



## 원격교육 시대의 인공지능 활용 온라인 평가

이용상(영남대학교)\*

신동광(광주교육대학교)\*\*

### <국문초록>

본 연구에서는 원격교육 상황에서 평가 방안으로 인공지능 챗봇을 활용한 쓰기 평가 모델을 제안하고 영어과의 사례를 통해 그 가능성을 탐색하였다. 이를 위해 현재 상용 서비스를 제공하고 있는 대표적인 챗봇 가운데 하나인 미추쿠를 활용하였으며, 학생들은 미추쿠와의 대화를 통해 글을 쓰기 위한 자료를 얻고 이를 바탕으로 글 쓰기를 완성하였다. 또한 인공지능을 활용한 온라인 평가 방식의 가능성을 탐색하기 위해 오프라인 평가의 점수와 온라인 평가의 점수를 비교·분석하였고 실험에 참여한 학생들을 대상으로 한 설문문을 통해 챗봇 기반 온라인 평가의 유용성과 개선점에 대한 의견을 조사하였다. 미추쿠를 활용한 온라인 쓰기 평가에서 학생들은 기존의 오프라인 평가보다 통계적으로 유의한 수준에서 더 높은 점수를 받았으며, 이와 같은 결과는 이러한 온라인 쓰기 평가가 학생들의 변별보다는 학생들이 더 나은 글을 쓰는 본질적인 목표에 초점을 둘 수 있다는 점에서 의미가 있다. 설문 결과에서 학생들은 미추쿠와의 대화를 통한 영어 글쓰기가 영어 실력의 향상에 기여하고, 원격수업에서 흥미도와 참여도를 높일 수 있을 것으로 전망하였다. 오늘날과 같은 인포데믹 사회에서 영어 쓰기 역량은 정보처리능력까지도 포함하는 능력으로 재정의되고 있다. 이런 측면에서 본 연구에서 제안한 인공지능 챗봇을 활용한 영어 쓰기 평가는 이러한 영어 쓰기 역량을 보다 타당하게 평가할 수 있는 방안인 동시에 타 교과의 원격교육에서도 학생들의 능력을 공정하고 객관적으로 평가할 수 있는 대안으로 적용 가능하리라 사료된다.

★ 주제어: 원격교육, 정보처리 역량, 인포데믹 사회, 온라인 쓰기 평가, 챗봇

\* 제1저자: 영남대학교 교육학과 조교수

\*\* 교신저자: 광주교육대학교 영어교육과 부교수(sdhera@gmail.com)

## I. 서론

최근 전세계에 확산된 코로나19 전염병으로 인해 많은 사상자를 발생했으며, 우리나라도 대구 신천지 31번 확진자를 기점으로 감염자가 폭발적으로 늘기 시작하였다. 이에 정부는 코로나19의 대유행으로 인한 학생들의 안전을 고려하여 전국 모든 유·초·중·고의 개학을 1주일 연기한 것을 시작으로(교육부, 2020.2.23.) 총 4차례에 걸쳐 개학을 연기한 바 있다. 그러나 각급 학교의 학사 일정상 더이상 개학을 연기할 수 없다는 판단하에 4월 9일부터 중학교 3학년과 고등학교 3학년을 시작으로 단계적인 개학을 결정하였으며, 다만 학생들의 등교로 인한 코로나19의 집단 감염 사태 발생을 방지하고자 전 학년 원격수업을 실시하는 것으로 결정하였다(교육부, 2020.3.31.). 사상 초유의 전 학년 원격수업을 위해 정부와 시도교육청 그리고 각급 학교는 원격수업을 위한 인프라 구축과 제도 개선 등을 신속하게 추진하였다. 특히 정부는 원격수업을 통한 정규수업이 가능하도록 ‘원격수업 운영 기준’을 마련하였으며(교육부, 2020.3.27.) 이후 원격수업 출결·평가·기록 가이드라인을 안내하여(교육부, 2020.4.7.) 원격수업의 현장 안착을 꾀하였다. 그러나 정부에서 마련한 ‘원격수업 운영 기준’이나 ‘원격수업 출결·평가·학교생활기록부 기재 지침’은 기본적으로 원격수업이 한 학기 전체 시행되는 시나리오를 가정한 것은 아니기 때문에 원격수업 상황에서의 평가 방안을 구체적으로 제시하지는 못했다. 요컨대 ‘원격수업 운영 기준’이나 ‘원격수업 출결·평가·학교생활기록부 기재 지침’에 의하면 학생 평가는 과제물에 대한 평가나 관찰을 통해서 수행하되 학생들의 성취도는 기본적으로 원격수업 종료 후 등교수업에서 지필 평가를 통해 확인하도록되어 있다. 그러나 이와 같은 방식은 원격수업이 일시적으로 이루어질 경우에는 유효하지만 한 학기 또는 한 학년 전체 기간 동안 등교수업이 이루어지지 못할 경우에는 적용이 불가능한 방안이다. 따라서 원격수업 상황에서 적용 가능한 온라인 평가 방안을 시급히 구안할 필요가 있다. 특히, 코로나19의 지속적인 확산으로 인해 2학기 대면 수업이 불투명하고, 향후에도 코로나19와 같은 전염병 확산이 언제든지 다시 발생할 수 있는 가능성이 상존하고 있으므로, 원격교육 환경에서 학생들의 성취도까지 평가할 수 있는 방안을 마련하기 위한 연구가 필요하다.

원격수업은 지금까지도 등교수업의 보완적인 수업 형태로 널리 활용되었으며, 등교수업을 대체하는 정규수업의 형태가 아니었기 때문에 평가 결과가 학교생활기록부에 기재될 필요가 없었다. 따라서 지금까지의 원격수업에서 평가는 민감도가 낮은 자기진단의 평가 수준에 머물렀다. 그러나 정규수업을 대체하는 현재의 원격수업에서는 자기진단평가 방식으로는 평가의 공정성과 신뢰성을 확보하기 어렵기 때문에 정상적인 학사과정 운영을 위해서는 무엇보다도 평가방식에 대한 새로운 접근이 요구될 수밖에 없다. 이런 측면에서 최근 인공지능(Artificial Intelligent, AI)의 발전은 원격교

육에서의 평가방식에 새로운 가능성을 보여주고 있다. 흔히 챗봇(Chatbot)으로 알려져 있는 인공지능로봇은 인간처럼 대화를 나눌 수 있는 프로그램으로 이러한 챗봇을 교수학습에 활용하고자 하는 연구(김재상, 2017; 김혜영, 신동광, 양혜진, 이장호, 2019; 신동광, 2019; 추성엽, 민덕기, 2019; Kim, 2016)가 다각도로 진행되고 있지만, 챗봇을 활용한 평가 연구는 전무하다.

본 연구의 목적은 원격교육에서 평가 방안을 제안하는 것이다. 이를 위해 본 연구에서는 챗봇을 활용한 평가 문항을 설계하고 적용해 봄으로써 원격교육 시대의 새로운 평가의 대안으로서 인공지능 기반 평가 문항의 가능성을 탐색하였다. 또한 학생들을 대상으로 한 설문을 통해 인공지능로봇 기반 온라인 평가 방식의 장단점과 개선사항 등을 파악하고, 본 연구에서 제안한 평가 방식의 안착을 위한 시사점을 도출하고자 하였다.

