

## 대규모 평가 서답형 문항 채점을 위한 문장 수준 자동채점 프로그램의 정확성 분석<sup>1)</sup>

송미영(한국교육과정평가원 연구위원)\*

노은희(한국교육과정평가원 연구위원)

성경희(한국교육과정평가원 부연구위원)

---

### <요 약>

본 연구는 문장 수준의 답안 작성을 요구하는 서답형 문항의 답안을 자동채점 프로그램으로 정확히 채점할 수 있는지 점검하고자 한다. 2014년 국가수준 학업성취도 평가의 국어·사회·과학 교과 서답형 6문항을 대상으로 문장 수준 자동채점 프로그램을 시범 적용하고, 교과전문가가 부여한 점수와 비교하여 채점 결과의 정확성을 분석함으로써 자동채점의 인간채점 대체 가능성을 검증하였다. 자동채점 결과의 정확성에 대한 통계량 모두 자동채점 과정에 채점자가 개입하는 경우가 개입하지 않는 경우보다 더 높은 수치를 보였는데, 인간채점 점수와 자동채점 점수 간 완전일치도는 96~99%, 카파계수는 0.76~0.99, 상관계수는 0.82~0.99, 정확률은 90~99%로서 전반적으로 문장 수준 자동채점 프로그램의 정확성은 매우 높았다. 한편, 기존 프로토타입을 보강한 채점 프로그램의 정확성 변화를 살펴보면, 채점자가 미개입하는 경우 인간채점과의 완전일치도, 카파계수, 상관계수, 정확률 평균은 기존 프로토타입 대비 각각 87%에서 96%로, 0.73에서 0.86로, 0.86에서 0.89로, 74%에서 83%로 높아졌다. 또한 자동채점과 인간채점 점수 간의 불일치 답안을 중심으로 교과별·문항별 채점 결과를 질적 분석한 결과, 학생 답안 유형 수, 채점기준의 조건과 점수 범주 등의 문항 특성에 따라 정확성이 다르게 나타남을 확인하였다. 이러한 분석 결과를 바탕으로 문장 수준 자동채점 프로그램을 적용한 대규모 평가 서답형 문항의 채점 설계를 제안하였다.

주제어 : 자동채점, 서답형 문항 채점, 채점 신뢰도, 채점자 간 일치도

---

1) 본고는 ‘한국어 문장 수준 서답형 문항 자동채점 프로그램 개발 및 적용’ (한국교육과정평가원 연구보고 RRE 2015-9)의 일부 내용을 보완하고 재구성한 것임.

\* 제1저자 및 교신저자, mysong@kice.re.kr

## I. 서론

창의적인 지식에 기반한 선진대국으로 도약하기 위해 정부는 2009 개정 교육과정을 도입한 이래 ‘창의인성’ 교육을 강조하면서 평가 방법의 변화를 추진해 오고 있다. 2009 개정 교육과정은 서술형이나 논술형 평가, 수행평가의 비중을 늘리도록 권고하고 있으며(교육과학기술부, 2009), 평가 방법에 대한 이러한 권고는 ‘창의융합형’ 인재 양성을 표방하는 2015 개정 교육과정에서도 이어지고 있다(교육부, 2015).

교육 현장에서 서답형 문항의 비중을 확대하도록 교육부에서 권고하는 이유는 학생이 생소한 바를 직접 구성하여 답하도록 하는 문항 특성에 있다. 이와 같은 특성으로 서답형 문항은 선택형 문항에 비해 문제해결력이나 창의력 등을 측정하는 데 적합하다고 알려져 있다. 그렇지만 실제 우리나라 교육 현장에서 서답형 문항의 활용은 저조한 편이다. 채점 결과의 객관성이나 공정성 확보 문제나 채점 소요 시간과 비용 등의 현실적인 문제가 서답형 문항의 활용을 제한한다.

서답형 문항을 효율적으로 채점하기 위해서 국외에서는 60년 이상의 역사를 가지고 자동채점 프로그램을 연구·개발하여 왔다. 특히 영어권을 중심으로 2000년대를 전후로 비교적 짧은 답을 요구하는 문항에 대한 자동채점 프로그램만 해도 37종이 개발되었다(Burrows et al., 2015). 자동채점의 타당성은 다양한 연구에서 입증되어(Shermis & Burstein, 2013) 여러 검사에 활용되고 있으며(Lewis, 2013; Zhang, 2013), 자동채점을 적용하면 대규모 평가의 채점 소요 비용을 대폭 줄어드는 것으로 보고되었다(Topol et al., 2014). 따라서 자동채점 프로그램은 서답형 문항의 채점을 돕는 효율적인 수단이 될 수 있으며, 이로써 국가수준이나 지역수준의 대규모 평가, 단위학교 내에서 이루어지는 평가에서 서답형 문항의 비중을 보다 확대할 수 있다.

영어권에서 축적된 자동채점 프로그램의 연구 결과들을 우리나라의 평가 장면에 단순히 수용하기에는 언어적인 특성상 무리가 있다. 또한 국내의 한국어 자동채점 관련 연구(예: 강원석, 2011; 권오영, 2004; 박희정, 강원석, 2003; 정동경, 2001; 조우진, 2006)에서 제시한 프로그램은 대규모 평가의 서답형 문항을 실제적으로 처리하기에는 한계가 있다. 따라서 한국어의 특징과 우리나라의 평가 특성에 맞는 서답형 문항 자동채점 프로그램의 개발이 필요하다.

이러한 필요성에 따라 단어·구 수준의 답안을 채점할 수 있는 프로그램 개발부터 시작하여(노은희 외, 2012; 2013; 2014), 문장 수준의 답안 작성을 요구하는 서답형 문항(이하 문장 수준 문항)의 답안을 채점하기 위한 자동채점 프로그램 개발 연구가 추진되고 있다(노은희 외, 2014). 본 연구는 지금까지 개발된 문장 수준 자동채점 프로그램으로 대규모 평가에서 서답형 문항의 채점을 어느 정도 정확히 처리할 수 있는지 점검하는 데 목적이 있다. 연구 목적을 달성하기 위하여 2014년 국가수준 학업성취도 평가의 문장 수준 문항을 대상으로 자동채점 프

로그그램을 시범 적용하고, 교과전문가에 의한 인간채점 점수와 비교하여 채점 결과의 정확성을 분석하고자 한다. 본 연구를 통해 자동채점의 인간채점 대체 가능성을 검증하고 대규모 평가에서의 문장 수준 자동채점 프로그램을 적용한 서답형 문항 채점 방안 설계를 위한 시사점을 도출할 수 있을 것이다.

## II. 한국어 문장 수준 자동채점 프로그램의 특징

### 1. 기계학습 기반의 자동채점

서답형 문항은 채점기준에 따라 각 답안에 대해 정오(正誤) 혹은 부분점수에 맞는 답안 유형인지를 판단하는 것으로 채점한다. 컴퓨터를 활용하여 서답형 문항을 자동채점한다는 것은 학생 답안들을 일정한 분류 기준에 의해 정답, 부분정답, 오답 등으로 분류하여 점수를 부여한다는 것이다. 기계학습 기반의 자동분류 방법은 지도학습, 비지도학습, 준지도학습, 앙상블 학습 등이 있다(Chapelle et al., 2006; Harrington, 2012). 기계를 통한 정확하고 일관성 있는 분류는 곧 채점의 정확성으로 연결된다. 따라서 한국어 문장 수준 자동채점 프로그램에 적합한 기계학습 방법을 선정하고자 각 방법의 장단점을 살펴보았다.

