다.

따라서 영어 교육 환경에서 테크놀로지를 활용하려는 노력이 지속되어 오고 있다. 1980년대 초반부터 컴퓨터 기반 언어 학습(computer-assisted language learning; CALL), 멀티미디어 기반 언어 학습(multimedia-assisted language learning; MALL) 등의 시도가 지속되어 있으며, 실시간 상호작용이 가능한 웹(Web 2.0)과 스마트폰 등 모바일 기술이 발달하여 교육을 위한 컴퓨터의 활용 범위가 엄청나게 확장되고 있다(Dudeny and Hockly 2012).

요즘은 특히 전세계적으로 상호작용이 가능한 인터넷과 소셜네트워크서비스의 발달로 인간이 직접 소통하지 않더라도, 컴퓨터가 매개가 된 의사소통(computer mediated communication, CMC; Peterson 2009)이 쉬워졌으므로, 온라인 채팅 프로그램, 이메일 교환, 게시판 작성, 음성인식 프로그램(이경량 2018), 실시간 대화형 챗봇(이장호 외 2019) 등을 통하여 학습자들의 영어 능력 향상을 시도하고 있다. 오히려 원어민과 직접 대면한 상호작용보다, CMC를 활용하여 걱정과 같은 부정적 정서적 요인을 제한하는 것이 더 효과적이라고도 한다(Lin, 2014).

Lin(2014)은 2000년부터 2012년까지 출간된 CALL 분야 논문들을 검색하여 25편의 논문을 메타 분석한 후, 예상과는 다르게 자연스러운 데이터(naturalistic data)를 활용하여 의사소통을 시도했을 때보다, 학습자들의 반응을 유도하는 데이터(elicited data)가 더 효과적이었다고 하였다. 따라서 컴퓨터 프로그램이 질문을 하여 학습자가 대답을 할 수 있도록 유도하는 문답 연습 프로그램의 개발이 필요할 것이다. 그리고 Ziegler(2016)가 1990년부터 2012년까지의 CMC 관련 연구를 검색하여 비교한 결과 대학생(79%)과 고등학생(21%)을 대상으로 연구를 진행하였으며, 이러한 연구들이 대부분은 쓰기(79%) 분야였지만, 유창성, 정확성, 어휘 사용 정도, 문장구조의 복잡성 등을 측정하였다(Huang and Hung 2010, Volle 2005). 요즘에는 특히 코로나로 인해, 컴퓨터나 스마트폰을 학습에 활용하는 연령이 점점 낮아지고 있는데도 불구하고, 중학생을 대상으로, 영어 질문에 대한 응답을 연습하여 문답 능력을 진단하고, 이를 영어 학습 성취도와 관련하여 분석한 경우는 거의 없었다.

2.2 상호작용 및 피드백 관련 연구

최근 기술의 발달로 미국, 중국, 유럽 및 우리나라에서는 인간의 감각(시각, 청각, 촉각, 화학적 감각 등)을 모방한 기계 출력 정보와 인간의 신체 움직임 정보를 이용하여 인간과 컴퓨터(기계) 간의 자연스러운 상호작용을 목표로 하는 멀티모달 상호작용 기술을 개발하고 있다. 국내에서도 음성 인식을 통한 실시간 상호작용 프로그램(예: KT 누구, 애완로봇, 챗봇 등)을 개발하고 있으

며, 응용언어학 및 외국어 학습을 돕는 제품 개발이 용이해지고 있다.

전통적인 영어 교실에서 교사와 학생이 상호작용을 할 때 가장 기본이 되는 것이 질문과 대답인데, 교사의 질문에 적절히 대답하지 못하면 대화의 단절이 생기기 때문에(Kim 2013), 문답 연습이 선행될 필요가 있다. 외국어 학습을 위해 챗봇을 활용한 연구(이장호 외 2019)에서 다양한 질문 유형을 제시하여 학습자들이 챗봇과 활발한 상호작용을 할 수 있다는 제언을 한 것처럼, 음성 상호작용 전에 텍스트 형태로 다양한 질문과 이에 대한 학생들의 대답 능력을 알아볼 필요가 있다고 사료되었다.

이처럼 영어 학습을 위해 컴퓨터나 모바일 기술을 활용하여 상호작용을 돕는 이유는 교사 일인이 여러 명의 학습자들을 대상으로 할 때보다 개별 학습자들에게 맞는 피드백을 즉각적으로 제공할 수 있기 때문이다. 피드백은 특히 즉각적으로 제공했을 때 효과적인데(Zahorik 1987), 학습 행동이 틀렸을 경우뿐만 아니라 문제가 없었을 경우에도 피드백을 즉각적으로 제공하면 그렇지 않았을 때보다 긍정적인 효과를 도출할 수 있다고 한다.

특히 학습자들은 영어로 작문을 할 때 맞았는지 틀렸는지 즉각적으로 피드백을 주는 것을 선호했으며, 내용에 대한 피드백보다 문법적 오류에 대한 피드백을 더 유용하다고 생각하였고(Hedgcock and Lefkowitz 1996, Saito 1994), 오류에 대한 수정은 도움이 되지 않고(Leki 1991), 전체적인 구조에 대한 피드백은 필요하다고 생각하는(Cohen 1991) 등 피드백을 받아야 하는 분야에 대한 생각은 다양하지만, 대부분의 학습자들은 교사의 피드백이 도움이 된다고 생각하였다(Ferris 1995; Saito 1994). 일반적인 피드백보다는 본인에게 특화된(personalized) 피드백을 선호하였으므로(Cramp 2011), 연구자들은 컴퓨터 적응형 검사 방식을 활용하여 즉각적이고 개별화된 피드백을 제공하는 방법을 모색하고 있다(Hattie and Timperley 2007, Parr and Timperley 2010, Wang 2011).

다시 말해, 컴퓨터를 활용한 영어 교육 분야에서 상대적으로 배제되어왔던(Ziegler 2016) 중학생들을 대상으로, 다양한 질문 유형 및 피드백을 포함하여 개발한 컴퓨터 프로그램을 활용하는 것이 영어 성취도 및 오류에 미치는 효과를 알아볼 필요가 있겠다. 이를 위해 본 연구는 다음과 같은 연구 질문을 설정하였다.

- 1) 우리나라 중학교 영어 학습자들은 영어 문답 프로그램으로 연습했을 때 영어 학습 성취도가 향상되는가? 영어 수준(상중하)과 성별에 따른 차이가 있는가?
- 2) 우리나라 중학교 영어 학습자들은 영어 문답 프로그램을 이용할 때 대답에서의 오류 현황과 종류에서 영어 수준과 성별 등에 따른 차이가 나타나는가?

3. 연구 방법

3.1 연구 참여자

서울에 위치한 중학교 3학년 총 108명(남 54명, 여 54명)이 본 연구에 참여하였다. 해당 학교의 영어 교사들에게 학생들이 영어 질문에 대한 대답에 어려움을 겪는지 물어보았고, 이 중 20년 이상 경력의 영어 교사가 학생들의 영어 질문 대답 능력 향상을 위하여 참여에 동의하였다.

참여 교사가 가르치는 학교에 컴퓨터실이 있고 1학년 담당 영어 교사들도 동의하여, 같은 주에 모든 1학년 학생들(138명: 1반 25명, 2반 23명, 3반 22명, 4반 22명, 5반 23명, 6반 23명)을 대상으로 영어 문답 연습 프로그램을 수행할 수 있었다. 그러나 이 중 일부 학생들은 컴퓨터 문제로

인하여 끝까지 수행하지 못하거나 적어도 1-1의 단계(아래 도구 참조)도 도달하지 못한 학생들이 있어 이들의 결과를 제외한 총 108명의 결과를 분석하였다.

입학 후 3월말 학습자들의 영어 수준을 평가하기 위한 시험을 시행한 결과가 있어 동의하에 성적 비교를 한 후, 상, 중, 하 세 그룹으로 나누어 비교하였으며 각 수준 및 성별에 따른 참여자의 수는 표 1과 같다.