## II. 이론적 배경

### 1. 학교 교육에서의 핵심역량

최근에 사회 각계에서는 핵심 키워드는 단연 4차 산업혁명이다. 4차 산업혁명의 시기가 도래함에 따라 교육 또한 큰 변화를 맞고 있다. 인공지능 즉 AI 기술의 급속한 발달은 우리의 생활에도 큰 변화를 가져오고 있다. 이미 많은 기업에서는 인간과 기계의 협업 또는 AI 기반의 기계들이 인간의 역할을 대체하고 있으며, 교육 분야에서도 정보 암기에 기반을 둔 교과는 큰 도전에 직면할 것으로 보인다. 대부분의 현대인들은 더 이상 전화번호를 암기하지 않는다. 불과 10~20년 전에는 전화번호의 암기력이 사회생활을 하는 데 있어 차별화된 능력으로 인정받았을 것이다. 이처럼 단순한 지식은 스마트폰과 같은 기기에 물어보면 언제 어디서든 답해주는 시대에 우리는 살고 있으며 그 변화는 기존 교육에도 큰 영향을 줄 수밖에 없다. 그렇다면 이러한 현대사회에 요구되는 능력에는 어떤 것이 있을까?

국가평생교육진흥원(2016)은 크게 기초문해, 역량, 인성 자질로 16가지의 현대사회에서 요구하는 핵심 능력을 제시하였다. 이는 지식에 기반한 선언적(declarative) 형태에 국한되는 것이 아니라 실제 사회에 진출하여 성공적인 삶을 살아가기 위해 필요로 하는 지식, 기능, 전문성 등을 포함하는 실행 가능한 과정적 지식(procedural knowledge)까지 포함한다는 점에서는 기존에 지식에 부여했던 가치와 차별화된다.

이미 세계 각국에서는 단순 지식기반의 사회에서 벗어나 그러한 지식을 어떻게 활용할 수 있는지의 실제적인 능력에 관심을 갖기 시작했고 이를 핵심역량이라 명명하고 교육과정에 반영하고자 지속적으로 노력하고 있다. 우리나라 또한 이미 2009 개정

교육과정에서 핵심역량을 반영해야 한다는 논의를 거쳐 2015 개정 교육과정에서 각 교과별로 4-6개의 핵심역량을 선정하였다. 또한 현대사회에서 요구되는 핵심적인 역량을 교육과정에 명시하고 이를 실제 교육 현장에서 어떻게 구현할 것인가에 대한 연구들이 활발히 수행되고 있다.

용어의 차이는 다소 있지만 2015 개정 교육과정에 포함된 교과역량은 기존의 연구에서 제시된 핵심역량을 교과별 특성에 맞게 재개념화한 것으로 영어과의 경우는 다음과 같이 4개의 교과역량으로 구성되어 있다.

첫째, 영어 의사소통 역량은 일상생활 및 다양한 상황에서 영어로 의사소통 할 수 있는 역량이며, 영어 이해 능력과 영어 표현 능력을 의미한다. 둘째, 자기관리 역량은 영어에 대한 흥미와 관심을 바탕으로 학습자가 자기 주도적으로 영어 학습을 지속할 수 있는 역량이며 영어에 대한 흥미, 영어 학습 동기, 영어 능력에 대한 자신감 유지, 학습전략, 자기 관리 및 평가를 포함한다. 셋째, 공동체 역량은 지역·국가·세계 공동체의 구성원으로서의 가치와 태도를 공유하여 공동체의 삶에 관심을 갖고 공동체가 당면하고 있는 문제를 해결하는 데 참여할 수 있는 능력이며 배려와 관용, 대인 관계 능력, 문화 정체성, 언어 및 문화적 다양성에 대한 이해 및 포용 능력을 의미한다. 끝으로 지식정보처리 역량은 지식정보화 사회에서 영어로 표현된 정보를 적절하게 활용하는 역량이며 정보 수집·분석 능력, 매체 활용능력, 정보 윤리를 포함하고 있다(교육부, 2015, pp. 3-4).

특히 2015 영어과 교육과정에서 주목할 부분은 언어교육에는 처음으로 정보처리 역량을 도입되었다는 점이다. 이미 이에 앞서 주현우와 신동광(2012)은 미래핵심역량으로서 영어능력에 정보검색 및 활용 능력을 포함해야 한다는 새로운 평가 구인을 제안하기도 하였다. 이는 기존 지필시험에서는 문항에서 제시된 정보 외에는 기타 정보에 대한 접근이 통제되었던 것과는 달리 실제 상황과 좀 더 유사한 환경 속에서 실제적 언어능력을 측정할 필요가 있다는 문제 제기였다. 최근 국내외 사회 환경을 비추어 보더라도 과거 정보화 사회에서는 많은 정보를 앞서 확보하는 능력이 중요했지만 이제는 넘쳐나는 정보의 인플레이션(inflation) 시대, 소위 인포데믹(infodemic) 환경에서 스스로 정보의 진위를 파악하고 내게 필요한 정보를 선별할 수 있는 능력이 곧 경쟁력이 된 만큼 이러한 정보처리 역량은 언어사용에서도 이미 중요한 자리를 차지하고 있다.

하지만 이러한 교과역량이라는 교육목표의 설정에도 불구하고 이를 어떻게 교수·학습과 연계하여 구현할 것인지 또는 어떠한 방식과 문항으로 교과역량을 평가할 것인가에 대한 구체적인 방법론은 부재한 상태이다. 이미 2009 개정 교육과정에서 미래 사회 핵심역량을 반영하고자 하는 시도가 있었지만 가시적인 효과를 볼 수 없었던 것이 이와 같은 이유에 있었다고 해도 과언이 아니다.

2000년대에 들어서면서 언어평가 관련 연구들은 외국어 평가의 핵심은 다른 사람

들과 의사소통하고 정보를 교환하는 것을 평가하는 것이라고 지적한다(Bachman, 2007; Bachman & Palmer, 2010). 전통적인 오프라인의 지필 평가는 급속한 기술 발전으로 인해 디지털 평가, 더 나아가 온라인 평가로 전환되고 있고 특히 영어과에서 수능 절대평가 도입 및 과정 중심 평가의 강화 등으로 인해 학생 변별을 위한 목적에서 벗어나 의사소통에 보다 초점을 둔 평가로 개선하고자 노력하고 있다. 그러나 이번 코로나19 바이러스의 확산으로 인해 원격수업이 확산되었음에도 불구하고 온라인 평가 방안은 여전히 부재한 상태이다. 이러한 사회적 기대와 교육계의 요구를 반영하여 본 연구에서는 포스트 코로나 시대의 온라인 환경에서 적용 가능한 온라인 평가 모델과 구체적인 적용 절차를 제시하고자 한다.