첫째, 지도학습(supervised learning) 방법은 기 채점된 학습용 답안의 정보기술적·통계적 특징을 분석하여 분류 기준을 찾고, 그에 따라 채점할 대상 답안을 특정 범주로 분류하는 것이다. 일반적으로 학습용 답안은 전문가를 활용하여 각 대상에 올바른 분류 결과를 부착하는데, 학습용 답안의 양이 충분할수록 채점 정확도는 높아진다. 즉, 지도학습 기반의 자동분류에 있어서 관건은 학습용 답안을 충분히 확보하여 투입하는 것이다(Rubin et al., 2012).

둘째, 비지도학습(unsupervised learning) 방법에 사용되는 학습용 답안은 정답/오답 등의 사전 분류 결과 정보가 없다. 학습용 답안을 내용상 유사한 것끼리 묶어 분류 기준을 마련한 후 채점해야 할 답안의 철자 및 의미 유사도를 고려하여 가장 유사한 집합에 포함시킨다. 각 집합이 정답인지 오답인지는 전문가가 최종 판단해야 하지만, 기계적으로 유사한 답안끼리 묶어서 제시함으로써 전문가가 일관성 있는 판단을 하도록 돕는다.

셋째, 준지도학습(semi-supervised learning) 방법은 지도학습 방법과 비지도학습 방법을 복합적으로 사용한다. 먼저 채점된 학습용 답안으로 분류 기준을 생성하고, 이를 바탕으로 미 채점된 학생 답안을 정답, 부분정답, 오답으로 분류한다. 분류된 답안 중에서 각 범주에 속할 확률이 높은(예, 99% 이상) 답안을 선별하여 기계학습에 활용한다. 확장된 학습용 답안으로부터 새로운 분류 기준을 마련하고 남아 있는 학생 답안을 분류한다. 이러한 과정을 반복하면서

학습용 답안의 크기를 점차 늘려 기계학습의 정확도를 개선할 수 있다(김수영, 조성배, 2002).

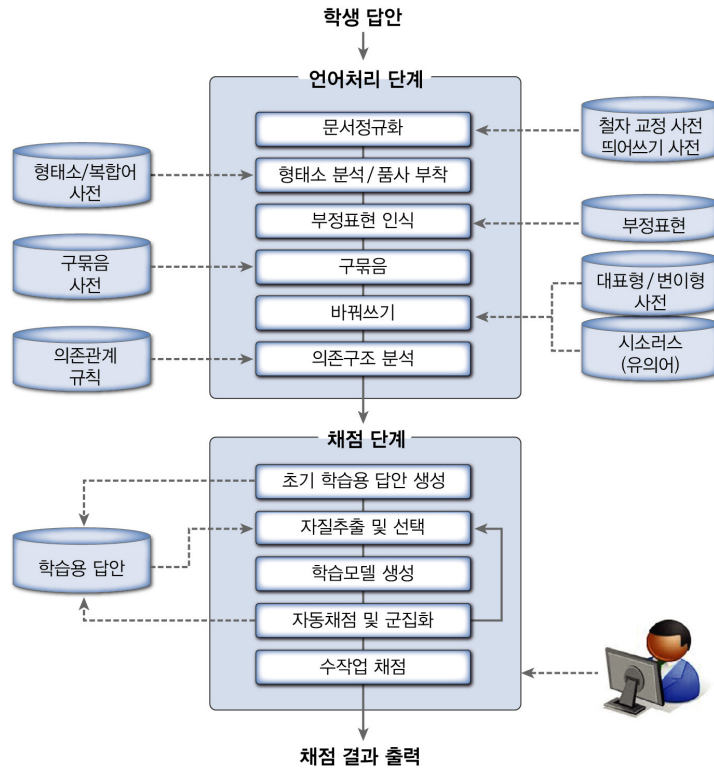
넷째, 앙상블학습(ensemble learning) 방법은 2가지 이상의 분류 기준을 생성하여 결과를 종합적으로 추론하는 방법이다. 서로 다른 기계학습 방법을 사용하거나 학습용 답안의 샘플을 다르게 구성하여 다양한 자동분류 기준을 만들고, 여러 방법에 의한 결과를 종합한다. 각 분류 방법에서 생성한 분류 결과를 조합하는 방법은 각 분류 방법의 특징에 따라 다수결이나 가중치 투표 방법 등을 활용할 수 있다(신현정 외, 2000). 앙상블학습 방법은 여러 개의 자동분류 방법을 종합하므로, 단일한 분류 기준만을 사용하는 것보다 정확도가 높은 결과를 얻을 수 있다.

이상에서 살펴본 기계학습 방법의 특징을 종합하면, 일반적으로 지도학습 방법이 준지도학습 방법보다 성능이 다소 우수한 것으로 알려져 있으나, 지도학습 방법은 그 성능을 높이기 위해 초기 학습용 답안 구축에 시간과 비용이 상당히 소요된다. 채점의 정확성은 물론 학습용 답안의 확보 용이성을 고려하여 어떠한 기계학습 방법을 활용할지 결정해야 한다. 학습용 답안이 충분하지 않은 상황에서 채점 성능이 담보되는 준지도학습 방법을 기반으로 하되, 다양한 기계학습 방법의 결과들을 조합하여 앙상블학습 방법을 선택하면 보다 정확한 채점 결과를 도출할 수 있다. 따라서 한국어 문장 수준 자동채점 프로그램은 준지도학습 방법과 앙상블 학습 방법을 기반으로 복합적으로 구현하였다.

## 2. 문장 수준 자동채점 프로그램의 채점 방식

국가수준 학업성취도 평가를 비롯한 우리나라 대규모 평가의 문항 유형과 한국어 처리 기술을 고려하여 문장 수준의 짧은 답안을 자동채점하는 프로그램의 개발 연구가 2014년부터 추진 중이다. 2014년에 설계한 한국어 문장 수준 자동채점 프로그램의 프로토타입(노은희 외, 2014)을 토대로, 2015년에는 각 단계별 세부 기능 구현에 필요한 기술을 보완·추가하여 문장 수준 자동채점 프로그램이 더 많은 답안을 정확히 처리할 수 있도록 개발하였다.

문장 수준 자동채점 프로그램의 전체적인 구조는 [그림 1]에 제시된 바와 같이 언어처리 단계와 채점 단계로 구성된다. 언어처리 단계에서는 학생 답안에 포함된 다양한 형태의 언어 정보를 분석하여 답안을 유형화한다. 그 다음 채점 단계는 학생 답안의 언어처리 결과로부터 군집화에 필요한 자질을 추출하고, 의미적으로 유사한 학생 답안을 군집화하여 채점을 지원하는 단계로서, ‘초기 학습용 답안 생성’, ‘자질추출’, ‘학습모델 생성’, ‘자동채점 및 군집화’, ‘수작업 채점’으로 이루어지며, 이때 기계학습 방법 중 준지도학습 방법과 앙상블학습 방법을 활용한다.



[그림 1] 2015년 개발 한국어 문장 수준 자동채점 시스템 구조도

문장 수준 자동채점 프로그램의 두드러진 특징은 반복 순환 구조의 채점 단계와 채점자 개입을 적절히 활용하여 자동채점 오류 가능성을 최소화한다는 점이다. 채점 단계에서는 먼저, 자질추출과 학습모델 생성을 통해 학생 답안을 분류하여 군집화하고, 채점이 완료된 학생 답안의 결과를 학습용 답안에 추가한다. 이와 같은 과정을 반복함으로써 자질의 신뢰성과 미채점 답안의 의미적 유사성을 높여 채점 결과의 정확성을 제고할 수 있다.