표 1. 참여자 현황

	합	상	중	하
남학생	54	27	20	7
여학생	54	33	15	6
합	108	60	35	13

3.2 연구 도구

(1) 영어 문답 연습 프로그램

영어 학습 도구 개발 경험이 있는 컴퓨터 전공 교수와 협업을 하여 문답 연습 프로그램을 개발하였다. 영어영문전공 학부생 두 명, 석사과정생과 약 5년 경력의 영어교육 분야 전문가와 함께 중학교 영어 교과서와 기존의 문법책들을 검토하여 다양한 질문 형태를 추출한 후 참여교사가 제안한 의문문 형태를 고려하여 크게 다섯 가지 종류로 분류하였다. 각 질문은 어디에서 추출한 것인지 참고문헌의 정보를 함께 표시하였다(예: 2-3-Y-030-0 Was the weather bad this morning?의 경우, 오른쪽 컬럼에 Cheryl Pavlik, Grammar Sense 1, Oxford, p. 215라는 정보를 기재함). 각 질문 형태에 맞는 가능한 대답을 모아 질문별로 정리한 후 컴퓨터 전문가와 약 7달간 지속적인 협의를 통해 프로그램으로 구현하기 적절한 형태로 질문과 답을 모아 텍스트 파일로 제공하였다.

표 2. 질문의 수준과 단계

수준-단계 (질문 수)	질문 종류	예
1-1 (110)	Do ... ?	Do you like this music?
1-2 (90)	Does ... ?	Does she like this music?
1-3 (152)	Did ...?	Did they like the music?
1-4 (30)	Don't ...? / Doesn't ...? / Didn't ...?	Don't they like the music?
2-1 (138)	Am ...? / Are ...?	Are you a student?
2-2 (144)	Is ...?	Is she a student?
2-3 (66)	Was ...? / Were ...?	Were they students?
2-4 (16)	Ain't ...? / Aren't ...? / Wasn't ...? / Weren't ...?	Aren't they students?
3-1 (90)	Have ...?	Have you heard the news?
3-2 (48)	Has ...?	Has she heard the news?
3-3 (44)	Had ...?	Had they heard the news?
3-4 (15)	Haven't ...? / Hasn't ...? / Hadn't ...?	Haven't they heard the news?
4-1 (38)	don't ...? / aren't ...? / ain't ...? / haven't ...? / doesn't ...? / isn't ...? / hasn't ...?	They are paper clips, aren't they? He likes this music, doesn't he?
4-2 (62)	didn't ...? / wasn't ...? / weren't ...? / hadn't ...?	They were students, weren't they?
4-3 (40)	do ...? / does ...? / did...? / am...? / is ...? /	She isn't listening to the
4-4 (64)	was ...? / were ...? / have ...? / has ...? / had ...?	teacher, is she?
5 (35)	Do you mind ...?	Do you mind if I ask you personal questions?

처음 추출한 질문의 종류는 훨씬 다양했지만, 컴퓨터 전문가가 프로그램(Heroku)으로 구현하려면 대답의 종류가 제한되어야 한다고 하여 '예/아니오'로 대답 가능한 질문으로 재설정하였다. 총 다섯 가지 종류(5수준: do동사 의문문, be동사 의문문, have동사 의문문, 부가의문문, 부정의문문)로 설정하여, 각 종류마다 다음과 같은 4단계를 거치도록 하였다: 1) 일인칭 주어(I)를 포함한 복수 주어일 경우(예: Do you like this music?, Are you a student?, Have you heard the news?, They are paper clips, aren't they?, and Do you mind if I ask you personal questions?), 2) 현재시제이면서 단수 주어일 경우, 3) 과거 시제일 경우, 4) 부정의문문일 경우.

본 연구가 부정의문문에 대한 어려움이 크다는 점에서부터 출발하였기 때문에 각 종류마다 4단계에서는 해당 동사 종류에 따른 부정의문문을 연습할 수 있도록 배치하였다. 마지막 부정의문문에 대해서는 형태로는 부정문이 아니지만 의미적으로 부정의문문인 'do you mind' 질문을 따로 설정하였으며, 이와 관련하여 'would you mind' 처럼 다른 조동사와 함께 쓰일 경우 대답이 무궁무진해지는 경우 프로그램으로 피드백을 주는 것이 매우 어려워져, 단계를 나누지 않고 'do you mind'로만 연습할 수 있도록 하였다. 요약하면, 총 5개의 수준 중 4개에 대해서는 4개의 단계를 거치므로, 16개의 단계가 있고, 다섯 번째 수준인 부정의문문에서는 프로그램 구현에 관한 제한에 따라 1개의 단계가 있어, 전체적으로는 5수준의 질문, 17개의 단계로 구성되어 있다(표 2).

이와 같이 구성된 질문의 수준과 단계들에 따라 교과서와 문법책들에서 추출된 질문들에 아래와 같은 체제로 번호를 붙여 컴퓨터가 인식하도록 하였다(표 3). 예를 들면, 1-1-Y-098-0이라는 번호의 경우, 첫 번째 수준, 첫 번째 단계의 질문이면서, 긍정의문문의 98번째 예이고, 이 때 마지막 0은 의문문이므로 'Do you want to go to the movie with us?'를 찾을 수 있도록 한다. 본 프로그램은 딥러닝을 할 수 있는 고도의 기술은 아닌 틀로 만들었기 때문에 예상 답변을 직접 추려 기입해두고, 이와 맞지 않는 답을 했을 경우 적절한 피드백을 할 수 있도록 데이터를 마련해 두었다. 그러나 답변의 경우 질문에 제시된 내용에 따라 예상 답변 체제에 따라 줄어들기도 늘어나기도 해서 최대 16가지의 답변이 필요한 경우도 있었다(예: 2-4-N-22-0인 'Aren't you supposed to be at the party?'에 대해 'Yes./Yes, I am./Yes, I am supposed to be at the party./Yes, I am supposed to be there./Yes, I'm supposed to be at the party./Yes, I'm supposed to be there./No./No, I ain't./No, I am not./No, I'm not./ No, I ain't supposed to be at the party./No, I ain't supposed to be there./No, I am not supposed to be at the party./No, I am not supposed to be there./No, I'm not supposed to be at the party./No, I'm not supposed to be there.'로 16개의 답을 미리 설정함).

표 3. 질문/답변 번호의 체제

1-	1-	Y-	001-	0~
의문문 종류 (do/be/have/ 부가/mind: 1-5)	하위(세부) 종류 (복수현재, 단수현재 과거, 부정: 1-4)	질문의 긍정(Y)/ 부정(N)	해당 질문에 부여된 고유 번호 (001-999)	질문(0)/답변 (1-16)을 나누는 번호

이러한 질문과 답변에 대한 체제에 맞추어 각 단계별 질문의 수는 표 2에 제시한 대로 1,182개이며, 이에 대한 답변은 8,000개 이상이었다. 학습자들의 대답이 맞거나 틀렸을 경우 문체와 단계에 맞도록 다양한 피드백(예: Excellent!/Wow!/It's okay./As a sentence!/Good, but don't repeat the verb./You don't need to repeat the verb though./Say don't, not do not./The shorter, the

better!/Sorry, I can't get it./Try again./Say something./I can't hear you./Don't be discouraged. Think again!/You did great!/I'm very proud of you. Great~/You might want to try again./Amazing!/I'm sure you made a simple mistake./You're almost there. 등)을 준비하였다.

학생들은 로그인을 한 후 시범 문제를 풀어보고 성공했을 경우 항상 2-1의 단계에서 시작하도록 설정하였고(그림 1), 이후 3개의 질문을 연속해서 맞게 대답했을 경우 2-2, 2-3, 2-4, 3-1의 단계로 넘어가고 5단계까지 연속해서 맞게 대답했을 경우 칭찬 피드백과 함께 종료하고, 2-1 단계 이외의 단계에서 연속해서 3개의 질문에 틀리게 대답하면 바로 직전의 단계로 돌아간다(예: 3-4단계에서 3문제 연속으로 틀리면 3-3으로 돌아간다). 그러나 2-1의 단계에서 3개의 질문에 연속해서 틀리게 대답했을 경우 1-1의 단계로 돌아가서 같은 방식으로 진행할 수 있지만, 1-1의 단계에서도 연속해서 3개의 질문에 틀리게 대답했을 경우 'Good job! Try again?'이라는 피드백으로 다시 시작하거나 종료할 수 있다.