## 2. 온라인 평가 현황 및 제한점

지금까지 교육현장에서 온라인 평가는 오프라인 평가를 대체하기보다는 오프라인 평가를 보완하는 방안으로 여겨져 왔다. 특히 최근 학교 현장에서 과정 중심 평가가 실시되면서 학생들의 학습 과정에 대한 평가와 피드백을 위한 온라인 평가 시스템 구축 관련 연구들이 지속적으로 수행되었다. 우선, 박종임(2015)은 학생 평가 패러다임이 ‘학습 결과의 평가’에서 ‘학습을 위한 평가’와 ‘학습으로서의 평가’로 변화되고 있음을 주목하면서 이에 따른 온라인 형성 평가 시스템의 필요성을 강조하였다. 동 연구에서는 형성 평가가 평가 장소의 유연성, 채점 및 평가 결과 제공의 용이성, 피드백 제공의 즉시성을 가지기 위해서는 온라인 평가 체제가 구축되어야 한다고 강조하였다. 한편 이문복과 주현우(2017)는 고등학교 영어 말하기 수행평가를 온라인 시스템으로 구현하고 시범적용 하였으며, 이를 통해 그동안 지필 시험을 통해 간접 측정되었던 영어 말하기 능력에 대한 평가의 한계를 극복하였다. 이 밖에도 최근 온라인 평가 시스템 구축 관련 연구들(김성숙, 박찬호, 김미경, 김창환, 2018; 김인숙, 박지현, 임은영, 2018)이 수행된 바 있으나 이들 연구들은 평가의 편의성을 높이고 동시에 학습자에 대한 환류 방안에만 초점을 맞추고 있어 기존의 오프라인 평가를 대체하면서 평가의 정확성과 공정성을 담보하기 위한 방안을 제시하고 있지는 못하다. 한편, 권형규(2002)는 학습자 중심의 자기주도적 온라인 평가가 가능하도록 하기 위해 학습자 본인의 필요나 목적에 따라 스스로 평가 문항을 제작하고 운영할 수 있는 온라인 평가 시스템 구현의 필요성을 강조하였고, 이를 위해 상호작용을 극대화할 수 있도록 인터페이스를 설계한 웹 기반 평가 시스템을 제안하였다. 권형규(2002)의 연구는 온라인 평가가 기존의 오프라인 지필시험에서 구현하지 못했던 학습자와 문항 간의 상호작용을 이끌어 낼 수 있는 가능성을 보여주었다는 점과 상호작용 촉진을 위한 온라인 평가 설계 방안을 제시하였다는 점에서 의의가 있다.

이상의 온라인 평가 연구들을 종합하면 오프라인 평가에 대한 보완재로서 온라인

평가를 접근하는 경향이 많았고, 학교 현장에 과정 중심 평가가 도입됨에 따라 형성 평가 또는 수행평가를 위한 기재로 온라인 평가 시스템을 구축하고 활용하는 방안을 제시하는 데 주된 목적이 있었다. 이와 같은 연구들은 오프라인 평가의 제한점을 일부 극복하고 평가의 편의성을 극대화하는 데만 초점을 맞추고 있어, 코로나19 사태로 인한 원격교육 전면 실시에 따른 온라인 평가 상황을 충분히 반영하고 있지 못하다. 따라서 온라인 평가가 수행평가나 형성평가 차원을 넘어 총괄평가로서 기존의 오프라인 평가를 대체할 수 있는 방안에 대한 연구가 필요하다.

온라인 평가는 텍스트 중심 평가, 하이퍼텍스트 중심 평가, 멀티미디어 중심 평가, 상호작용성을 이용한 평가로 진화되었으며(황상연, 김두규, 임병민, 김정훈, 이재무, 1999), 이중 상호작용성을 이용한 평가가 가장 진일보한 평가라 할 수 있다. 최근 학습에 있어서 상호작용의 중요성을 인식하여 온라인 교수학습 과정에서 상호작용을 극대화하여 학습을 극대화하기 위해 인공지능 챗봇을 이용하고자 하는 연구들(김재상, 2017; 김혜영, 신동광, 양혜진, 이장호, 2019; 신동광, 2019; 추성엽, 민덕기, 2019)이 다수 있지만 아직까지 평가에서 챗봇에 기반한 연구는 거의 찾아 볼 수 없었다. 따라서 본 연구에서는 상호작용 중심의 평가를 통해 학습을 극대화하고 동시에 총괄평가로서 평가의 정확성과 공정성을 담보할 수 있도록 인공지능기반 온라인 평가 방안을 제시하였다.

### III. 평가 문항 개발 및 절차

#### 1. 평가 참여 대상

본 연구의 온라인 평가에 참여한 학생들은 광주권 A 대학의 2019년 2학기에 영작문 과목을 수강한 3학년 학생들이었다. 대학수학능력시험 영어영역에서 1등급에 해당하는 학생들이나 영작문 과목은 처음 수강하는 학생들이었다. 온라인 평가는 학생들의 동의하에 진행되었으며, 본 연구는 온라인 평가 모델의 제시에 초점을 둔 만큼 실험에 참여한 인원은 28명으로 제한적이다.

#### 2. 평가 도구: 인공지능 챗봇 미츠크(Mitsuku)

본 연구의 실험에서는 챗봇을 활용한 쓰기 평가를 구안하였다. 이를 위해 자유 주제에 대한 대화가 가능한 미츠크라는 챗봇을 채택하였다. 미츠크는 2005년 Steve Worswick이 개발한 챗봇으로 전신인 A.L.I.C.E.(Artificial Linguistic Internet Computer Entity)의 개발에 사용된 데이터 입력 시스템인 AIML(Artificial

Intelligence Modelling Language)을 기반으로 개발되었다(Wikipedia, 2020). 또한 현존하는 가장 우수한 챗봇으로 챗봇 개발에 있어 가장 권위있는 로브너 상(Loebner Prize)을 2013, 2016, 2017, 2018, 2019년까지 5회나 수상하여 최다 수상 챗봇으로 기네스북에도 등재되었다(Guinness World Records, 2020). 미추쿠는 사이버 인간과 같이 챗봇 캐릭터를 설정하여 인간과 같이 일관성있게 대화가 가능하도록 설계되어 있다. 미추쿠는 AIML 시스템에 저장된 정보를 이용하여 대화자의 발화에 포함된 핵심어를 인식·분석한 후 응답하거나 사용자가 대화 중 생성한 발화를 수집하여 응답에 활용하기도 한다. 미추쿠가 타 챗봇(예, Cleverbot)과 구별되는 가장 두드러진 특징은 AIML 시스템에 저장되어 있지 않은 정보를 대화자가 요구할 경우 웹검색을 하여 정보를 제공한다는 점이다. 본 연구도 이 기능에 착안하여 평가 과업을 구안하였다.

### 3. 평가 절차

#### 가. AI 챗봇 기반 온라인 쓰기 평가 문항 개발 및 적용

본격적인 실험에 앞서 참여자들의 영어 쓰기 능력을 진단하기 위해 <표 1>과 같은 쓰기 문항으로 오프라인 평가를 실시하였으며, 이후 동일한 학생을 대상으로 미추쿠를 활용한 온라인 평가를 실시하였다.