또한, 채점 단계의 ‘초기 학습용 답안 생성’, ‘자동채점 및 군집화’, ‘수작업 채점’ 과정에 채점자가 개입할 수 있도록 설계하였다. 기계학습에 사용할 초기 학습용 답안을 생성하는 과정부터 언어처리 단계를 거쳐 유형화된 학생 답안을 채점자가 채점하게 함으로써 자동채점의 정확성을 높일 수 있다. 뿐만 아니라, 기계학습을 통해 채점된 점수를 채점자가 반복 채점의 각 라운드마다 검토 후 확정지을 수 있도록 하여, 혹시라도 발생할 수 있는 자동채점 오류에 대한 점검의 기회를 제공하였다. 반복 채점을 거듭하여도 더 이상 자동 처리되지 않는 최종 미채점 답안을 채점자가 채점하는 ‘수작업 채점’을 끝으로 전체 자동채점 과정은 종료된다.

### Ⅲ. 연구 방법

#### 1. 분석 자료

문장 수준 자동채점 프로그램은 기계학습 자료원으로 교과전문가의 채점 결과가 사전 투입되어야 하고, 인간채점 결과는 자동채점의 정확성 평가를 위한 준거 역할을 하게 된다(Zhang, 2013). 이러한 점에서 본 연구는 해당 교과의 현직 교사가 기계점한 국가수준 학업성취도 평가 문항 중 최근 시행된 2014년 국어·사회·과학 교과의 서답형 문항을 시범 적용 대상으로 하였다.

<표 1> 문장 수준 자동채점 프로그램 시범 적용 대상 서답형 문항 정보

학년	교과	문항 번호	예시 정답	점수 범주	학생 답안 수
중3	국어	6-(2)-㉠	어려움을 극복할 수 있는 힘을 기를 수 있을 것이다.	0, 1	7,453
중3	사회	2	빙하가 침식 작용을 하였다.	0, 1, 2, 3	7,442
중3	과학	2-(2)	시간에 따라 점점 커진다.	0, 2	7,440
		8-(1)	가운데가 부풀어 오른 납작한 원반	0, 1, 2	7,440
고2	국어	2	할머니께서 동생에게 선물을 주셨다.	0, 1, 2	7,965
		6-(1)	우리는 자연에 순응해야 한다고 생각한다.	0, 1	7,965

\* 각 문항의 구체적인 내용은 [부록] 참조

문장 수준 자동채점 프로그램의 시범 적용 대상은 <표 1>의 문장 수준 문항에 대한 학년별·교과별·문항별 학생 7,440~7,965명의 답안이다. 현행 국가수준 학업성취도 평가는 지필 평가 체제에서 시행되므로, 자동채점을 시범 적용하기 위해서 학생들의 지필 답안을 전산 입력하였다. 이때 입력자 2명이 독립적으로 각각 입력하고 대조·수정하는 절차를 거쳐 답안 입력 오류를 최소화하였다.

동일 문항에 대해 총 4회의 자동채점이 이루어졌다. 2014년 프로토타입과 2015년에 정교화 개발한 프로그램의 정확성을 비교하기 위해 두 프로그램 모두 적용하였고, 채점자 개입 여부에 따른 영향을 파악하고자 두 프로그램의 반복 채점 과정에 채점자가 개입하는 경우(이하 ‘채점자 개입’)와 그렇지 않은 경우(이하 ‘채점자 미개입’)로 나누어 각각 채점하였다.

학생 답안을 자동채점하기 위한 초기 학습용 답안은 전체 문항에 대해 언어처리 단계 이후 동일하게 공백 답안과 이를 제외한 고빈도 답안 유형 10개를 수작업 채점하여 생성하였다. 이와 같은 기본값을 설정하여 문항별 57.3~91.99%의 답안이 초기 학습용 답안으로 포함되었다.

## 2. 분석 방법

<표 1>에 제시한 서답형 6문항에 대한 약 8,000개씩의 답안을 문장 수준 자동채점 프로그램으로 문항별 4회씩 채점한 결과를 대상으로, 자동채점의 정확성(accuracy)을 분석하였다. 자동채점 프로그램의 정확성 평가 준거로 교과전문가가 채점한 점수를 사용하는 것이 일반적이다(Shermis, 2010; Zhang, 2013). 본 연구에서도 각 문항에 대한 자동채점의 점수와 인간채점의 점수를 비교하였다. 여기서 인간채점 점수는 2014년 국가수준 학업성취도 평가에서 해당 교과와 현직 교사가 부여한 최종 점수를 가리킨다. 국가수준 학업성취도 평가의 서답형 문항별 최종 점수는 복수 채점한 점수를 재차 확인한 점수이므로 개별 채점자의 채점 오류 및 비밀관성이 제거된 기준 점수로 상정하고 비교 준거로 삼았다.

자동채점에 사용된 프로그램과 채점자 개입 여부의 각 조건에서 자동채점과 인간채점에 의한 점수가 어느 정도 일치하는지를 분석하기 위해 채점자 간 신뢰도 계수로서 상관계수, 완전 일치도(exact agreement), 카파계수(kappa coefficient)와 정보검색·분류 시스템의 성능 지표로서 정확률(precision)과 재현율(recall)을 산출하였다(Shermis & Burstein, 2013). ‘초기 학습용 답안 생성’과 ‘수작업 채점’에는 채점자의 판단으로 채점되기 때문에 [그림 2]와 같이 채점 답안을 분류하여 정확률과 재현율의 원래 의미를 해치지 않는 선에서 각각 다음과 같이 계산하였다.

$$\cdot \text{정확률: } P = \frac{a}{a+b}$$

$$\cdot \text{재현율: } R = \frac{a}{a+b+c+d}$$

		채점 결과가 올바른 답안 (인간채점 결과와 일치)	채점 결과가 틀린 답안 (인간채점 결과와 불일치)
자동 채점 대상 답안	시스템으로 자동채점된 답안	<i>a</i>	<i>b</i>
	시스템으로 자동채점되지 않은 답안 (=수작업 채점한 답안)	<i>c</i>	<i>d</i>
초기 학습용 답안		<i>e</i>	<i>f</i>

*a*: 올바르게 자동채점한 답안 수

*b*: 틀리게 자동채점한 답안 수

*c*: 올바르게 수작업 채점한 답안 수

*d*: 틀리게 수작업 채점한 답안 수

*e*: 올바르게 채점한 초기 학습용 답안 수

*f*: 틀리게 채점한 초기 학습용 답안 수

※ 초기 학습용 답안이 정확하게 생성되었으면 *f* = 0 이 됨.

[그림 2] 자동채점의 정확성 평가를 위한 답안 분류

정확률 *P*는 ‘자동채점 시스템으로 채점한 답안’ 중에서 실제로 ‘올바르게 자동채점한 답안’의 비율로서, 채점자의 판단 과정이 포함되지 않고 자동채점 시스템에 의한 채점 결과가 얼마

나 정확한가를 나타낸다. 재현율  $R$ 은 ‘전체 답안’에서 ‘초기 학습용 답안’을 제외한 ‘자동채점 대상 답안’ 중 ‘올바르게 자동채점한 답안’의 비율로서, 자동채점 대상 답안 중에서 자동채점 시스템이 정확히 채점한 정도를 나타낸다. 채점자가 개입하는 경우는, 자동채점의 반복 과정에서 채점자가 채점 결과를 확인하고 오류를 정정하므로, 채점자 개입 수를 제거하여 시스템으로 자동채점한 답안 수를 계산하였다.

본 연구에서는 인간채점 결과를 준거로 자동채점의 정확성을 평가하기 위해서 이상과 같이 평가 집합의 크기와 계산 방법이 다른 5가지 통계량을 산출하였다. 채점자 간 신뢰도 계수 3가지는 전체 답안이, 정보검색·분류 시스템의 성능 지표 중 재현율은 전체 답안에서 초기 학습용 답안을 제외한 자동채점 대상 답안이, 정확률은 자동채점 대상 답안에서 수작업 채점 대상으로 남는 답안을 제외한, 즉 기계학습으로 자동채점된 답안이 평가 집합이 된다.