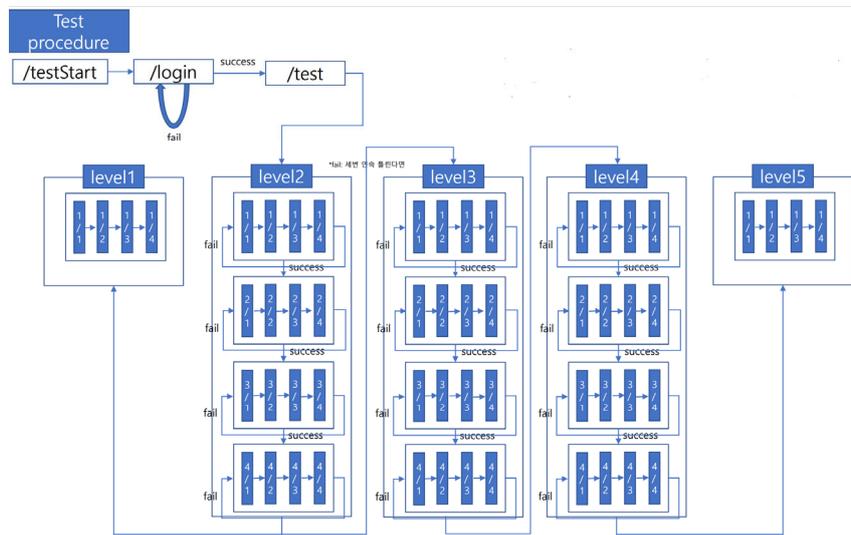


그림 1. 진행 순서

컴퓨터를 이용하여 홈페이지를 방문 후 로그인한 후 문답 연습을 해야 중간에 그만두고 다시 시작하더라도 학습자의 전체 학습 결과를 기록하여 분석할 수 있으므로, 회원가입과 로그인 절차를 함께 진행하도록 안내받았고, 사람이 질문하는 느낌을 주기 위해 참여 학부생이 직접 그린 캐릭터 그림을 질문과 함께 제시한 단순화된 화면을 활용하였다(그림 2).



그림 2. 로그인 화면과 캐릭터

(2) 영어 성취도 평가

연구 참여 중학생들의 영어 수준은 입학 후 3월말에 모든 신입생을 대상으로 보는 시험 결과를 활용하였다. 단순히 의문문만을 묻는 시험은 아니고 일반적인 영어 능력을 측정하기 위한 시험이었으며, 본 연구를 통해 궁극적으로는 영어 성취도 수준을 향상 시키는 것이 목적이므로 연구 참여 전후의 일반적 영어 능력 향상도를 비교하는 것이 타당하다고 사료되었다. 사전 성취도 평가는 70점 만점이었으며, 참여한 108명의 평균 점수인 51.53 ($SD = 9.68$)을 기준으로 하(0-41.00), 중(41.01-51.00), 상(51.01-70)으로 나누었다(표 1).

이후 본 연구가 진행된 직후 학교에서 시행된 중간고사 점수도 의문문만을 포함하는 것이 아니라 일반적인 영어 성취도 능력을 평가하는 형태였으므로, 중간고사 점수를 활용하여 향상 정도를 비교하였다. 사전 성취도 평가 점수가 70점 만점이었으므로, 중간고사 점수도 100점 만점을 70점 만점으로 환산하여 비교하였다.

3.3 자료 수집 절차

본 연구는 중간고사 시행 직전인 5월 첫째 주에 프로그램 개발 및 분석에 도움을 준 컴퓨터공학 전공 학생이 해당 학교를 방문하여 프로그램 진행 전반에 도움을 주었다. 컴퓨터 실습실에서 학생들에게 본 연구에 대해 설명을 한 후 회원가입과 로그인 절차를 진행하였다. 해당 영어 선생님이 학생들에게 일단 프로그램을 시작하게 되면 대문자로 시작하고 마침표로 끝나는 문장 형태로 답을 타이핑해야 하며, 의문문에 대한 대답이므로 Yes나 No로 반드시 시작을 해야 한다고 강조하였다. 예를 들면, 'Of course, I am.'과 같이 답할 경우 실제 대답으로는 가능하지만 프로그램에서는 정답으로 인식하지 못하는 것을 안내하고 시범문제에 성공적으로 답을 하면 2-1단계부터 시작할 수 있었다.

문답 연습 프로그램 다운로드부터 회원가입, 로그인, 질문 실행, 대답 타이핑 등 전반의 과정에 문제가 있을 경우 방문한 프로그램 개발 학생이 즉각 도움을 줄 수 있었다. 학생들의 로그인 정보는 연구 참여 후에도 활용을 원하는 학생들이 기억하기 좋고, 시험 성적과의 매칭도 효율적으로 할 수 있도록 학교에서의 학번을 이용하도록 하였다.

일단 로그인을 하면 2-1단계부터 주어진 의문문에 대한 답을 하고, 멈추어도 다시 로그인할 경우 이어서 진행할 수 있었다(그림 1). 각 단계마다 질문들이 학생마다 무작위로 제시되도록 하였다. 모든 학생들이 컴퓨터실에서 동일한 수업 시간 내에 진행을 하도록 하였으며, 컴퓨터 문제 등으로 모든 과정을 마치지 못했던 학생들과, 1-1단계의 과정도 마치지 못했던 학생들의 결과는 분석에 포함하지 않았다.

3.4 자료 분석 절차

본 프로그램은 모든 학생들의 틀린 대답까지도 저장하도록 설계되어 있었으므로, 프로그램 개발 학생이 모든 참여 학생의 데이터를 엑셀 파일에 추출하였다. 회원 가입 시 학번과 성별 등의 정보를 함께 기입하도록 하였으므로 이러한 배경 정보를 활용하여 사전 성취도 시험 및 중간고사 시험 결과를 매칭할 수 있었다.

영어 의문문 데이터를 수집하며 종류 설정 및 과정을 함께 한 영어영문학과 전공 학생, 대학원

생, 영어교육 전문가가 함께 참여 중학생들이 타이핑한 대답들을 모두 검토하며 공통적으로 보이는 오류를 코딩하기 위해 논의하였다. 대학생들의 영작문에서의 오류들을 상중하 수준별로 3개의 대분류, 15개의 중분류, 63개의 세부 항목으로 나누어 분석했던 기존의 연구를 참고하여, 참여 학생들의 오류를 검토하다가 공통으로 많이 보이는 오류들을 추려내어 32가지(예: ,를 .로 잘못 기재, 다른 조동사 사용, 오타, 동사 시제 틀림, 동사 수일치 틀림, 긍정 부정 반대로 기재 등)로 결정하였다. 각 오류들을 학생들의 답안별로 분석하여 코딩하였고, 지속적인 검토를 통해 틀린 문제수와 총 10개의 오류 범주¹로 결정하였다. 이후 영어 성취도 수준별 비교를 통해 그룹별 유의미한 차이가 나타난 6개의 오류 범주(1) 전체적으로 틀린 문제의 수는 프로그램 개발 학생이 제공한 엑셀 파일에서 자동 추출되었고, 2) 주어 오류, 3) 동사 오류, 4) 맞춤법 오류, 5) 긍정-부정 매칭 오류, 6) 기타 오류)를 확정하여 세부적인 분석을 하였다(결과 참조). 학원에서 영어를 가르친 경력이 약 2년인 영어교직 전공자와 영어교육 전문가가 각각 코딩을 하였고, 두 평가자간의 코딩 차이는 없었다.

엑셀 파일에 총 32가지의 오류로 각각 학생들의 대답을 코딩한 후 분석을 위해서는 10개의 오류 범주로 합산하여 통계 프로그램을 이용하여 그룹별 비교를 하였다. 그룹별 통계 비교를 하기 위해서 사회과학분야 통계 분석 프로그램인 SPSS(18 버전)를 활용하였다($p < .05$). 영어 수준에 따른 세 그룹 비교를 위해서는 ANOVA 및 ANCOVA(사후 분석 포함), 성별에 따른 비교를 위해서는 독립표본 t -test, 사전과 사후 비교를 위해서는 데이터 분할 후 대응표본 t -test를 실행하였다.

4. 결과 및 논의

4.1 중학생 참여자들의 영어 문답 프로그램 활용 결과

본 연구 참여 중학생들의 사전 영어 수준별 세 그룹으로 나눈 후 영어 문답 연습 프로그램으로 반복적인 연습을 진행하였고, 연습 직후 사후 영어 시험을 본 후 그룹간에 차이가 있는지 알아보 고자 하였다. 두 영어 시험이 비슷한 형태와 질문으로 구성되어 있기는 하지만, 학기초의 시험은 보다 더 전반적인 영어 능력 평가를 위한 시험이었고, 사후의 시험은 의문문이 들어간 교과서의 내용을 포함한 중간고사였으므로, 70점 만점으로 환산하여 비교하였다.