<표 1> 영어 쓰기 오프라인 평가 문항

---

학교에 CCTV 카메라를 설치하는 것의 장점(지지) 또는 단점(반대)에 대한 자신의 입장 한 가지를 선택하여 서론, 본론(근거 포함), 결론의 형식에 맞게 글을 작성하시오.  
(300-400단어 내외)

---

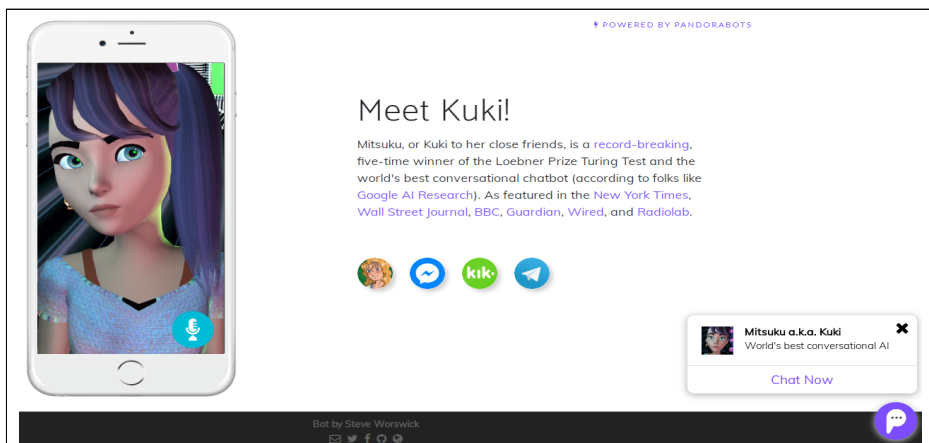
실험에 앞서 미추쿠와 채팅을 시도한 결과, 자유로운 채팅 활동은 학생 간 참여도에서 큰 편차가 발생한다는 사실을 확인하였다. 이러한 분석을 바탕으로 실제 쓰기 평가에서는 <표 2>와 같이 8단계의 구체적인 세부 과업을 설정하여 진행하였으며, 검색결과의 원문을 그대로 사용할 경우 표절로 처리되어 감점되므로 자신의 표현으로 바꿔 글을 작성하도록 학생들에게 공지하였다.

&lt;표 2&gt; 온라인 쓰기 평가의 세부 과업 수행 절차

| 단계 | 수행 과업   |
|----|---|
| 1  | 인사하기  |
| 2  | 자기소개(이름 등)하기 및 미추쿠의 신상정보 알아보기(예, 사는 곳, 취미, 결혼 여부, 가족 등)   |
| 3  | 하고 있는 일(직업) 알아보기, 사람들을 돕기 위해서 미추쿠가 할 수 있는 일 알아보기  |
| 4  | 미추쿠가 좋아하는 것들(이유 포함) 알아보기, 그 밖의 궁금한 점 물어보기   |
| 5  | 태국의 대표적인 관광지, 축제, 음식, 기후 등의 유무형의 문화적 정보 알아보기  |
| 6  | 스페인의 대표적인 관광지, 축제, 음식, 기후 등의 유무형의 문화적 정보 알아보기   |
| 7  | 대화 마무리하고 인사하기   |
| 8  | 대화 내용 복사하여 워드파일에 붙이고 그 아래 미추쿠가 제공한 검색 내용을 바탕으로 두 문화에 대한 특징을 비교하는 글을 300-400단어로 요약하여 기술하고 그 결과 파일을 구글 드라이브를 통해 공유폴더에 업로드하시오. |

도입부의 과업으로는 실제 채팅 상황과 유사하게 신변잡기를 대화의 주제로 제시하였다. 본 연구의 온라인 쓰기 평가의 범위에는 채팅 부분을 포함하지 않았고 다만 8단계에 제시한 300-400 단어의 비교글 쓰기에 한정하였다. 이에 따라 궁극적으로는 학생들이 직접 웹검색 엔진을 통해 정보를 찾아 비교 글을 작성하는 활동과 큰 차이가 없을 수 있다. 하지만 상호작용에 기반한 의사소통 과정이 반영된 채팅을 통해 일상적인 대화를 나누고 필요한 정보를 챗봇이라는 친구에게 질의하여 얻는 절차는 기존 쓰기 평가에서는 적용된 바 없으며 실제 상황적 맥락 속에서 구현된 평가라는 점에서 의의가 있다. 또한 이를 통해 학생들의 동기유발이나 참여도 개선 효과도 기대된다.

학생들은 위의 <표 2>에 제시한 과업지시서를 바탕으로 미추쿠 웹사이트(<https://www.pandorabots.com/mitsuku/>)에 직접 접속하여 평가에 참여하였다.

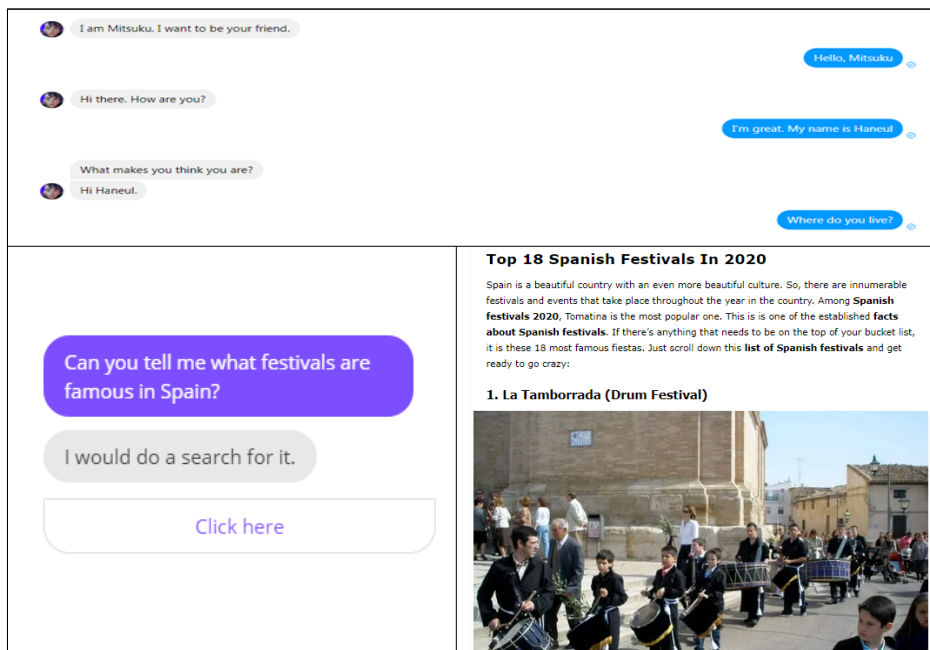


[그림 1] 미추쿠 홈페이지



[그림 1]에 제시된 바와 같이 미추쿠와의 채팅을 위해서는 구글의 크롬(Chrome) 브라우저를 통해 미추쿠 홈페이지에 접속해야 하며 우측 하단의 “Chat Now” 버튼을 클릭하여 데스크톱 또는 스마트폰을 통해 채팅이 가능하다.

다음의 [그림 2]는 <표 2>에서 제시한 세부 과업에 따라 실제 미추쿠와 채팅을 수행한 예시이다.



[그림 2] 과업지시서에 기반한 채팅 및 검색 결과 예시

[그림 2]에서 상단의 캡처는 1단계와 2단계의 과정 예시이고 좌측 하단의 캡처는 6단계에서 스페인의 축제 정보를 묻는 과정을 보여준다. 그리고 우측 하단의 캡처는 미추쿠가 제공한 스페인 축제 정보 사이트의 예시이다. 위의 절차를 통해 미추쿠와의 채팅이 종료되면 미추쿠가 제공한 태국과 스페인의 문화관련 사이트를 바탕으로 두 문화의 비교글을 작성하면서 영어 쓰기 평가가 마무리된다.