추가적으로 자동채점과 인간채점 점수 간 불일치 답안을 중심으로 현재 자동채점 프로그램의 타당성과 한계를 가늠해 보기 위하여 각 서답형 문항과 답안 특성이 채점 결과에 끼치는 영향이 어떠한지를 질적 분석하였다. 또한 복수채점으로 진행되는 인간채점의 채점자 1인을 대체하여 자동채점을 적용하는 경우의 타당성을 확인하고자, 인간채점의 채점자 2인, 자동채점과 채점자 1인의 채점자간 신뢰도 계수를 각각 산출하였다.

## IV. 분석 결과

### 1. 자동채점의 정확성

문장 수준 자동채점 프로그램의 정확성은 실제 국가수준 학업성취도 평가에서 교과전문가가 여러 단계를 거쳐 확정한 인간채점 점수와 비교하여 분석하였다. 2014년 학업성취도 평가 국어·사회·과학 서답형 6문항의 인간채점 점수와 2015년 프로그램 및 2014년 프로토타입에 의한 자동채점 점수가 상호 일치하는 정도에 대한 5가지 통계량을 문항별로 산출한 결과와 그 평균은 <표 2>와 같다.

기존 프로토타입을 보강한 2015년 프로그램에 의한 자동채점 점수와 인간채점 점수 간 ‘완전일치도’의 범위는 채점자 미개입의 경우 93~99%, 채점자 개입의 경우 96~99%로 전체 문항에서 매우 높았다. ‘카파계수’는 채점자 미개입의 경우 .58~.99, 채점자 개입의 경우 .76~.99이었고, 중3 과학 8-(1)번을 제외하면 채점자 개입의 카파계수와 상관계수 모두 .92 이상으로 높았다. 자동채점 시스템에 의한 채점 답안 중에서 인간채점 점수와 일치하는 정도를 나타내는 정확률은 중3 과학 8-(1)번도 90%로 높은 정확률을 보였다. 따라서 전반적으로 문



장 수준 자동채점 프로그램의 정확성은 매우 높은 편이라 할 수 있다.

<표 2> 문장 수준 자동채점 프로그램의 문항별 채점 정확성

문항 번호	구분 채점자	2014 프로토타입					2015 프로그램				
		완전 일치도 (%)	카파 계수	상관 계수	정확률 (%)	재현율 (%)	완전 일치도 (%)	카파 계수	상관 계수	정확률 (%)	재현율 (%)
중3 국어 6-(2)-㉠	미개입	99.5	.990	.990	97.5	92.7	99.3	.985	.985	96.6	95.6
	개입	99.5	.991	.991	97.6	91.1	99.7	.994	.994	98.8	94.6
중3 사회 2	미개입	47.5	.278	.807	84.1	78.4	93.6	.888	.920	87.2	64.8
	개입	99.3	.987	.993	97.9	74.8	97.5	.957	.963	97.7	59.1
중3 과학 2-(2)	미개입	92.6	.781	.787	70.7	64.1	92.7	.784	.789	71.3	63.6
	개입	98.9	.964	.964	94.7	72.8	97.5	.921	.921	97.0	58.7
중3 과학 8-(1)	미개입	93.9	.570	.678	84.2	82.8	94.6	.579	.687	84.9	84.4
	개입	96.3	.762	.837	89.9	84.1	96.1	.756	.821	90.3	82.8
고2 국어 2	미개입	87.9	.770	.901	28.4	25.1	97.8	.953	.973	75.5	31.3
	개입	99.5	.989	.993	97.1	84.6	99.6	.992	.994	99.2	58.1
고2 국어 6-(1)	미개입	98.4	.966	.967	78.1	71.0	98.2	.982	.963	80.6	53.7
	개입	99.2	.982	.982	86.9	68.7	99.0	.963	.978	93.0	47.1
평균	미개입	86.6	.726	.855	73.8	69.0	96.0	.862	.886	82.7	65.6
	개입	98.8	.946	.960	94.0	79.4	98.2	.931	.945	96.0	66.7

<표 2>에 제시된 바와 같이 자동채점 결과의 정확성에 대한 통계량 모두 채점자가 개입하는 경우가 개입하지 않는 경우보다 더 높은 수치를 보였다. 이는 ‘채점자 미개입’에서는 자동채점의 반복 수행을 시스템에 의존하는 데 비해, ‘채점자 개입’에서는 채점 도중에 채점자가 자동채점 오류를 정정할 수 있기 때문이다. 즉, 기계학습에 의한 자동채점이 반복되는 동안에 채점자가 개입할 수 있도록 문장 수준 프로그램을 설계함으로써 채점의 정확성을 높인 것으로 볼 수 있다.

‘채점자 미개입’에서 자동채점의 반복 과정은 전적으로 기계에 의해서 진행되었기에 두 프로그램의 정확성을 비교하면, 채점자 개입 영향이 통제된 상태에서 프로그램 개선에 따른 정확성 향상 정도를 파악할 수 있다. <표 2>와 <표 3>에서 두 프로그램의 채점 정확성을 나타내는 통계량별 차이를 살펴보면, 2014년 프로토타입보다 2015년 프로그램에 의한 채점 결과의 정확성이 더 높은 편이었다. ‘채점자 미개입’의 경우 프로토타입에 비하여 인간채점과의 완전일치도는 평균적으로 87%에서 96%, 카파계수는 .73에서 .86, 상관계수는 .86에서 .89, 정확률은 74%에서 83%로 높아져 프로그램의 채점 정확성이 향상되었다. 특히 중3 사회 2번 문항의 경우 인간채점과의 완전일치도는 46%p나 대폭 향상되었다. 반면, 재현율은 3%p 감소하였

는데, 자동채점 대상 답안 중에서 미판단 답안으로 남겨진 답안이 프로토타입에 비하여 증가하였기 때문이다. ‘채점 단계’의 개선은 재현율보다 정확률에 더 무게를 두고 기계학습에 의한 자동채점의 오류를 줄이도록 기능을 강화한 결과이다.

<표 3> 2015년 문장 수준 자동채점 프로그램의 전년 대비 정확성 증감폭

구분 문항 번호	채점자 미개입					채점자 개입				
	완 전 일치도	카파계수	상관계수	정확률	재현율	완 전 일치도	카파계수	상관계수	정확률	재현율
중3 국어 6-(2)-㉠	▼0.2	▼.005	▼.005	▼0.9	▲2.9	▲0.2	▲.003	▲.003	▲1.2	▲3.5
중3 사회 2	▲46.1	▲.610	▲.113	▲3.1	▼13.6	▼1.8	▼.03	▼.03	▼0.2	▼15.7
중3 과학 2-(2)	▲0.1	▲.003	▲.002	▲0.6	▲.5	▼1.4	▼.043	▼.043	▲2.3	▼14.1
중3 과학 8-(1)	▲0.7	▲.009	▲.009	▲0.7	▲1.6	▼0.2	▼.006	▼.016	▲0.4	▼1.3
고2 국어 2	▲9.9	▲.183	▲.072	▲47.1	▲6.2	▲0.1	▲.003	▲.001	▲2.1	▼26.5
고2 국어 6-(1)	▼0.2	▲.016	▼.004	▲2.5	▼17.3	▼0.2	▼.019	▼.004	▼6.1	▼21.6
평균	▲9.4	▲.136	▲.031	▲8.9	▼3.3	▼0.6	▼.015	▼.015	▼0.1	▼12.6

\* ▲: 증가(+), ▼: 감소(-)

## 2. 문항별 채점 특성

자동채점 프로그램에 의한 채점 결과의 통계 분석을 바탕으로, 각 서답형 문항의 특성이 채점 결과에 끼치는 영향을 심층적으로 분석함으로써 현재 자동채점 프로그램의 한계를 가늠해 보았다. 문장 수준 프로그램의 적용 대상 문항 중 교과별로 정확성이 낮게 나타난 문항에 대해 자동채점과 인간채점 점수 간 불일치 답안을 중심으로 채점 결과를 분석한 결과는 다음과 같다.