표 4에서 볼 수 있듯이 영어 문답 프로그램을 활용하기 이전의 영어 성취도 시험 결과는 세 그룹이 유의미하게 달랐으며($p = .000$), 사후 분석 결과(표 5), 상 그룹의 성적이 중과 하보다 유의미하게 높았으며, 중 그룹의 성적이 하보다 유의미하게 높았다(상 > 중 > 하; $p = .000$).

표 4. 사전 평가 결과

시험	그룹	N	M	SD	df	F	p
사전	하	13	33.38	8.27	2	207.35	.000
	중	35	46.49	2.68			
	상	60	58.40	4.03			
	합계	108	51.53	9.68			

¹ 틀린 문제 수, 구두점 오류, 대소문자 오류, do 대신 be, 주어 오류, 동사 오류, 맞춤법 오류, 내용 오류, 입력값과 맞지 않는 오답, 긍정-부정 매칭 오류, 기타 오류

표 5. 사전 평가 결과 사후 검정

시험	종속변수	(I) 점수등급	(J)점수등급	평균차(I-J)	표준오차	p
사전	Scheffe	하	중	-13.10*	1.43	.000
			상	-25.02*	1.34	.000
		중	하	13.10*	1.43	.000
			상	-11.91*	.93	.000
		상	하	25.02*	1.34	.000
			중	11.91*	.93	.000

이렇게 사전 시험의 결과가 달랐으므로, 문답 프로그램으로 연습을 한 직후 본 사후 시험 결과에 대해 ANCOVA를 시행하였다. 그 결과, 조정되기 전의 기술통계로는 상 그룹이 가장 높아보이지만, 조정 평균에서는 상 그룹이 오히려 가장 낮아진 걸 볼 수 있었다(표 6). 사전 시험 결과를 통제했을 때 문답 프로그램을 활용한 후의 수준별 영어 성취도 시험의 결과는 유의미하게 달랐으며($F = 85.75, df = 2, p = .000$), 흥미롭게도 중 그룹의 급격한 향상이 놀라웠다(표 7). 다시 말해, 문답 프로그램으로 연습을 하기 전, 상, 중, 하 세 그룹은 수준에 따라 유의미하게 성적 차이가 났었는데(상 > 중 > 하), 문답 프로그램으로 연습을 한 후 상 그룹을 제외하고 중과 하 그룹이 매우 향상되었으며, 특히 중 그룹이 상 그룹이나 하 그룹보다 유의미하게 높은 성적을 보였다(상 = 하 < 중). 그리고 각 그룹이 사전에 비해 사후에 향상되었는지 알아보기 위해 들린 대응표본 t-test 결과, 상 그룹을 제외하고($p = .121$) 중과 하 그룹은 모두 유의미하게 상승하였다($p = .000$). 본 프로그램으로 문답 연습한 것이 중과 하 그룹에게 효과가 있고 그 중에서도 중 그룹에게 가장 효과가 있었으며, 상 그룹에게는 효과가 없었던 것으로 볼 수 있겠다. 따라서 수준이 높은 학습자들에게는 의문사(why와 how 등)를 활용한 다양한 질문 유형을 제공하는 것이 바람직하다는 기존 연구(이장호 외 2019)의 제언을 반영하여 다른 시도를 할 필요가 있겠다.

표 6. 사후 평가 결과 및 조정 평균 (ANCOVA)

시험	그룹	N	M	SD(SE)	df	F	p
사후	하	13	40.15	7.12	2	82.55	.000
	중	35	57.74	4.66			
	상	60	58.67	4.25			
	합계	108	56.14	7.62			
사후 (조정 평균)	하		56.13	1.39	2	85.75	.000
	중		62.18	.58			
	상		52.62	.57			

표 7. 사후 평가 결과 대응별 비교 (ANCOVA)

시험	종속변수	(I) 점수등급	(J)점수등급	평균차(I-J)	표준오차	p
사후 (조정 평균)	Scheffe	하	중	-6.01*	1.24	.000
			상	3.51	1.81	.055
		중	하	6.01*	1.24	.000
			상	9.56*	.97	.000
		상	하	-3.51	1.81	.055
			중	-9.56*	.97	.000

성별로는 어떤 차이가 있는지 알아보기 위해 t-test를 들린 결과, 사전과 사후 모두 여학생의 점수가 남학생보다 높은 듯이 보이지만 통계적으로는 유의미한 차이가 없었다(표 8). 사전에 비해 사후에는 여학생은 여학생대로, 남학생은 남학생대로 모두 유의미하게 향상되었지만($p = .000$), 각 성취도 수준 내에서의 성별 차이(상 그룹의 남녀 차이, 중 그룹의 남녀 차이, 하 그룹

의 남녀 차이) 또한 없었다. 다시 말해, 본 연구에 참여한 남학생과 여학생은 영어 문답 프로그램으로 연습을 하기 전과 후의 영어 성취도는 서로와 큰 차이가 없었다.

표 8. 성별에 따른 영어 성취도 평가 결과

	성별	N	M	SD	t	df	p
사전	남학생	54	50.28	10.67	1.35	106	.181
	여학생	54	52.78	8.50			
하	남학생	7	30.71	10.11	1.29	11	.223
	여학생	6	36.50	4.46			
중	남학생	20	46.35	2.68	.34	33	.735
	여학생	15	46.67	2.77			
상	남학생	27	58.26	4.69	.24	58	.809
	여학생	33	58.52	3.48			
사후	남학생	54	55.72	8.28	.56	106	.572
	여학생	54	56.56	6.94			
하	남학생	7	39.43	8.58	.38	11	.709
	여학생	6	41.00	5.62			
중	남학생	20	57.25	4.87	.72	33	.478
	여학생	15	58.40	4.45			
상	남학생	27	58.81	4.73	.24	58	.810
	여학생	33	58.55	3.89			

4.2 중학생 참여자들의 대답에서의 오류 현황

영어 문답 프로그램 수행 중 나타난 오류들의 종류와 수 등 오류 현황은 성별이나 수준별 의문문에 대한 대답 수행에 어떤 차이가 있는지 알아보았다. 특히 틀린 문제의 수는 학생이 문답 연습 프로그램에서 전체적으로 같은 단계를 지났더라도 다른 특성을 드러내 줄 수 있어 분석에 시사점이 있을 것으로 기대되었다. 예를 들면, 두 학생이 똑같이 2-4단계까지 갔다고 해도, 한 학생이 2-3단계에서 두 개의 문제를 틀린 후 3개의 문제를 연속해서 맞춰 2-4단계로 갔고, 다른 학생은 2-3단계에서 하나도 틀리지 않고 연속 3개를 맞춰 2-4단계로 갔다고 하면, 결과는 같지만 후자의 학생은 0개, 전자의 학생은 2개의 틀린 문제가 있기 때문에 향후 학습에 대한 다른 계획을 수립할 필요가 있다.

표 9에서 볼 수 있듯이 틀린 문제 수와 총 10개의 범주에서 유의미한 그룹 차이를 나타낸 6개 (틀린 문제 수와 5개 범주: *로 표시)에 대해 사후 분석을 실시하였다. 오류 항목을 선별하기 위해 전체적으로 오류를 훑어볼 때, 가장 어려운 수준으로 설정한 'do you mind'로 묻는 부정의문문에 대한 대답을 할 때의 오류에서 다른 단계의 부정의문문 수준(1-4, 2-4, 3-4, 4-4; 표 2)에 대한 대답을 할 때의 오류와는 다른 오류가 있을 것이라 기대했었지만 결과는 그렇지 않았다. 따라서 긍정-부정 매칭 오류에 포함되어 함께 분석되었으며, 'Yes/No'로 대답을 하지 않아도 되는 다양한 부정의문문에 대한 대답을 연습해 볼 수 있는 교실 내 다른 활동이 필요할 것으로 보인다. 또는 딥러닝 기술을 활용한 프로그램을 개발하는 단계에서 부정의문문에 대해 다양하게 대답하는 데이터를 수집하여 자연스러운 대화를 할 수 있도록 도울 필요가 있겠다.