#### 나. AI 챗봇 기반 쓰기 평가 문항 채점 및 피드백 제공

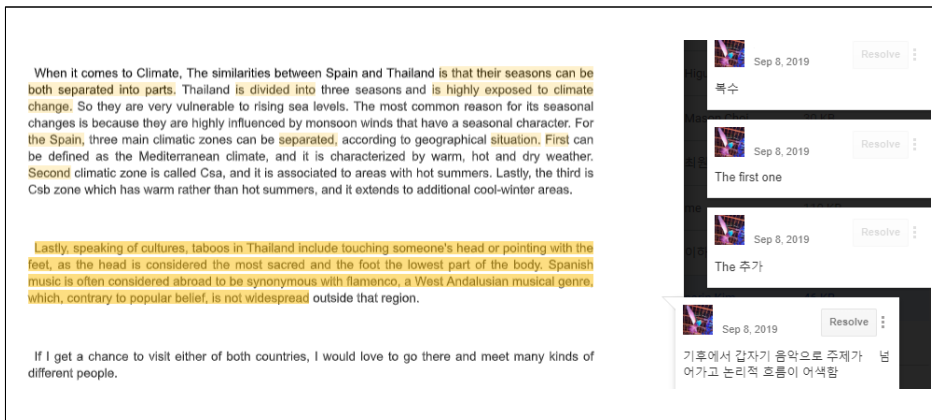
영어 쓰기 오프라인 평가 및 챗봇 활용 온라인 영어 쓰기 평가의 채점기준은 다음의 <표 3>와 같이 동일하다. 다만 <표 3>에 제시한 ‘과제수행’ 영역의 채점요소는 챗봇 활용 영어 쓰기 평가 문항을 기준으로 제시하였다.

<표 3> 영어 쓰기 채점기준

| 채점영역 | 과제수행             | 내용          | 구성       | 언어사용           |
|------|------------------|-------------|----------|----------------|
| 척도   | 0-3              | 0-5         | 0-3      | 0-5            |
| 채점요소 | -태국 문화 소개        | -내용의 주제 연계성 | -전개의 논리성 | -어휘사용의 적절성     |
|      | -스페인 문화 소개       | -내용의 적절성    | -비교의 적절성 | -문법사용의 정확성     |
|      | -300-400 단어 글 작성 | -부연설명의 충실도  | -내용의 일관성 | -철자 및 구두점의 정확성 |

채점영역은 4개로 과제수행, 내용, 구성, 언어사용을 측정하며 채점이 불가한 답안은 0점을 부여하도록 하였다. 또한 과제에 제시한 조건의 수행 여부를 보는 ‘과제수행’은 3점 척도, 부연 설명의 충실도를 주로 측정하는 ‘내용’은 5점 척도, 논리적 전개를 측정하는 ‘구성’은 3점 척도, 그리고 어휘사용의 적절성과 문법의 정확성 등을 측정하는 ‘언어사용’은 5점 척도를 적용하는 등 영역별로 차별화하여 채점기준을 설정하였다. 채점은 국가수준의 영어 쓰기채점에 경험이 풍부한 2명의 채점자가 채점 후 편차가 있을 경우 제 3의 채점자가 조정하여 확정하였다.

본 연구에서는 온라인 평가의 수행 및 피드백 제공에 모두 구글 드라이브를 활용하였다. 구글 드라이브에 강좌 공유방을 만들고 여기에 온라인 문항을 올리면 학생들이 미추쿠와 채팅을 한 후 비교글을 작성하여 업로드하였다. 구글 드라이브에 업로드된 문서는 구글 문서를 통해 다음의 [그림 3]과 같이 온라인 상태에서 읽고 피드백을 바로 제공할 수도 있다.



[그림 3] 구글 문서를 통한 피드백 제공 예시

피드백은 일반 워드프로세서의 메모 기능과 유사하나 한 가지 차이점은 교수자가 제공한 피드백에 채팅처럼 댓글을 달 수 있다는 점이다. 이를 이용하면 추가 질문이

나 동의/비동의 여부 표시 또는 동료 피드백 제공도 가능하여 온라인 교육 시 매우 유용하다.

## VI. 평가 결과

### 1. 온라인 영어 쓰기 채점 자료 분석 결과

본 연구에서는 앞서 설명한 바와 같이 오프라인 평가와 챗봇 활용 온라인 평가 결과를 비교하기 위해 2회에 걸친 영어 쓰기 평가를 실시하였다. 오프라인 평가에서는 전통적인 방식으로 외부 정보가 통제된 상태에서 쓰기 평가가 이루어졌고, 온라인 평가에서는 챗봇 미추쿠가 제공한 온라인 사이트의 정보를 활용한 쓰기 평가가 시행되었다. 오프라인 평가와 온라인 평가 점수를 비교한 결과 <표 4>와 같이 구성 영역을 제외한 과제수행, 내용, 언어사용 영역에서 모두 온라인 평가의 평균 점수가 높은 것으로 나타났으며, 결과적으로 이들 4개 영역의 점수를 합친 총점도 온라인 평가 점수가 높게 나타나 학생들이 미추쿠를 활용한 온라인 영어 쓰기 평가에서 더 나은 성과를 보여주고 있음을 알 수 있었다. 이는 영어 쓰기의 궁극적인 목적이 학생 변별이 아닌 더 나은 쓰기 결과물을 도출하는 것이라 점을 감안하면 매우 긍정적인 결과라고 해석할 수 있다.

<표 4> 영어 쓰기 오프라인/온라인 평가 결과 기술통계량(N=28)

| 구분       | 과제수행       |           | 내용         |           | 구성         |           | 언어사용       |           | 총점         |           |
|----------|------------|-----------|------------|-----------|------------|-----------|------------|-----------|------------|-----------|
|          | 오프라인<br>평가 | 온라인<br>평가 | 오프라인<br>평가 | 온라인<br>평가 | 오프라인<br>평가 | 온라인<br>평가 | 오프라인<br>평가 | 온라인<br>평가 | 오프라인<br>평가 | 온라인<br>평가 |
| 평균       | 2.86       | 3         | 3.71       | 4.32      | 2.75       | 2.5       | 3.32       | 3.82      | 12.64      | 13.64     |
| 표준<br>편차 | 0.36       | 0         | 0.98       | 0.61      | 0.44       | 0.58      | 0.67       | 0.77      | 1.75       | 1.16      |

다음으로 오프라인 평가와 온라인 평가 간 점수 차이가 통계적으로 유의한지를 검증하기 위해 대응표본 t검정을 실시하였으며 그 결과는 <표 5>와 같다.

<표 5> 영어 쓰기 오프라인/온라인 평가 점수 평균 차 검정

| 평가영역 | 대응차   |      |                 |                 |      | t     | 자유도 | 유의확률<br>(양쪽) |
|------|-------|------|-----------------|-----------------|------|-------|-----|--------------|
|      | 평균    | 표준편차 | 평균의<br>표준<br>오차 | 차이의 95%<br>신뢰구간 |      |       |     |              |
|      |       |      |                 | 하한              | 상한   |       |     |              |
| 과제수행 | -.14  | .36  | .07             | -.28            | .00  | -2.12 | 27  | .043         |
| 내용   | -.61  | .92  | .17             | -.96            | -.25 | -3.51 | 27  | .002         |
| 구성   | .25   | .80  | .15             | -.06            | .56  | 1.66  | 27  | .109         |
| 언어사용 | -.50  | .96  | .18             | -.87            | -.13 | -2.75 | 27  | .011         |
| 총점   | -1.00 | 1.74 | .33             | -1.68           | -.32 | -3.04 | 27  | .005         |

<표 5>에서 보여주듯이 과제수행, 내용, 언어사용 영역의 평균 차이는 모두 통계적으로 유의한 것으로 나타났으며(과제수행  $p<0.05$ , 내용  $p<0.01$ , 언어사용  $p<0.05$ ), 오프라인 평가에 비해 온라인 평가의 점수가 떨어진 구성 영역은 평균 차이가 유의하지 않은 것으로 나타났다. 영역별 점수를 합친 총점의 평균은 온라인 평가가 평균 1점 높은 것으로 나타났으며 통계적으로 유의한 것으로 나타났다( $p<0.01$ ). 이와 같은 결과는 학생들이 미추쿠를 활용한 온라인 평가에서 더 나은 영어 쓰기 능력을 보여주고 있음을 의미하며, 이는 학생들이 더 나은 영어 쓰기를 연습할 수 있는 학습 도구로도 본 연구에서 제시하는 온라인 평가가 활용될 수 있다는 가능성을 보여준다.