먼저, 국어 문항 중 일치도 및 정확률이 가장 낮게 나타난 고2 국어 6-(1)번 문항의 예시 답안은 “우리는 자연에 순응해야 한다고 생각한다.”이고, 주어진 문장에서 ‘자연에’라는 부사어를 추가하여 답안을 완성하면 된다. 점수 범주의 수는 2개(0, 1점)이고, 채점기준도 비교적 간략하여 점수 부여를 위한 판단이 용이한 편에 속한다. 그렇지만 채점 불일치 답안이 82개(전체 답안의 1%)로 예상보다 다소 많아, 정확률이 93%로 낮은 편이었다. 이 중 절반 이상이 입력 오류에 기인하는데, ‘자연율’과 ‘자연에’라는 표현이 상호 잘못 입력된 경우가 대다수였다. 또한 채점기준에 따르면, ‘자연에’ 대신 ‘그것에’, ‘그에’, ‘이에’ 등과 같은 부사어를 유사 답안

으로 인정하는데, 자동채점 프로그램이 이를 포섭하지 못해 채점 불일치가 발생하였다. 이는 유사 답안을 추가하면 비교적 간단히 해결될 수 있는 문제이다. 한편, 이 문항의 경우, 인간채점의 채점자 간/내 일관성 문제도 다수 발견되었다. “우리는 자연에 순응해야 한다고 생각했다.”, “우리는 자연에 순응해야 한다고 생각합니다.”, “우리는 자연에 순응한다고 생각한다.”와 같은 동일한 답안에 교과전문가들은 0점과 1점으로 달리 점수를 부여하였다.

다음으로, 중3 사회 2번 문항은 완전일치도가 98%, 정확률이 98%로 나타났다. 이 문항의 예시 답안은 “빙하가 침식 작용을 하였다.”인데, ‘빙하’ 대신 ‘얼음’을, ‘침식 작용을 하였다’ 대신 ‘깎는다’, ‘마모되었다’, ‘부수다’, ‘파이게 하다’ 등과 같은 표현을 유사 답안으로 인정한다. 국어 교과와 달리 답안의 진술 조건에 문법적·형식적 제약이 없어 유사 표현이 다양하게 나타날 뿐만 아니라, 점수 범주의 수가 4개(0, 1, 2, 3점)로 부분 점수를 세밀하게 판정해야 하므로 채점에 어려움이 가중된다. 실제로 186개(전체 답안의 2%)의 채점 불일치 답안 목록을 살펴보면, 동일한 표현을 찾아볼 수 없을 정도로 학생들의 답안이 다양하게 나타났다. 여기에 정오 판정과 관련 없는 불필요한 정보까지 같이 진술되어 있는 경우도 불일치 답안의 상당 부분을 차지하고 있었다. 이러한 학생 답안의 높은 다양성은 결국 정확한 군집화를 어렵게 만들고 자동채점의 오류로 이어진 것으로 추측된다.

마지막으로, 중3 과학 8-(1)번 문항은 자동채점과 인간채점 점수 간 완전일치도가 96%로 다른 문항에 비해 다소 낮고, 채점 불일치 답안이 290개(전체 답안의 4%)로 다른 문항에 비해 많은 편이다. 이는 인간채점과 자동채점 간 채점기준의 적용 및 해석이 각각 달라 발생한 것으로 추정된다. 가령 이 문항의 예시 답안은 “가운데가 부풀어 오른 납작한 원반”인데, 채점기준의 유의사항에 따르면 ‘원반’이라는 단어가 반드시 포함되어야 한다고 명시되어 있다. 그러나 실제 인간채점에서는 ‘원반’ 모양과 유사한 형태를 지칭하는 ‘접시’, ‘쟁반’, ‘원판’ 등도 유사 답안으로 인정하였다. 이는 자동채점 과정에서 유사 답안으로 처리하면 간단히 해결될 수 있는 문제로 자동채점의 오류라고 보기는 어렵다. 물론 자동채점 프로그램의 한계로 나타난 불일치 답안도 존재한다. 우리 은하의 모습을 그림으로 표현한 경우 자동채점 프로그램은 처리하지 못했다. 또한 자동채점의 군집화 오류도 발견되는데, ‘반원’에 대해 자동채점은 1점을 부여하였다. 이는 ‘원반’이라는 용어의 영향으로 인해 잘못 군집화한 것으로 추정된다. 뿐만 아니라 ‘직선 중간에 구가 있는’, ‘원 양쪽에 선 있는’ 등과 같이 학습모델 생성을 통해 군집화하기 어려운 특이 답안은 자동채점이 0점을 부여하였는데, 이 역시 답안 군집화 단계의 개선이 요구되는 사안이라 하겠다.

질적인 문항 분석 결과를 종합해 보면, 자동채점은 인간채점과 비교하여 문항 특성에 영향을 크게 받는다. 즉, 답안 작성의 조건, 채점기준, 점수 범주, 다양한 학생 답안 반응 등 각 문항의 특성이 자동채점 결과에 복합적으로 영향을 끼친다. 문장 수준 답안의 경우 다양한 유사 표현을 정확하게 군집화하여 해당 점수를 산출하는 것이 관건인데 자동채점 프로그램의 군집

화 성능이 이를 따라가지 못할 경우에는 채점 정확도가 떨어진다. 이에 여러 교과목의 다양한 문항에 대한 유사 답안을 포섭할 수 있도록 보다 고도화된 군집화 방법의 보완이 요구된다.

### 3. 자동채점의 채점자 대체 적용

본 연구의 자동채점 시범 적용 대상인 국가수준 학업성취도 평가의 서답형 문항은 총 4차에 걸쳐 채점이 진행된다. 동일 답안을 2인의 채점자가 독립적으로 채점하여 채점자 간 점수가 일치하면 최종 점수로 확정하고, 점수 차이가 있는 답안은 다음 차수에서 재채점한다. 본 연구에서는 1차 채점에서 채점자 1인을 자동채점으로 대체하는 것으로 가정하고, 채점자 간 신뢰도를 산출하여 인간채점의 경우와 비교하였다. 기존 인간채점의 채점자별 점수와 자동채점 점수 간 완전일치도와 상관계수는 <표 4>와 같다.