영어 교사들이 수업 시간 중 학생들에게 하는 질문의 종류를 비교한 기존 연구(Chaudron 1988, Long and Sato 1983, Long 1983)에 따르면, 본 연구에서 사용한 전시형 질문(display question)을 훨씬 더 많이 활용하고 있다고 하였다. 이러한 전시형 질문은 학습자와의 진정한 상호작용이라기보다는 질문에 대한 대답이 미리 정해져 있어 학습자의 학습 결과를 확인하는 의미

로 많이 활용된다. 궁극적으로는 인공지능 등을 활용하여 학습자에 대한 정해지지 않은 내용을 묻는 참조형 질문(referential question)을 연습하는 것이 더 적절하겠지만, 상기했듯이 본 연구에서 활용하는 프로그램에서는 우선 학습자들의 영어 문답 능력에 필요한 문법 능력을 향상시키기 위한 중간 단계의 목적이 있었으므로, 전시형 질문을 다양하게, 답으로 인정되는 것을 최대한 많이 설정하여 연습할 수 있도록 하였다. 그러나 향후 연구에서는 학생들과 진정성 있는 상호작용을 할 수 있도록 참조형 질문(Lee 2005)을 더욱 다양한 유형으로(이장호 외 2019) 개발할 필요가 있겠다.

위에서 유의미한 결과가 나와 사후 분석을 시도한 항목들에 대해 살펴보면, 우선 전체 문답 연습을 하면서 틀린 문제의 수는 하 그룹이 17.23($SD = 9.44$)개로 가장 많았으며 중 그룹($M = 4.83$, $SD = 3.06$)이 상 그룹($M = 4.03$, $SD = 2.60$)보다 조금 많았다. 이러한 세 그룹의 차이는 모두 통계적으로 유의미했으며(표 9), 하 그룹의 틀린 문제 수가 평균 17개 이상으로 대부분의 하 그룹의 학생이 상이나 중 그룹의 학생들보다 4배 이상의 문제를 틀린 것으로 영어 의문문의 대답 연습이 매우 많이 필요한 상태라는 것을 알 수 있었다. 다른 관점으로는 본 프로그램에서 난이도 별로 설정한 각 단계를 넘어가기 위해서는 연속으로 3개의 문제를 맞혀야 하고, 연속으로 3개의 문제를 틀렸을 경우 아래 단계로 돌아갔다가 다시 위 단계로 가야 하므로 많이 틀린 학생일수록 더 많은 문답 연습을 한 것이라, 단계를 다양하게 설정하여 하 그룹의 학생들이 상당히 많은 수의 질문에 틀렸어도 상대적으로 지루하지 않게 지속적인 연습을 할 수 있었다는 긍정적인 효과가 있었다고 볼 수 있겠다. 따라서 위의 성취도에서 사후 시험에서 상 그룹과 중 그룹의 차이가 없어진 결과는, 상 그룹 학생들보다 중 그룹의 학생들이 영어에 대한 연습량이 더 많아진 것이 사후 시험에서의 성취도에 영향을 미친 것으로 볼 수도 있겠다.

이러한 상황은 긍정과 부정의 매칭 오류에서도 똑같이 드러나는데(상 < 중 < 하), 전체 틀린 문제 수보다는 하 그룹의 오류 정도가 두드러지게 많지는 않더라도(하: $M = 2.46$, $SD = 1.20$, 중: $M = 1.08$, $SD = 1.20$, 상: $M = .40$, $SD = .99$), 중 그룹의 두 배 이상, 상 그룹의 8배 이상의 오류가 나타났다. 이는 영어로 부정의문문을 대답하는 것이 가장 어렵다고 했던 포커스그룹의 논의와 참여교사의 경험에서도 드러났듯이 우리나라와 대답하는 방식의 차이가 가장 컸기 때문이라고 보인다. 다시 말해, 하 그룹에서 특히 'Didn't you read the book?'과 같은 부정의문문에 대해 '응, 안 읽었어.'로 대답하듯이 영어로도 'Yes, I didn't.' 또는 'No, I did.'로 매칭 오류를 더 많이 하게 되었다고 생각되며, 영어 교사들은 긍정과 부정의 매칭에 대한 연습 기회를 학생들의 영어 성취도 수준을 고려하여 더 제공하는 것이 필요할 것으로 사료된다.

주어와 동사에 대한 오류에서도 실제로는 같은 양상을 보였는데(상 = 중 < 하), 상 그룹과 중 그룹이 차이가 거의 없고 이들에 비해 하 그룹은 약 4-5배 이상의 오류를 범했고, 기타 다른 오류들도 상 그룹에 비해 2배 이상이었던 것으로 보아, 하 그룹은 의문문에 대한 문답 연습과 함께 문법 연습이 선행되어야 한다고 보인다. 다시 말해, 상 그룹이나 중 그룹은 영어 문답 연습을 한 후 음성인식이나 챗봇 등을 활용해서 실제 원어민과의 음성 대화 상황으로 넘어가도 되겠지만, 하 그룹에서는 다른 요소를 더 연습할 수 있도록 지도하면서 본 프로그램 활용을 더 할 필요가 있다.

표 9. 영어 성취도 수준별 오류 현황

종류	그룹	N	M	SD	df	F	p
틀린 문제의 수*	하	13	17.23	9.44	2	56.35	.000
	중	35	4.83	3.06			
	상	60	4.03	2.60			
	합계	108	5.88	5.88			
구두점 오류	하	13	4.62	2.63	2	2.02	.116
	중	35	4.91	4.21			
	상	60	3.45	3.06			
	합계	108	4.06	3.47			
대소문자 오류	하	13	1.69	2.06	2	1.72	.183
	중	35	2.37	2.28			
	상	60	1.48	2.29			
	합계	108	1.80	2.27			
do 대신 be	하	13	.15	.55	2	.27	.763
	중	35	.34	.97			
	상	60	.28	.72			
	합계	108	.29	.79			
주어 오류*	하	13	1.46	1.76	2	12.43	.000
	중	35	.40	.74			
	상	60	.23	.46			
	합계	108	.44	.89			
동사 오류*	하	13	2.31	.75	2	17.31	.000
	중	35	.54	1.01			
	상	60	.47	1.11			
	합계	108	.71	1.19			
맞춤법 오류*	하	13	7.69	3.61	2	8.07	.001
	중	35	9.11	5.61			
	상	60	4.98	4.75			
	합계	108	6.65	5.26			
내용 오류	하	13	.08	.28	2	.033	.967
	중	35	.06	.24			
	상	60	.07	.25			
	합계	108	.06	.25			
입력값 다른 오류	하	13	.23	.60	2	.516	.599
	중	35	.09	.28			
	상	60	.32	1.38			
	합계	108	.23	1.06			
긍정-부정 매칭 오류	하	13	2.46	1.20	2	20.18	.000
	중	35	1.09	1.20			
	상	60	.40	.99			
	합계	108	.87	1.27			
기타 오류*	하	13	11.77	11.44	2	5.13	.008
	중	35	8.69	6.95			
	상	60	5.81	4.68			
	합계	108	7.47	6.85			

가장 흥미로운 것은 맞춤법 오류인데, 중 그룹($M = 9.11, SD = 5.61$)이 상 그룹($M = 4.98, SD = 4.75$)보다 두 배 정도의 오류를 범했고, 이는 하 그룹($M = 7.69, SD = 3.61$)보다 수치상으로는 더 많았지만 통계적으로 유의미한 차이는 아니었다. 대학생들의 영작문을 성취도 수준별로 비교했던 다른 연구(이경량 2014)에서는 중 그룹의 학생들이 더 많은 시도를 하는 과정에서 오류

를 범하게 된다고 했던 것과 비슷한 결과라고 볼 수 있겠다. 또한 중 그룹이 영어 성취도에서도 가장 큰 향상도를 보인 것도 많은 시도를 하면서 참여했기 때문으로도 볼 수 있으므로 틀려도 많이 시도할 수 있는 여건을 마련하는 것이 필요하겠다.