2. AI 챗봇 기반 쓰기 평가에 대한 설문 결과

본 연구에서는 또한 챗봇 활용 쓰기 평가에 대한 설문 조사를 실시하였다. 설문지는 설문 문항은 5점 리커트 척도의 선다형 설문 문항 5개와 개방형 설문 문항 4개로 구성하였다. 선다형 문항에서는 인간 대화와의 유사성, 영어 교육의 효과성, 정보수집의 용이성, 문화 교육의 효과성, 문화 교육의 동기유발성에 대해 질문하였으며 그 결과는 <표 6>과 같다.

&lt;표 6&gt; 챗봇 활용 쓰기 평가에 대한 설문 결과

| 질문  | 응답 빈도(비율)  |             |              |              |              |
|---|------------|-------------|--------------|--------------|--------------|
|   | ①          | ②           | ③            | ④            | ⑤            |
| • 미추쿠와의 대화 시 상대가 실제 어느 정도 인간처럼 느껴졌습니까?  |            |             |              |              |              |
| ① 매우 다름 ② 다름 ③ 보통   | 1<br>(3.6) | 5<br>(17.9) | 9<br>(32.1)  | 9<br>(32.1)  | 4<br>(14.3)  |
| ④ 유사함 ⑤ 매우 유사함  |            |             |              |              |              |
| • 미추쿠와의 대화가 영어 능력 향상에 어느 정도 도움이 될 수 있다고 생각하십니까?                                 |            |             |              |              |              |
| ① 전혀 도움 안됨 ② 거의 도움 안됨 ③ 보통  | 0<br>(0.0) | 0<br>(0.0)  | 4<br>(14.3)  | 17<br>(60.7) | 7<br>(25.0)  |
| ④ 어느 정도 도움됨 ⑤ 매우 도움 됨   |            |             |              |              |              |
| • 미추쿠와의 대화를 통해 주어진 과제에서 요구한 문화 정보의 수집이 쉬웠습니까?                                   |            |             |              |              |              |
| ① 매우 어려움 ② 어려움 ③ 보통   | 0<br>(0.0) | 2<br>(7.1)  | 10<br>(35.7) | 8<br>(28.6)  | 8<br>(28.6)  |
| ④ 쉬움 ⑤ 매우 쉬움  |            |             |              |              |              |
| • 미추쿠와의 대화가 문화교육에 어느 정도 도움이 될 수 있다고 생각하십니까?                                     |            |             |              |              |              |
| ① 전혀 도움 안됨 ② 거의 도움 안됨 ③ 보통  | 0<br>(0.0) | 1<br>(3.6)  | 7<br>(25.0)  | 15<br>(53.6) | 5<br>(17.9)  |
| ④ 어느 정도 도움됨 ⑤ 매우 도움 됨   |            |             |              |              |              |
| • 인공지능 챗봇 기반 평가를 활용한 문화교육이 기존의 영어 교과서에서 다루는 문화교육보다 흥미도나 참여도 유발에 도움이 된다고 생각하십니까? |            |             |              |              |              |
| ① 전혀 도움 안됨 ② 거의 도움 안됨 ③ 보통  | 1<br>(3.6) | 2<br>(7.1)  | 1<br>(3.6)   | 10<br>(35.7) | 13<br>(46.4) |
| ④ 어느 정도 도움됨 ⑤ 매우 도움 됨   |            |             |              |              |              |

설문 결과를 보면 미추쿠와의 대화가 영어 능력 향상에 어느 정도 도움이 될 수 있다고 생각하느냐에 대한 질문에 대해 85.7%는 도움이 된다고 응답(어느 정도 도움이 됨, 매우 도움이 됨)한 반면, 도움이 되지 않는다고 응답(거의 도움 안됨, 전혀 도움 안됨)한 학생은 없어 미추쿠와 같은 챗봇 기반 평가 문항이 학생들의 영어 실력 향상에 도움이 될 수 있음을 확인하였다. 또한 미추쿠와의 대화를 통해서 평가 과제로 수행한 문화 정보의 수집이 쉬웠다고 응답한 비율은 57.2%(쉬움, 매우 쉬움)인 반면에 그렇지 않다고 응답한 비율은 7.1%(어려움, 매우 어려움)밖에 되지 않아 학생들이 미추쿠와의 대화를 통해 정보 수집이 용이했던 것으로 파악된다. 마지막으로 인공지능 챗봇 기반 평가를 활용한 문화교육이 기존의 영어 교과서에서 다루는 문화교육보다 흥미도나 참여도 유발에 도움이 되는지 여부를 묻는 질문에 82.1%가 도움이 된다고 응답(어느 정도 도움이 됨, 매우 도움이 됨)하여 챗봇 기반 온라인 쓰기 문항이 문화교육이라는 평가내용에서 학생들의 흥미도와 참여도를 보다 효과적으로 증진시키는 역할을 충분히 할 수 있음을 보여주었다. 이상의 설문 결과는 비록 제한된 인원이 참여한 결과였기 때문에 일반화할 수는 없으나, 미추쿠와 같은 챗봇을 활용한 온

라인 평가가 기존의 평가 방식보다 더 유용할 수 있음을 시사하며 또한 원격교육 상황에서의 평가 방안으로 챗봇 기반 평가가 활용될 수 있는 가능성을 보여준다.

본 연구의 설문에서는 또한 학생들이 미추쿠와의 대화에서 좋았던 점이나 미추쿠가 영어 능력 향상에 도움이 된다고 생각하는 이유 그리고 향후 개선점에 대해 학생들에게 서술형 문항으로 질문하였다. 그 중 대표적인 응답을 하나씩 살펴보면 우선 미추쿠와의 대화에서 좋았던 점에 대해서 학생 A는 영어 사용 시 오류에 대한 두려움을 줄이고 자신감을 가지고 영어를 구사할 수 있는 환경이라는 점을 꼽았다. 이는 기존 연구(신동광, 2019; Fryer & Carpenter, 2006)의 결과와도 일치한다.

#### 학생 A

영어를 이용한 대화를 할 때 사람들이 가장 두려워하는 것들 중 하나는 바로 언어를 사용할 때 실수를 하는 것이다. 이런 의미에서 미추쿠와의 채팅을 오류가 발생해도 미추쿠가 이해하고 잘못을 정정해 주기 때문에 학생들이 더 자신있게 언어를 사용할 수 있는 것 같다.

다음으로 미추쿠와의 대화가 영어 능력 향상에 도움이 된다고 느낀 이유를 묻는 질문에 대해서 학생 B는 미추쿠와의 온라인 평가는 읽기와 쓰기 능력을 모두 필요로 하기 때문에 두 가지 언어 능력을 동시에 향상 시킬 수 있다는 점과 일방향이 읽기 형식이 아닌 상호작용의 대화라는 형식이 좀 더 효과적으로 영어를 공부 할 수 있도록 해준다고 응답하였다.