<표 4> 인간채점 대비 자동채점의 정확성

구분 문항 번호	인간채점 (채점자1 vs. 채점자2)		자동채점의 채점자 1인 대체			
	완전일치도(%)	상관계수	(채점자1 vs. 자동채점)		(채점자2 vs. 자동채점)	
			완전일치도(%)	상관계수	완전일치도(%)	상관계수
중3 국어 6-(2)-㉠	99.2	.983	99.3	.987	99.4	.988
중3 사회 2	98.7	.986	97.1	.957	97.0	.958
중3 과학 2-(2)	99.2	.976	97.4	.917	97.3	.914
중3 과학 8-(1)	99.5	.982	96.1	.820	96.0	.813
고2 국어 2	99.5	.993	99.5	.993	99.5	.992
고2 국어 6-(1)	98.8	.974	98.5	.969	98.7	.974

자동채점 시범 적용 대상 6문항에 대한 인간채점의 두 채점자 간 신뢰도를 살펴보면, 완전 일치도는 99%, 상관계수는 .97~.99이고, 채점자 1인과 자동채점 프로그램에 의한 두 점수 간 완전일치도는 96~99%, 상관계수는 .81~.99이었다. 상대적으로 낮은 상관계수를 보이는 문항은 중3 과학 8-(1)번 문항으로, 이 문항을 제외하면 .92 이상으로 높은 편이었다. 전반적으로 채점자 1인을 자동채점으로 대체하여도 채점자 간 신뢰도가 기존과 유사하며 그 폭이 크지 않았다. 그 차이는 문항별 채점 특이 사항을 분석한 결과에서 기술하였듯이 답안 입력 오류 정정이나 채점기준의 세부 지침 명시 등으로 해소될 수 있을 것이며, 두 점수 간 불일치 답안을 채점하는 2차 채점에서 더 높은 일관성을 확보할 수 있는 여지가 있다. 이와 같은 결과는 실제 자동채점을 적용하여 채점하는 방안이 적절한지에 대한 경험적 증거가 될 수 있다.

## V. 결론 및 제언

본 연구는 한국어 문장 수준 자동채점 프로그램을 대규모 평가 서답형 문항에 시범 적용하여 채점 정확성을 분석함으로써 자동채점 적용의 타당성을 점검하는 데 목적이 있다. 한국적 대규모 평가 상황을 고려하여 설계한 자동채점 프로그램의 프로토타입과 세부 기능을 보강한 프로그램을 모두 적용하여 2014년 국가수준 학업성취도 평가 국어·사회·과학 교과와 서답형 6문항에 대한 문장 수준 답안을 채점하고 그 정확성을 분석하였다.

채점 신뢰도 계수를 산출한 결과, 인간채점 점수와 자동채점 점수 간 완전일치도, 카파계수, 상관계수 모두 매우 높았고, 자동채점 과정에 채점자가 개입하는 경우가 전혀 개입하지 않는 경우보다 채점 정확성을 나타내는 통계량 모두 더 높은 수치를 보였다. 우리나라 대규모 평가의 채점 오류는 결정적인 문제가 될 수 있기 때문에 본 연구의 자동채점 프로그램은 기술적으로 보강된 동시에 채점 진행 과정마다 채점자의 판단과 그에 따른 조정을 통해 정확도를 기하여, 결과적으로 매우 양호한 수준의 신뢰로운 점수를 산출한 것으로 분석된다.

한편, 자동채점 과정의 채점자 개입 효과를 통제하고 기계학습에 의존한 자동채점 결과와 인간채점과의 완전일치도, 카파계수, 상관계수, 정확률은 기존 프로토타입에 비하여 각각 평균적으로 10%p, .10, .03, 9%p만큼 높아진 것으로 나타났다. 정교화한 프로그램의 채점 정확성이 프로토타입보다 향상된 것으로 평가할 수 있으며, 이는 기계학습에 의한 자동채점 오류를 줄이도록 프로그램의 세부 기능을 강화한 정교화 방향과 상통한다.

교과별·문항별 채점 결과의 정확성은 차이를 보이는데, 답안 작성의 조건 및 채점기준, 점수 범주, 학생의 다양한 답안 반응 등에 영향을 받는 것으로 분석되었다. 답안 작성의 형식적 조건이 모호하거나 채점기준이 복잡할수록 다양한 학생 답안을 유사한 답안끼리 군집화하고 점수 판정을 내리기가 용이하지 않게 된다. 또한 채점기준에 교과별 유사 답안을 구체화하여 제시하지 않은 경우, 자동채점 프로그램은 이를 자율적으로 판단하지 못한다. 이러한 결과는 모든 서답형 문항에 자동채점 프로그램을 일괄적으로 적용하기에는 한계로 작용할 수 있다. 따라서 자동채점 프로그램을 보다 많은 서답형 문항에 적용하여 채점 결과를 누적하고, 정확한 채점이 가능한 문항과 그렇지 못한 문항의 특성을 체계적으로 분류하여, 이를 바탕으로 프로그램의 성능을 개선하고 자동채점 적용의 세부 지침을 마련할 필요가 있다.

한편, 자동채점과 인간채점 간 불일치 답안 분석에서 채점자의 피로도와 채점기준의 주관적 적용으로 인한 인간채점의 일관성 문제가 발견되었다. 특정 문항에 대해 일관된 기준으로 장시간 채점해야 하는 경우 인간채점의 일관성을 점검하는 데 자동채점을 활용할 수 있다. 뿐만 아니라 자동채점 프로그램은 채점자를 대체하여 일관되고 신속하게 채점할 수 있다. 단, 문항 유형이나 특성 등에 따라 가용 프로그램 및 채점 정확성, 채점 결과에 대한 민감도 등을

고려하여 자동채점 의존도와 적용 방식을 결정하여야 한다(Lottridge et al., 2013).

고부담 대규모 평가에 자동채점 프로그램을 적용하는 방안으로 먼저, 현재 동일 답안의 복수채점에 동원되는 교과전문가 1인의 역할을 자동채점이 대체하는 방안부터 고려해 볼 수 있다. 본 연구 결과에 따르면 자동채점 프로그램은 인간채점과 매우 높은 수준의 일치도를 보이므로, 채점자를 대체할 수 있을 것이다. 교과전문가 2인에 의한 복수채점의 부분 대체 수단으로 자동채점을 적용하는 경우, 채점자 1인은 자동채점 프로그램으로 채점하고, 다른 채점자는 기존과 같이 인간채점을 수행한다. 자동채점과 인간채점을 병행하여 동일 답안을 복수채점한 두 점수가 일치하면 최종 점수로 확정하고, 불일치한 경우는 다음 시기에 재채점하는 방식을 취할 수 있다. 본 연구에서 채점자 1인을 자동채점으로 대체하는 경우의 채점자 간 신뢰도를 인간채점의 경우와 비교할 때 매우 양호한 결과를 확보할 수 있음을 확인하였다.

다음으로 교과전문가 2인 이상이 자동채점 프로그램으로 각각 채점하는 방안의 운영도 가능하다. 자동채점 과정에 채점자의 판단이 요구되는 부분이 있으므로, 고부담 대규모 평가인 경우 채점자에 따라 다른 결과를 부여할 가능성을 줄여야 한다. 채점자 2인이 자동채점 프로그램으로 각자 독립 채점하여 두 점수가 일치하면 최종 점수로 확정하고, 불일치 답안을 다시 자동채점한 다음, 2차 자동채점 결과가 불일치한 경우는 제3의 채점자가 점검하여 점수를 확정하는 절차를 설계해 볼 수 있다. 이러한 절차에 따라 2015년 국가수준 학업성취도 평가 중 학교 3학년 사회 교과와 표집 채점에 단어·구 수준 자동채점 프로그램을 실제 적용한 바 있다(노은희 외, 2015).

대규모 평가나 교실 평가에서 자동채점을 적용하게 되면 서답형 문항의 비중 확대를 기대할 수 있다. 그런데 지필평가 체제에서 자동채점을 수행하려면 학생들이 직접 수기로 작성한 답안의 전산 입력이 필수적이다. 아직 지필 답안을 기계에 의해 전자 답안으로 자동 변환하는 수단은 확보되지 못하여 본 연구에서는 수작업으로 입력하는 방법을 택하였다. 수작업 입력의 수고를 덜고 자동채점을 폭넓게 적용하기 위해서는 지필 답안을 전산화하는 도구가 마련되어야 한다. 기술적 측면에서 답안의 전산화 및 자동채점이 가능한 환경이 되더라도 평가 주체나 대상이 자동채점을 불신하거나 거부감을 가지면 자동채점 적용의 걸림돌이 된다. 따라서 각종 평가의 관련자와 다양한 이해 당사자들에게 자동채점이 지니는 강점과 더불어 보완적 조치에 대해 분명히 안내하여 자동채점의 적용에 대한 긍정적 인식을 확산해야 할 것이다.