표 10. 오류 현황 사후 검정

종류	종속변수	(I) 점수등급	(J)점수등급	평균차(I-J)	표준오차	p
틀린 문제 수	Scheffe	하	중	12.40*	1.34	.000
			상	13.20*	1.26	.000
		중	하	-12.40*	1.34	.000
			상	.80	.88	.664
		상	하	-13.20*	1.26	.000
		중	-.80	.88	.664	
주어 오류	Scheffe	하	중	1.06*	.26	.000
			상	1.23*	.25	.000
		중	하	-1.06*	.26	.000
			상	.17	.17	.625
		상	하	-1.23*	.25	.000
		중	-.17	.17	.625	
동사 오류	Scheffe	하	중	1.76*	.34	.000
			상	1.84*	.32	.000
		중	하	-1.76*	.34	.000
			상	.08	.22	.943
		상	하	-1.84*	.32	.000
		중	-.08	.22	.943	
맞춤법 오류	Scheffe	하	중	-1.42	1.60	.676
			상	2.71	1.51	.205
		중	하	1.42	1.60	.676
			상	4.13*	1.05	.001
		상	하	-2.71	1.51	.205
		중	-4.13*	1.05	.001	
긍정-부정 매칭 오류	Scheffe	하	중	1.38*	.35	.001
			상	2.06*	.33	.000
		중	하	-1.38*	.35	.001
			상	.69*	.23	.015
		상	하	-2.06*	.33	.000
		중	-.69*	.23	.015	
기타 오류	Scheffe	하	중	3.08	2.17	.367
			상	5.96*	2.02	.016
		중	하	-3.08	2.17	.367
			상	2.88	1.46	.146
		상	하	-5.96*	2.02	.016
		중	-2.88	1.46	.146	

본 연구의 영어 문답 프로그램을 활용하면 하 그룹의 학생들이 영작문 과업에서는 실수를 범하지 않기 위해 또는 단순히 영어로 작문을 많이 하지 못했기 때문에 오류가 상대적으로 적었던 것(이경량 2014)과 달리, 영어 문장 산출의 양을 조절하지 않고 의문에 대해서는 비슷한 양을 대답해야 하는 상황(틀릴 경우 더 여러 개의 질문을 연습할 수 있지만 각 질문에 대해 한 문장으로 대답을 한다는 의미에서)에서는 하 그룹이 확실히 다른 그룹에 비해 많은 오류를 범하게 되어 영어 성취도 수준별 오류에 대한 더 정확한 진단이 가능하다고 볼 수 있겠다. 따라서 더 정확한 진단에 기초하여 교수/학습 계획을 세워야 하는 중학교 영어 교실 상황에서는 영어 성취도 수준별 학생들이 비슷한 양과 내용을 산출할 수 있도록 설계되어 있는 본 문답 연습 프로그램을 활용하여 학생들의 틀린 문제 수와 산출한 대답에서 나타나는 오류의 현황을 살펴보는 것이 필요할 것으로 사료된다.

이러한 오류 현황이 성별에 따른 차이를 보이는지 살펴본 결과(표 11), 기존의 연구(Ellis 2012, Pavlenko and Piller 2008)와는 달리 여학생($M = 6.02, SD = 6.98$)이 남학생($M = 5.74, SD = 4.60$)에 비해 더 적은 오류를 보이지는 않았다(틀린 문제 수: $t = .24, df = 106, p = .808$). 그러나 오류의 종류에서는 주어 오류와 긍정과 부정의 매칭 오류에서 여학생보다 남학생의 답변에서 유의미하게 더 많은 오류를 볼 수 있었다. 특히, 긍정과 부정의 매칭 오류는 영어 수준 내에서 성별 비교를 해 본 결과, 상 그룹에서만 오류가 거의 없었던 여학생($M = .03, SD = .17$)에 비해 남학생($M = .85, SD = 1.35$)의 오류가 유의미하게 높았던 유일한 항목이었고, 나머지에서는 성별 차이가 없었다. 이처럼 오류 현황에서 성별 차이가 미미했던 결과로 미루어보아, 영어 의문문에 대한 답변에 대한 연습을 할 때 남학생과 여학생을 나누어 생각할 필요가 없이 동일한 교수/학습 계획을 수립할 수 있겠다.

표 11. 성별에 따른 오류 현황

종류	성별	N	M	SD	t	df	p
틀린 문제 수	남학생	54	5.74	4.60	.24	106	.808
	여학생	54	6.02	6.98			
주어 오류	남학생	54	.61	1.12	2.09	106	.006
	여학생	54	.26	.52			
동사 오류	남학생	54	.57	.98	1.21	106	.362
	여학생	54	.85	1.37			
맞춤법 오류	남학생	54	7.11	5.76	.92	106	.228
	여학생	54	6.19	4.69			
긍정-부정 매칭 오류	남학생	54	1.20	1.46	2.82	106	.039
	여학생	54	.54	.95			
기타 오류	남학생	50	8.24	8.33	1.11	106	.268
	여학생	52	6.73	5.01			

표 12에서는 위에서 언급한 오류 현황에 대해서, 영어 수준별, 성별 내 영어 수준별, 성별, 영어 수준 내 성별 비교에 대한 통계적으로 유의미한 결과를 요약하였다. 성별 차이보다는 영어 수준별 오류 현황의 차이를 더 염두에 두어야 할 것이다. 오류 현황에서 주어와 동사 오류에서 상과 중 그룹의 차이가 없었는데, 일반적 영어 상황에서는 긍정-부정 매칭 오류나 맞춤법 오류 보다(일반 시험에서는 쓰기 항목보다는 이미 맞는 철자로 주어진 문제를 해석하고 활용하는 문제가 더 많으므로) 필요한 능력이 주어와 동사를 문법적으로 잘 활용할 수 있는 것으로 볼 수 있겠다. 또한 긍정부정 매칭 오류에서는 수준별 및 성별 차이가 두드러지게 나타나므로, 수준이 낮을수록, 남학생이 많을수록 더 많이 연습할 수 있는 기회를 제공할 필요가 있겠다.

표 12. 오류 현황 결과 요약

종류	성취도 수준별 요약	성별 내 성취도 수준별 요약		성별 요약	성취도 수준별 성별 요약		
		남학생	여학생		하	중	상
틀린 문제 수	상<중<하	상=중<하	상=중<하	-	-	-	-
주어 오류	상=중<하	상=중<하	-	남>여	-	-	-
동사 오류	상=중<하	상=중<하	상=중<하	-	-	-	-
맞춤법 오류	상<중	-	상<중	-	-	-	-
긍정-부정 매칭 오류	상<중<하	상=중<하	상<중<하	남>여	-	-	남>여
기타 오류	상<하	상<하	-	-	-	-	-

그러므로 일반적인 영어 능력에서 하 그룹과 남학생들에게 주어와 동사를 잘 활용하여 오류 없이 쓸 수 있는 연습을 가장 우선적으로 할 수 있도록 하고, 긍정과 부정 매칭을 포함한 나머지 문답 상황과 다른 과업 상황을 고려하여 영어 성취도 수준별 계획을 세워 더 효율적인 교수/학습을 도모할 필요가 있겠다.

5. 결론

본 연구는 인공지능 등 테크놀로지의 급격한 발달로 IT, 금융, 제조 등 다양한 분야에서 활발한 혁신을 하고 영어와 같은 외국어 학습 분야에서도 챗봇 등의 대화형 상호작용 프로그램을 개발하여 학습을 촉진시키기 위한 다양한 노력을 하고 있는 상황에서(이장호 외 2019), 우리나라 중학교 영어 학습자를 돕기 위한 영어 문답 연습 프로그램을 개발하여 그 효과를 알아보기 위해 계획되었다. 향후 실시간 채팅 및 음성으로도 의미 기반 대화가 가능한 상호작용 프로그램에 적응하기 위한 중간 단계로서, 실시간 채팅을 하는 환경으로 구성된 통제된 영어 문답 연습을 통해 대비할 수 있도록 하였다. 컴퓨터를 활용한 영어 교육 분야에서 상대적으로 배제되어왔던(Ziegler 2016) 중학생들을 대상으로, 특히 우리나라 사람들이 어려워하는 부정의문문을 포함한 영어 대답 연습을 돕기 위해, 다양한 질문 유형을 난이도에 따라 단계적으로 설정하고 개별 수준에 맞는 영어 대답과 피드백을 즉각 제공하는 과정을 경험할 수 있는 프로그램을 개발하여, 영어 수준과 성별에 따라 프로그램을 통한 연습이 영어 성취도와 오류에 미치는 효과를 알아보려고 하였다.

서울의 한 중학교 영어 교사의 자발적 참여로 해당 중학교 1학년 모든 학생(138명)이 컴퓨터 실습실에서 프로그램 개발자의 진행 도움으로 영어 문답 프로그램을 활용하여 영어로 문답 연습을 시행하였다. 컴퓨터 등의 문제로 5단계까지 프로그램을 종료하지 못했거나, 3문제를 연속으로 틀려서 1-1단계도 끝내지 못하고 종료한 학생들은 제외한 총 108명(남 54명, 여 54명)의 결과를 분석하였다.