#### 학생 B

영어를 공부하는 데에는 우선적으로 읽기가 수반되고, 그 다음 심화된 과정으로 쓰기가 이어지는데 미추쿠와의 대화는 읽기 능력과 쓰기 능력을 모두 필요로 하기 때문에 채팅을 통해서 자연스럽게 영어 능력을 향상시킬 수 있겠다고 느꼈다. 또한, 미추쿠와 즐겁게 대화를 이어가고 나에게 필요한 정보를 얻기 위해서 등 나의 필요에 따른 영어 사용이기 때문에 책을 이용한 영어 활동보다 보다 효과적으로 영어 공부를 할 수 있다.

마지막으로 향후 개선점에 대해서 학생 C는 미추쿠에 더 많은 자료를 축적할 필요성을 제기하였다. 본 연구의 목적은 챗봇의 기능을 검증하는 데 있지 않고 이를 이용한 온라인 대화와 평가 문항의 개발 및 적용 가능성을 탐색하는 데 있다. 그럼에도 불구하고 미추쿠와 같은 챗봇이 아직 많은 부분에서 완벽하지 못하여 특히 학생 C가 지적한 바와 같이 데이터베이스를 확충하여 다양한 상황에서 보다 많은 자료를 학생들에게 제시할 수 있도록 개선되어야 할 것이며 이는 후속 연구를 통해서 이루어질 필요가 있다.

## 학생 C

미추쿠에 대한 정보를 알기 위해서 많은 질문을 해 보았는데, 미추쿠가 나의 질문에 대한 답으로 사진을 제시하였을 때 정말 사람같다는 생각이 들었다. 놀라워하면서 다른 사진은 없는지 "Send me another picture."라는 부탁을 했는데 준비된 사진이 없는지, 아니면 나의 명령어가 잘 못 되었는지 계속 동일한 사진을 보내는 부분이 있었다. 미추쿠의 데이터 베이스에 다양한 사진 자료가 제공된다면 더욱 효과적인 대화를 할 수 있을 것 같다.

## V. 결론 및 논의

본 연구에서는 원격교육 상황에서 평가 방안으로 인공지능 챗봇을 활용하여 대화 맥락을 제공한 영어 쓰기 평가 모델을 제안하고 그 가능성을 탐색하였다. 본 연구에서는 현재 상용 서비스를 제공하고 있는 대표적인 챗봇 가운데 하나인 미추쿠를 이용하여 영어의 쓰기 능력을 평가하기 위한 문항을 제안하였다. 또한 학생들을 대상으로 오프라인 평가와 챗봇 기반 온라인 평가를 통해 그 결과를 비교·분석하였으며, 설문을 통해 학생들에게 챗봇 기반 온라인 평가의 유용성과 개선점에 대한 의견을 조사하였다.

본 연구에서 제안한 온라인 쓰기 평가에서는 학생들이 미추쿠와의 대화를 통해 영어 글쓰기를 위한 초안을 잡을 수 있도록 돕고, 이후 미추쿠가 제공한 정보를 활용하여 두 문화를 비교하는 글을 완성하게 하였다. 평가에 참여한 학생들은 1차 쓰기 평가로 전통적인 오프라인 평가를, 2차 쓰기 평가로는 챗봇 기반 정보활용 쓰기를 수행하였다. 또한 이 두 평가 간 점수 차이를 보기 t 검정을 실시하였으며, t검정 결과 온라인 쓰기 평가가 대부분의 평가 영역(과제수행, 내용, 언어사용)에서 오프라인 평가보다 통계적으로 유의한 수준에서 높은 점수를 보였다. 뿐만 아니라 설문 결과에서도 학생들은 온라인 평가가 학생들의 전반적인 영어 능력 향상이나 영어 쓰기에 대한 흥미와 참여를 유발하는 데 도움이 될 수 있을 것이라고 응답하였다.

현대사회에서 요구되는 영어 쓰기 능력은 단순히 영어로 글을 쓰는 능력뿐만 아니라 정보검색 및 활용능력이 복합적으로 상호보완적인 관계를 유지해야 하며(주현우, 신동광, 2012), 본 연구 결과는 미추쿠를 이용한 온라인 평가가 이러한 영어 쓰기 능력을 적절히 평가할 수 있는 방안이 될 수 있음을 보여준다. 더욱이 온라인 평가에서는 정보통신기기를 이용한 학생들의 부정행위를 통제할 수 없기 때문에 학생들의 영어 쓰기 능력을 공정하고 정확하게 평가하는 데 많은 제한점이 있다. 따라서 온라인 평가가 공정성과 객관성을 담보하기 위해서는 필연적으로 문제해결식의 개별화된 평가가 이루어져야 하며 이런 측면에서 본 연구에서 제안하는 챗봇 기반 평가 모델은 유용한 대안이 될 수 있을 것이다. 또한 소규모 학급의 원격교육에서는 교사가 실시간 화상대화를 통해 본 연구와 같은 방식으로 평가할 수 있지만 교사와의 대화는 불

편하고 어려울 수 있다. 그러나 챗봇과의 대화는 심리적 부담이 적고 무한반복이 가능하다는 장점이 있기 때문에(양혜진, 김혜영, 신동광, 이장호, 2019) 소규모 학급에서도 챗봇을 활용한 영어 쓰기 평가가 여전히 유용하게 활용될 수 있을 것이다.

본 연구에서는 연구범위에 포함되지 않았지만 두 나라의 문화에 대한 비교글 외에 채팅 부분 또한 평가의 범위에 포함하기 위한 평가기준의 개발도 후속 연구를 통해 고려해 볼 수 있을 것이다. 지금까지 영어 쓰기 평가에서는 문장 쓰기 또는 문단 수준의 에세이 쓰기가 주를 이루었고 채팅과 같은 대화형식의 쓰기 문항은 평가에 반영되지 못하였다. 과거에는 파일 첨부는 이메일에서나 가능했으나 현재는 채팅앱을 통해서도 파일 첨부가 가능하고 실시간 접속 및 확인이 편리하며 의사소통의 트렌드가 음성 통화에서 채팅과 같은 문자 형태로 선호도가 변해감에 따라 채팅앱의 사용 비중이 급격히 늘고 있다. 이러한 사회적 현상을 고려하여 실제적인 쓰기 평가 구인을 반영한다면 채팅 형태로 영어 쓰기 능력을 평가하는 방식은 평가의 타당성에 매우 부합한다고 판단된다.

본 연구에서 제시한 챗봇 기반 온라인 평가 모델은 타 교과에도 적용 가능할 것이다. 예를 들어, 과학이나 사회과의 경우도 챗봇을 활용한 상호작용 모델을 적용하면 의사소통역량과 탐구역량을 동시에 신장시킬 수 있으며 발견식 또는 문제해결식 평가 문항의 개발이 용이해 질 수 있다. 또한 이 두 교과에서도 채팅 부분을 평가 범위에 포함하면 교과별 특성을 반영한 의사소통능력(예, 과학적 의사소통능력)을 평가할 수 있고 검색된 정보를 활용하는 부분에서는 비판적 사고능력이나 정보활용능력 등을 복합적으로 평가할 수 있을 것이다.