## 참 고 문 헌

- 강원석(2011). 질의문 유형 분석을 통한 서답형 자동채점 시스템. **한국콘텐츠학회논문지**, 11(2), pp. 13-21.
- 교육과학기술부(2009). **2009 개정 교육과정 총론**(교육과학기술부 고시 제 2009-41호).
- 교육부(2015). **초·중등학교 교육과정 총론**(교육부 고시 제 2015-74호).
- 권오영(2004). 웹 기반 주관식 평가문항 채점 알고리즘 설계 및 구현. 한서대학교 교육대학원 석사학위 논문.
- 김수영, 조성배(2002). 레이블이 없는 문서를 이용한 SVM 기반의 점증적 지도학습. **한국정보과학회 학술발표논문집**, 29(1B), pp. 301-303.
- 노은희, 심재호, 김명화, 김재훈(2012). 대규모 평가를 위한 서답형 문항 자동채점 방안 연구. 한국교육과정평가원 연구보고 RRE 2012-6.
- 노은희, 김명화, 성경희, 김학수(2013). 대규모 평가를 위한 서답형 문항 자동채점 프로그램 정교화 및 시범 적용. 한국교육과정평가원 연구보고 RRE 2013-5.
- 노은희, 이상하, 임은영, 성경희, 박소영(2014). 한국어 서답형 문항 자동채점 프로그램 개발 및 실용성 검증. 한국교육과정평가원 연구보고 RRE 2014-6.
- 노은희, 송미영, 성경희, 박소영(2015). 한국어 문장 수준 서답형 문항 자동채점 프로그램 개발 및 적용. 한국교육과정평가원 연구보고 RRE 2015-9.
- 박희정, 강원석(2003). 유의어 사전을 이용한 주관식 문제 채점 시스템 설계 및 구현. **컴퓨터교육학회 논문지**, 6(3), pp. 207-216.
- 신현정, 장민, 조성준, 이봉기, 임용업(2000). 앙상블 학습알고리즘의 일반화 성능 비교: OLA, Bagging, Boosting. **한국정보과학회 학술발표논문집**, 27(1B), pp. 226-228.
- 정동경(2001). 벡터 유사도와 시소러스를 이용한 주관식 답안의 채점 방법. 동국대학교 교육대학원 석사학위 논문.
- 조우진(2006). 의미 커널과 한글 워드넷에 기반한 지능형 채점 시스템. 한림대학교 대학원 석사학위 논문.
- Burrows, S., Gurevych, I., & Stein, B. (2015). The Eras and Trends of Automatic Short Answer Grading. *International Journal of Artificial Intelligence in Education*, 25(1), 60-117.
- Chapelle, O., Schölkopf, B., & Zien, A. (2006). *Semi-supervised Learning*. The MIT Press.

- Harrington, P. (2012). *Machine Learning in Action*. Manning Publications.
- Lewis, J. K. (2013). Ethical Implementation of an Automated Essay Scoring (AES) System: A Case Study of Student and Instructor Use, Satisfaction, and Perceptions of AES in a Business Law Course. Retrieved May 20, 2015, from [http://digitalcommons.salve.edu/cgi/viewcontent.cgi?article=1047&context=fac\\_staff\\_pub](http://digitalcommons.salve.edu/cgi/viewcontent.cgi?article=1047&context=fac_staff_pub)
- Lottridge, S. M., Winter, P. C., & Mugan, L. (2013). *The AS Decision Matrix: Using Program Stakes and Item Type to Make Informed Decisions about Automated Scoring Implementations*. Final research report delivered by Pacific Metrics Corporation. Monterey, CA.
- Rubin, T. N., Chambers, A., Smyth, P., & Steyvers, M. (2012). Statistical Topic Models for Multi-label Document Classification. *Machine Learning*, 88(1-2), 157-208.
- Shermis, M. D. (2010). Automated Essay Scoring in a High Stakes Testing Environment. In V. J. Shute, & B. J. Becker (Eds.), *Innovative Assessment for the 21st Century* (pp. 167-185). New York: Springer.
- Shermis, M. D., & Burstein, J. (2013). *Handbook of Automated Essay Evaluation: Current Applications and New Directions*. New York: Routledge.
- Topol, B., Olson, J., & Roeber, E. (2014). Pricing Study: Machine Scoring of Student Essays. Retrieved July 13, 2015, [http://www.assessmentgroup.org/uploads/ASAP\\_Pricing\\_Study\\_Final.pdf](http://www.assessmentgroup.org/uploads/ASAP_Pricing_Study_Final.pdf)
- Zhang, M. (2013). Contrasting Automated and Human Scoring of Essays. *ETS R&D Connections*, 21, Retrieved April 2, 2015, from <http://www.ets.org/Media/Research/>

·논문접수 : 2016.01.05. / 수정본접수 : 2016.02.01. / 게재승인 : 2016.02.16.



## ABSTRACT

### Analysis on the Accuracy of Automated Scoring for Korean Large-scale Assessments

**Mi-Young Song**

Research Fellow, Korea Institute for Curriculum and Evaluation

**Eun-Hee Noh**

Research Fellow, Korea Institute for Curriculum and Evaluation

**Kyung-Hee Sung**

Associate Research Fellow, Korea Institute for Curriculum and Evaluation

In order to effectively manage the scoring of constructed-response (CR) items for Korean large-scale assessments, this study aims to implement the automatic scoring (AS) system for sentence-level responses based on the prototype designed in 2014 and to score CR items for the National Assessment of Educational Achievement (NAEA) 2014 using the AS system.

We scored answers to six CR items of Korean language, social studies, and science in the NAEA 2014 using the AS system for sentence-level responses. The scores from the AS system were highly consistent with the scores by human scoring showing 96.1 to 99.7% of the exact agreement and 0.82 to 0.99 of correlation coefficients between the scores from two scoring methods. The exact agreements for the AS system this year were higher than those for the prototype. This result means that the performance of the AS system has been improved.

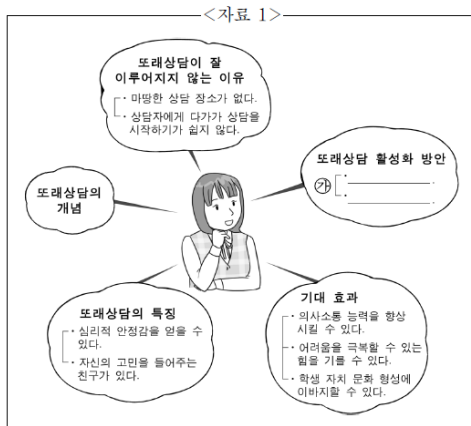
The AS program for Korean CR items is designed a human-machine collaborative and stepwise scoring method to guarantee the accuracy of scoring. This study provides evidence that automated scoring might be a reliable and efficient and could serve as a useful complement to human scoring for a large-scale assessment.

*Key Words : Automated scoring, Rater consistency*

## [부록] 한국어 문장 수준 자동채점 프로그램 적용 문항 및 예시 정답

- 2014년 중3 국어 6-(2)-㉠번 문항

**【서답형 6】** 토래상담에 대한 글을 학교 신문에 실고자 한다. 다음 물음에 답하시오.



- <자료 2>
- ㄱ. 토래상담을 위해 교내에 편안한 분위기의 장소를 마련해야 한다.
  - ㄴ. 토래상담은 교육 받은 토래가 상담자가 되는 비전문가 상담이다.
  - ㄷ. 상담자와 참여자가 정서적으로 과도하게 밀착하여, 문제 상황을 객관적으로 인식하지 못한다.
  - ㄹ. 상담자의 역할에 대한 교육이 증대되어야 하고, 토래상담 프로그램이 다양화되어야 한다.