본 연구에서는 영어영문학 전공 학부생들과 대학원생, 영어교육 전문가, 영어교사, 컴퓨터 전공 전문가, 프로그램 개발 학생들 등 많은 수의 연구 참여자들과 함께, 장기간에 걸쳐 데이터를 수집하고 분류하여 본 프로그램을 개발하였다. 의문문의 종류를 영어 교과서와 문법서 등을 참고하여 5개의 수준, 17개의 단계로 다양화하여 총 1,182개를 수집하고 이에 대한 답변은 8,000개 이상을 준비하고, 답변 종류에 따라 개별화된 것처럼 느낄 수 있는 피드백을 즉각 제동하여, 최대한 지루하지 않게 연습을 할 수 있도록 하였다. 가장 낮은 단계(1-1)가 아니라 2-1의 단계에서 시작하여 연속으로 3개의 맞는 대답을 했을 때 다음 단계로 넘어가 최종 5단계까지 갈 수 있고, 2-1의 단계에서 연속으로 3개 틀렸을 경우 1-1의 단계로 내려간 후, 여기서도 3문제를 연속으로 틀려 1-2의 단계로 넘어가지 못할 경우는 프로그램이 종료되도록 설계하였다.

우선, 우리나라 중학교 영어 학습자들은 영어 문답 프로그램으로 연습했을 때 영어 학습 성취도가 향상되는지를 알아보려고 하였다. 사전 시험 결과를 바탕으로 상중하 그룹을 구분하였으므로 세 집단의 영어 수준 차이는 유의미하게 나타났었으나, 영어 문답 연습 프로그램 활용 직후 시행된 중간고사에서는 중 그룹이 상과 하 그룹보다 유의미하게 높은 점수를 보였다. 상 그룹을 제외한 중과 하 그룹 학생들은 사전에 비해 유의미하게 향상되어 본 프로그램은 상 그룹보다는 중과 하 그룹에게 활용하는 더 적절해보였다. 본 프로그램에서 난이도 별로 설정한 각 단계를 넘어가기 위해서는 연속으로 3개의 문제를 맞추어야 하고, 연속으로 3개의 문제를 틀렸을 경우 아래 단계

로 돌아갔다가 다시 위 단계로 가야 하므로 많이 틀린 학생일수록 더 많은 문답 연습을 한 것이므로, 상 그룹 학생들보다 중 그룹의 학생들이 영어에 대한 연습량이 더 많아진 것이 사후 시험에서의 성취도에 영향을 미친 것으로 볼 수 있겠다. 수준이 높은 학습자들에게는 의문사(why와 how 등)를 활용한 다양한 질문 유형을 제공하는 것이 바람직하겠다(이장호 외 2019). 학습자들의 성별에 따라서는 각 수준별 및 수준 내 차이는 미미하였다.

이와 더불어 영어 문답 프로그램 연습을 통해 나타난 대답에서의 오류 현황과 종류가 영어 수준과 성별 등에 따라 차이가 나타나는지 알아보았다. 학습자들의 오류는 대학생들의 영어 작문 오류를 분석하기 위해 분석틀을 마련한 기존 연구(이경량 2014)를 활용하여 탐색적으로 코딩하였고, 틀린 문제 수와 10가지 종류의 오류 범주를 통계적으로 비교하였다. 이 중 성취도 수준별 유의미한 차이가 나타난 6가지 범주(틀린 문제 수와 5가지 종류의 오류 범주)에 대해 사후 분석을 하였다. 틀린 문제 수와 긍정-부정 매칭 오류에서 하 그룹이 중 그룹보다, 중 그룹이 상 그룹보다 유의미하게 많았으며(상 < 중 < 하) 주어와 동사 오류에서는 하 그룹의 오류가 많은 것은 똑같지만 상 그룹과 중 그룹의 오류 정도가 비슷했다(상 = 중 < 하). 이러한 경향은 정도의 차이가 약간 있을 뿐 성별 내 영어 수준별(남학생의 상중하 비교와 여학생의 상중하 비교)로 봤을 때 거의 비슷하게 나타났다. 그러나 전체 학생의 성별에 다른 오류 현황 차이는 거의 없었으며(주어 오류와 긍정-부정 매칭 오류에서만 남 > 여), 영어 수준 내 성별(하 그룹의 남녀 비교, 중 그룹의 남녀 비교, 상 그룹의 남녀 비교) 비교에서는 상 그룹의 긍정-부정 매칭 오류에서의 차이(남 > 여)를 제외하면 어떠한 차이도 없었다.

본 프로그램은 방대한 양의 질문과 대답 및 피드백을 장기간 계획하고 설계하여 만든 영어 문답 연습을 위한 것으로서, 상대적으로 짧은 시간에 학생들의 오류 현황에 대해 진단하고 향후 교수/학습 계획을 수립하기 위한 정보를 수집하기에 효과적이었다. 또한 단계마다 연속으로 3개의 문제를 맞힐 수 있었던 상 그룹의 학생들은 상대적으로 적은 연습량으로 인해 다른 그룹에 비해 효과가 적었다고 볼 수 있었던 반면, 하 그룹의 학생들은 문제가 틀리더라도 다시 시도하여 마지막 단계까지 지속할 수 있었으므로 서면 시험보다 훨씬 더 많은 연습을 할 수 있었던 긍정적인 효과가 있었다.

그러나 본 연구에서는 학교에서 일반적으로 시행되는 중간고사 시험을 통해 프로그램의 효과를 알아보려고 했었지만, 영어 문답 프로그램의 즉각적인 효과를 좀 더 명시적으로 알아보기 위해서는 과업 중심 평가를 시행하는 것이 더 좋았을 것이라 사료되었다. 향후 동일한 프로그램을 활용하기 전후에 학생들이 영어 의문문에 대해 대답을 하는 능력을 위주로 평가하는 시험을 개발하여 직접적인 향상도를 알아본 후에 일반적인 영어 성취도 평가 시험을 시행하여 그 관계를 알아 볼 수 있는 후속 연구가 필요할 것이다.

또한 기존의 연구들에서 전시형 질문보다 참조형 질문이 더 상호작용에 효과적이라고 했었지만(Chaudron 1988, Lee 2005), 본 프로그램은 미리 가능한 모든 대답을 정해두어야 했었기 때문에 'Yes/No'로 대답해야 하는 전시형 질문으로 한정했던 한계가 있었고, 이로 인해 'do you mind' 외의 다양한 대답이 가능한 부정의문문을 포함하지 못했던 한계도 있었으므로, 향후 인공지능 등을 활용하여 딥러닝이 가능한 프로그램 개발이 가능해지면 참조형 질문으로도 대답의 문법도 함께 확인할 수 있기를 기대한다. 또한 상 그룹의 학생들도 더 많은 양의 연습을 할 수 있도록 5단계를 마친 후 다시 연습을 희망하도록 더 흥미를 유발시킬 수 있는 방법을 모색해야 할 필요도 있겠다. 본 연구와 같은 프로그램 개발 연구들의 결과를 참고하고, 4차 산업의 발달과 코로나로 촉발된 교육에서의 테크놀로지 도입의 필요성을 적극 반영할 경우, 영어 교육 분야에서도 궁극적으로

는 실시간 상호작용을 하여, 교실 밖에서도 꾸준한 영어 학습이 가능할 것으로 사료된다.