## 참고문헌

- 권형규 (2002). 온라인 상호작용 및 사용자 인터페이스에 따른 자기 주도적 평가 및 운영을 위한 웹 기반 평가 시스템의 설계 및 구현. **교육공학연구**, 18(2), 123-155.
- 교육부 (2015). **영어과 교육과정**(교육부 고시 제2015-74호). 세종: 교육부
- 교육부 (2020.2.23.). 전국 모든 유·초·중·고 신학기 개학 연기 결정 및 중국 입국 유학생 보호관리 방안 보완 조치 마련. 교육부 보도자료. <https://www.moe.go.kr/boardCnts/fileDown.do?m=020402&s=moe&fileSeq=5c0ed063d7c72672af3eada54b8129d5>에서 2020년 5월 1일 인출.
- 교육부 (2020.3.27.). 체계적인 원격수업을 위한 운영 기준안 마련. 교육부 보도자료. <https://www.moe.go.kr/boardCnts/fileDown.do?m=020402&s=moe&fileSeq=5ef226dc1c945b15474cf3415bd0ae79>에서 2020년 5월 1일 인출.
- 교육부 (2020.3.31.). 처음으로 초·중·고·특 신학기 온라인 개학 실시. 교육부 보도자료. <https://www.moe.go.kr/boardCnts/fileDown.do?m=020402&s=moe&fileSeq=8e7c9ed8c7fe1e590919393fc7a12237>에서 2018년 5월 1일 인출.
- 교육부 (2020.4.7.). 원격수업 출결평가기록 가이드라인 안내. 교육부 보도자료. <https://www.moe.go.kr/boardCnts/fileDown.do?m=020402&s=moe&fileSeq=3a85a3894ca95175c5c61eedf702ca03>에서 2020년 5월 1일 인출.
- 국가평생교육진흥원 (2016). **2016 글로벌 평생교육동향**(이슈페이퍼 11-B552881-0000 02-10). 서울: 국가평생교육진흥원. [http://www.ilec.go.kr/upfiles/Fboard\\_11/2016%20%EA%B8%80%EB%A1%9C%EB%B2%8C%ED%8F%89%EC%83%9D%EA%B5%90%EC%9C%A1%EB%8F%99%ED%96%A5\\_%EB%82%B4%EC%A7%80.pdf](http://www.ilec.go.kr/upfiles/Fboard_11/2016%20%EA%B8%80%EB%A1%9C%EB%B2%8C%ED%8F%89%EC%83%9D%EA%B5%90%EC%9C%A1%EB%8F%99%ED%96%A5_%EB%82%B4%EC%A7%80.pdf)에서 2020년 5월 1일 인출.
- 김성숙, 박찬호, 김미경, 김창환 (2018). 문항구성 모의실험을 통한 온라인 학업성취도 평가 시스템 분할점수 자동설정 방안 탐색. **교육평가연구**, 31(3), 631-651.
- 김인숙, 박지현, 임은영 (2018). 형성평가시스템 활용을 통한 단위학교의 형성평가 지원 방안. **교육과정평가연구**, 21(3), 181-199.
- 김재상 (2017). 인간과 인공지능 기기의 상호작용이 EFL 환경에서 초등학생들의 협업적 언어습득에 미치는 영향. 광주교육대학교 교육대학원 석사학위논문.
- 김혜영, 신동광, 양혜진, 이장호 (2019). 영어교과 보조 도구로서의 AI 챗봇 분석 연구. **학습자중심교과교육연구**, 19(1), 89-110.
- 박종임 (2015). 평가 패러다임 변화에 따른 온라인 형성평가 시스템의 발전 방향: 국어과 온라인 형성평가 시스템 사례 분석을 중심으로. **교원교육**, 31, 41-60.
- 신동광 (2019). 영어 쓰기 능력 향상을 위한 AI 챗봇 활용 방안 탐색. **교원교육**,

35(1), 41-55.

양혜진, 김혜영, 신동광, 이장호 (2019). 인공지능 음성챗봇기반 초등학교 영어 말하기 수업 연구. **멀티미디어언어교육**, 22(4), 184-205

이문복, 주현우 (2017). 온라인 기반 고등학교 영어 말하기 수행평가 지원 시스템 개발 및 시범적용. **멀티미디어언어교육**, 20(1), 167-185.

주현우, 신동광 (2012). 영어 쓰기 시험에서 콘코던서(Concordancer) 활용이 학생들의 영어 쓰기 수행력에 미치는 영향 (연구보고 RRE 2012-19). 서울: 한국 교육 과정평가원.

추성엽, 민덕기 (2019). 상호작용 촉진을 위한 과업 기반 AI 챗봇 활용 및 학생 발화 분석. **Primary English Education**, 25(2), 27-52.

황상연, 김두규, 임병민, 김정훈, 이재무 (1999). 웹을 기반으로 한 학습자 진단 및 조언 시스템 구현. **한국정보교육학회 동계학술대회 자료집**, 193-201.

Bachman, L. F. (2007). What is the construct? The dialectic of abilities and contexts in defining constructs in language assessment. In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. E. Turner, & C. Doe (Eds.). *Language testing reconsidered* (pp. 41-71). Ottawa, Canada: University of Ottawa Press.

Bachman, L., & Palmer, A. (2010). *Language assessment in practice*. Oxford, UK: Oxford University Press.

Fryer, L., & Carpenter, R. (2006). Bots as language learning tools. *Language Learning and Technology*, 10(3), 8-14.

Guinness World Records (2020, May 1). Most Loebner Prize wins. Retrieved May 4, 2020, from <https://www.guinnessworldrecords.com/world-records/603076-most-loebner-prize-wins/>.

Kim, N-Y. (2016). Effects of voice chat on EFL learners' speaking ability according to proficiency levels. *Multimedia-Assisted Language Learning*, 19(4), 63-88.

Wikipedia (2020, May 1). Mitsuku. Retrieved May 1, 2020, from <https://en.wikipedia.org/wiki/Mitsuku>.

논문 접수: 2020년 5월 6일

논문 심사: 2020년 7월 9일

게재 승인: 2020년 7월 27일

<ABSTRACT>

## A Study on the Online Assessment using Artificial Intelligence for Distance Education

Lee, Yongsang (Yeungnam University)

Shin, Dongkwang (Gwangju National University of Education)

In the present study, we proposed an online writing assessment model using an artificial intelligence chatbot in the distance education environment and explored its feasibility in English education. To this end, Mitsuku, one of the most advanced chatbots was adopted. The participants obtained cultural information through some chitchat with Mitsuku and referred to the information for writing their essay. In order to investigate the possibility of the chatbot-based online assessment for distance education, we compared the results of the traditional paper and pencil writing test and the chatbot-based online writing test. In addition, we also surveyed the students on the usefulness and improvement of the chatbot-based online test. The results show that the students gained significantly higher scores in the online test compared to those of the offline test. This is a critical finding, which means that the students could have a chance to improve their writing by chatting with Mitsuku. The survey results also show that the students believe chatting with Mitsuku would contribute to improving their English competence and increasing their interest and participation in distance education. In the current infodemic society, the competence of writing in English includes the information processing ability, and thus the English writing assessment using an AI chatbot proposed in this study would be an more appropriate way for assessing this kind of English competence. In conclusion, it was also suggested that the online assessment using AI could be applied to other subjects as well.

★ **Key words:** distance education, information-processing ability, infodemic society, online writing assessment, chatbot