(2) <자료 1>의 '기대 효과'를 글로 써 보았다. ㉠과 ㉡을 각각 한 문장으로 쓰되, ㉡은 <자료 1>의 '토래상담의 특징'을 근거로 쓰시오.

< 글 >

토래상담의 기대 효과는 크게 세 가지 측면에서 살펴볼 수 있다. 상담자는 의사소통 능력을 향상시킬 수 있을 것이다. 왜냐하면 상담을 하면서 경청하고 공감하는 자세를 기를 수 있기 때문이다. 참여자는 ㉠ . 왜냐하면 ㉡ . 학교 내에서 학생 자치 문화가 형성될 것이다. 왜냐하면 학생들 스스로 문제를 해결할 수 있는 능력이 증대되기 때문이다.

㉠ : \_\_\_\_\_.

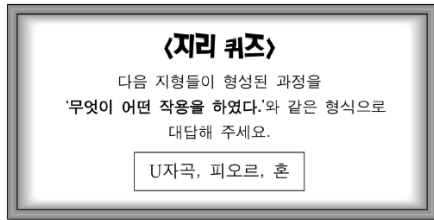
㉡ : \_\_\_\_\_.

- 예시 정답

㉠ : 어려움을 극복할 수 있는 힘을 기를 수 있을 것이다.

- 2014년 중3 사회 2번 문항

【서답형 2】 (가)에 들어갈 지형 형성 과정을 한 문장으로 쓰시오.



정답은 (가)입니다.

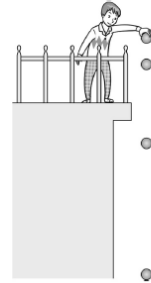
네, 정답입니다.

(가) ( )

- 예시 정답  
빙하가 침식 작용을 하였다.

- 2014년 중3 과학 2-(2)번 문항

【서답형 2】 그림은 낙하하는 공의 위치를 일정한 시간 간격으로 나타낸 것이다.

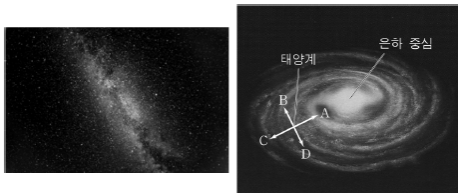


(2) 낙하하는 동안 공의 속력은 어떻게 변하는지 쓰시오.

- 예시 정답  
(2) 시간에 따라 점점 커진다.

- 2014년 중3 과학 8-(1)번 문항

【서답형 8】 그림 (가)는 여름철 은하수의 모습이고, (나)는 위에서 본 우리 은하의 모습을 나타낸 것이다.



(가)

(나)

(1) ( )안에 들어갈 알맞은 내용을 쓰시오.

(가)와 같이 은하수가 밝은 띠 모양으로 관측되는 이유는 우리 은하를 옆에서 보면 ( ) 모양이고, 태양계가 우리 은하의 중심에서 약 3만 광년 떨어진 나선팔에 위치하기 때문이다.

- 예시 정답  
(1) 가운데가 부풀어 오른 납작한 원반

- 2014년 고2 국어 2번 문항

[8, 서답형 2] 다음 글을 읽고 물음에 답하시오.

(가) [이·제] 나·라·히 仁義之道(인의지도)를 ·퍼 [쓰시·며] 詩書之教(시서지교)를 ·크·기 ① 너·기·시·느·니 :네·빅·화 :사·롭 도의·여 ·즈·라·나 급·데·하·야 벼·슬·하·야 나·라 :들·스·와 [닐·굴] :진·심·하·야 섬·기·수·오·며 ·부·못·의 [효·도·하·속·오·며] 가문·을 ·빚·내·요·미 :엇·더·홀·고·이 ·곧·내 몸·일·위·나 道(도)를 ·행·하·야 :후·세·에·일·홈 :배·퍼 내·여 ·부·모·를 나·토·와 내·요·미 :효·도·의 [민·초·미·나·라]

- <번역박통사(翻譯朴通事)>(1510년경) -

(나) [이제] 나라가 인의(仁義)의 도를 퍼 [쓰시며] 시서(詩書)의 가르침을 크게 여기시니, 네가 배워 사람이 되어 커서 급제하여 벼슬하여 나라를 도와 [일굴]을 충심으로 섬기며 부모께 [효도하며] 가문을 빚냄이 어떠하겠는가. 내 몸을 일으켜 도를 행하여 후세에 이름을 떨쳐 부모를 나타내는 것이 곧 효도의 마침이니라.

- 예시 정답  
할머니께서 동생에게 선물을 주셨다.

[서답형 2] ①에 나타난 높임 표현을 학습하기 위해 현대 국어의 문장을 만들고자 한다. <조건>에 맞게 적절한 문장을 쓰시오.

<조 건>

- 다음 말들 중 4개를 선택하여 한 문장으로 쓸 것.  
할머니께서, 할머니께, 동생에게, 동생이, 선물을, 주다, 드리다
- ①에 나타난 높임 표현과 같은 종류의 높임 표현으로 쓸 것.
- ‘-다.’로 끝나는 과거 종결형 문장으로 쓸 것.

\_\_\_\_\_.

- 2014년 고2 국어 6-(1)번 문항

[서답형 6] <자료 2>는 <자료 1>을 검토한 결과이다. <자료 2>를 반영하여 <자료 1>의 ㉠-㉣를 각각 고쳐 쓰시오.

<자료 1>

동양과 서양의 건축물에는 자연에 대한 사람들의 생각이 녹아 있다. 서양 사람들은 자연을 개발해야 할 대상이라고 생각하지만 우리는 자연의 도움을 받고 산다고 생각한다. 또 서양 사람들은 자연을 극복해야 한다고 생각하지만 ㉠ 우리는 순응해야 한다고 생각한다.

㉡ 이처럼 자연을 대하는 생각이 틀리기 때문에 건축물의 형태에도 차이가 생긴다. 서양의 건축물은 자연을 극복한다는 의식을 가지고 만들어진 반면에 우리의 전통적인 건축물은 자연과의 조화를 고려하여 지었다. 지붕만 보더라도 서양의 경우는 높고 웅장하며 하늘을 향해 솟아나는 듯한 직선의 미를 보여 주지만, 우리의 경우는 부드럽고 완만한 산등성이 같은 ( ㉢ )을/를 보여 준다.

<자료 2>

선생님: 먼저 어휘와 문장 수준에서 <자료 1>을 검토해 봅시다. ㉠에는 필요한 문장 성분이 빠져 있으니, 그 성분을 보충하여 해당 문장을 다시 쓰세요. 또 ㉡에는 적절하지 못한 단어가 하나 있으니, 그 단어를 찾아 고쳐 해당 문장을 다시 쓰세요.

이번에는 문단 수준에서 봅시다. 두 번째 문단에서는 서양의 건축물과 우리의 건축물이 내용상 대조를 이루고 있어요. 앞 내용을 고려하여 2어절의 표현으로 ㉢를 채워 보세요. 앞에 나온 어구와 대구가 되는 어구를 쓰면 대조의 느낌을 더 살릴 수 있을 거예요.

(1) ㉠: \_\_\_\_\_  
\_\_\_\_\_.

- 예시 정답  
(1) ㉠: 우리는 자연에 순응해야 한다고 생각한다.