참고 문헌

- 이경량(Lee, K. R.). 2014. 영어 학습자들의 영작문에 나타난 문법 오류 분석 연구(Profiling Korean English learners' writing errors). 《영어학》(*Korean Journal of English Language and Linguistics*) 14(2), 301-321.
- 이경량(Lee, K. R.). 2018. 음성인식 피드백 프로그램을 활용한 영어 학습 프로그램 개발 연구(Developing English learning program using speech recognition-based feedback). 《영어학》(*Korean Journal of English Language and Linguistics*) 18(4), 486-508.
- 이장호, 김혜영, 신동광, 양혜진(Lee, J., H. Kim, D. Shin and H. Yang). 2019. 외국어학습을 위한 대화형 챗봇의 담화 분석을 통한 개선 방안 연구(A study on the conversational interaction with chatbot for foreign language learning). *Multimedia-Assisted Language Learning* 22(1), 132-153.
- Abraham, L. 2008. Computer-mediated glosses in second language comprehension and vocabulary learning: A meta-analysis. *Computer Assisted Language Learning* 21, 199-226.
- Chaudron, C. 1988. *Second Language Classroom: Research on Teaching and Learning*. New York: Cambridge University Press.
- Chen, J. F. 1996. Gender differences in Taiwan business writing errors. *The Internet TESL Journal* 2(10), 1-8.
- Childs, M. 2010. *Learners' Experience of Presence in Virtual Worlds*. Doctoral dissertation, University of Warwick, UK.
- Chou, C. C., L. Block and R. Jesness. 2012. A case study of mobile learning pilot project in K-12 schools. *Journal of Educational Technology and Learning with Technology* 1(2), 13-25.
- Clarke, L. and L. Abbott. 2013. Hands on the iPad, heads in the cloud: Teaching and learning with iPads in a Belfast primary school In R. McBride and M. Searson, eds., *Proceedings of Society for Information Technology & Teacher Education International Conference 2013*, 3672-3680. Chesapeake, VA: AACE.
- Cohen, A. D. 1991. Feedback on writing. *Studies in Second Language Acquisition* 13(2), 133-159.
- Cramp, A. 2011. Developing first-year engagement with written feedback. *Active Learning in Higher Education* 12, 113-124.
- Dudeny, G. and N. Hockly. 2012. ICT in ELT: How did we get here and where are we going? *ELT Journal* 66(4), 533-542.
- Eggen, P. and D. Kauchak. 2004. *Educational Psychology: Windows on Classrooms*. Pearson Prentice Hall, Upper Saddle River.
- Ellis, R. 2012. *The Study of Second Language Acquisition* (2nd ed.). Oxford: Oxford

- University Press.
- Falloon, G. and E. Khoo. 2014. Exploring young students' talk in iPad-supported collaborative learning environments. *Computers & Education* 77, 13–28.
- Ferris, D. R. 1995. Student reactions to teacher response in multiple-draft composition classrooms. *TESOL Quarterly* 29(1), 33–53.
- Hattie, J. and H. Timperley. 2007. The power of feedback. *Review of Educational Research* 77, 81–112.
- Hedgcock, J. and N. Lefkowitz. 1996. Some input on input: Two analyses of student response to expert feedback in L2. *The Modern Language Journal* 80(3), 287–308.
- Hess, T. S. and G. A. Gunter. 2013. Serious game-based and non-game-based online courses: Learning experiences and outcomes. *British Journal of Educational Technology* 44(3), 372–385.
- Hewitt, E. and J. Stephenson. 2011. Foreign language anxiety and oral exam performance: A replication of Phillip's *MLJ* study. *The Modern Language Journal* 96(2), 170–189.
- Hsu, C.-K., G.-J. Hwang and C.-K. Chang. 2013. A personalized recommendation-based mobile learning approach to improving the reading performance of EFL students. *Computers & Education* 63(1), 327–336.
- Huang, H.-T. D. and S.-T. A. Hung. 2010. Effects of electronic portfolios on EFL oral performance. *Asian EFL Journal* 12(2), 192–212.
- Kim, J.-S. 2013. On the directions of improving EFL learners' interactional competence in the primary classroom discourse. *Korean Journal of Linguistics* 72(1), 159–181.
- Lee, J. and S. Choi, 2010. The effects of using multimedia title on preschool children's English listening skills, vocabulary, story recall ability and affective domains. *Multimedia-Assisted Language Learning* 13(3), 237–252.
- Lee, K. R. and S. Kwon. 2014. Effects of vocabulary memorizing tools on L2 learners' vocabulary size. *The Journal of Asia TEFL* 11(2), 125–148.
- Lee, S. 2005. Comparison of teacher talks from three different primary school ESOL teachers. *English Teaching* 60(2), 161–183.
- Leki, I. 1991. The preferences of ESL students for error correction in college-level writing classes. *Foreign Language Annals* 24, 203–218.
- Li, J. 2010. Learning vocabulary via computer-assisted scaffolding for text processing. *Computer Assisted Language Learning*, 23, 253–375.
- Lin, H. 2014. Computer-mediated communication (CMC) in L2 oral proficiency development: A meta-analysis. *ReCALL* 27(3), 261–287.
- Long, M. H. 1983. Native-speaker/non-native-speaker conversation and the negotiation of comprehensible input. *Applied Linguistics* 4(2), 126–141.
- Long, M. H. and C. J. Sato. 1983. Classroom foreigner talk discourse: Forms and functions of teachers' questions. In H. W. Seliger and M. H. Long, eds., *Classroom-oriented Research in Second Language Acquisition*, 268–285. Rowley, MA: Newbury House.
- MacIntyre, P. D. and R. C. Gardner. 1991. How does anxiety affect second language

- learning? A reply to Sparks and Ganschow. *The Modern Language Journal* 79(1), 90–99.
- McCorkle, B. and J. Palmeri. 2016. Lessons from history: Teaching with technology in 100 years of *English Journal*. *English Journal* 105(6), 18–24.
- Oxford, R., P. O. Young, S. Ito and M. Sumrall. 1993. Japanese by satellite: Effects of motivation, language learning styles and strategies, gender, course level, and previous language learning experience on Japanese language achievement. *Foreign Language Annals* 26, 359–271.
- Parr, M. J. and H. Timperley. 2010. Feedback to writing, assessment for teaching and learning and student progress. *Assessment Writing* 15, 68–85.
- Pavlenko, A. and P. Piller. 2008. Language education and gender. In S. Mary, ed., *The encyclopedia of language and education: Vol. 1* (2nd ed.), 57–69. New York: Springer.
- Peterson, M. 2009. Learner interaction in synchronous CMC: A sociocultural perspective. *Computer Assisted Language Learning* 22(4), 303–321.
- Saito, H. 1994. Teachers' practices and students' preferences for feedback on second language writing: A case study of adult ESL learners. *TESL Canada Journal* 11(2), 46–70.
- Sharples, M., I. Arnedillo–Sanchez, M. Milrad and G. Vavoula. 2009. Mobile learning: Small devices, big issues. In N. Balacheff, S. Ludvigsen, T. de Jong, A. Lazonder and S. Barnes, eds., *Technology–enhanced Learning: Principles and Products*, 233–249. Dordrecht, Netherlands: Springer.
- Wang, T. H. 2011. Implementation of web–based dynamic assessment in facilitating junior high school students to learn mathematics. *Computers & Education* 56, 1062–1071.
- Ulusoy, M. 2006. The role of computers in writing process. *The Turkish Online Journal of Educational Technology* 5(4), 58–66.
- Volle, L. M. 2005. Analyzing oral skills in voice e–mail and online interviews. *Language Learning & Technology* 9(3), 146–163.
- Xu, Y. 2018. Construction of a multiple English teaching mode based on cloud technology. *iJET* 13(8), 239–253.
- Yekta, R. R. and S. Dafeian. 2016. A study of the effects of time lag between learners' errors and teachers' feedback on the depth of vocabulary knowledge. *Theory and Practice in Language Studies* 6(11), 2109–2116.
- Yip, F. W. and A. C. Kwan. 2006. Online vocabulary games as a tool for teaching and learning English vocabulary. *Educational Media International* 43(3), 233–249.
- Yoshii, M. 2006. L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language Learning and Technology* 10, 85–101.
- Yun, J. 2011. The effects of hypertext glosses on L2 vocabulary acquisition: A meta–analysis. *Computer Assisted Language Learning* 24, 39–58.
- Zahorik, J. A. 1987. Reacting. In M. J. Dunkin, ed., *International Encyclopedia of Teaching and Teacher Education*, 416–423. Oxford: Pergamon Press.

- Ziegler, N. 2016. Synchronous computer-mediated communication and interaction. *Studies in Second Language Acquisition* 38, 553–586.
- Zimmerman, S. and B. Howard. 2013. Implementing iPads into K–12 classrooms: A case study. In R. McBride and M. Searson, eds., *Proceedings of Society for Information Technology & Teacher Education International Conference 2013*, 2512–2516. Chesapeake, VA: AACE.
- Zoghi, M., Kazemi, S. A. and Kalani, A. 2013. The effect of gender on language learning. *Journal of Novel Applied Sciences* 2(S4), 1124–1128.

예시 언어(Examples in): 영어(English)

적용 가능 언어(Applicable Languages): 영어(English)

적용 가능 수준(Applicable Level): 중학생(middle school students)