

서답형 문항 온라인 채점 시스템의 개발과 적용

조지민[†] · 김경훈^{††}

요 약

서답형 문항이 포함된 대규모 평가에서 학생들의 답안이 어떻게 채점자들에게 분배되고 어떠한 체계적인 채점 과정을 거쳐야 하는 지에 대한 채점의 효율성은 전반적인 평가의 성공에 중요한 사안이다. 서답형 문항 채점에서 채점자 신뢰도를 확보하기 위해 채점자의 채점 결과가 얼마나 일관성이 있으며, 또한 채점자 간의 채점 결과 또한 서로 일관성을 유지해야하는 것이 중요하다. 이 연구의 목적은 지필 평가로 시행된 서답형 문항 채점을 위한 온라인 채점 시스템을 설계하고 개발한 후, 이를 시행한 결과를 기반으로 온라인 채점 시스템에 대한 신뢰도를 확인하고자 하였다. 이러한 서답형 문항 온라인 채점은 채점자들의 채점 과정에 대한 정보로, 기존의 전통적인 채점 방식과의 비교와 함께 채점자 내 및 채점자 간의 신뢰도를 제공하고 있다. 그 결과로 온라인 채점 시스템이 신뢰롭고 타당한 서답형 문항 채점 결과를 얻는데 효과적이었다는 것을 알 수 있었다.

주제어 : 온라인 채점, 채점자 신뢰도, 서답형 문항

Development and Application of an Online Scoring System for Constructed Response Items

Jimin Cho[†] · Kyunghoon Kim^{††}

ABSTRACT

In high-stakes tests for large groups, the efficiency with which students' responses are distributed to raters and how systematic scoring procedures are managed is important to the overall success of the testing program. In the scoring of constructed response items, it is important to understand whether the raters themselves are making consistent judgments on the responses, and whether these judgments are similar across all raters in order to establish measures of rater reliability. The purpose of this study was to design, develop and carry out a pilot test of an online scoring system for constructed response items administered in a paper-and-pencil test to large groups, and to verify the system's reliability. In this study, we show that this online system provided information on the scoring process of individual raters, including intra-rater and inter-rater consistency, compared to conventional scoring methods. We found this system to be especially effective for obtaining reliable and valid scores for constructed response items.

Keywords : Online Scoring, Rater Reliability, Constructed Response Items

† 정 회 원: 한국교육과정평가원 선임연구위원
 †† 중신회원: 한국교육과정평가원 선임연구위원 (교신저자)
 논문접수: 2013년 10월 16일, 심사완료: 2013년 12월 12일, 게재확정: 2014년 3월 15일

1. 서론

컴퓨터를 이용한 정보통신 기술의 비약적인 발전과 이에 관련된 인프라의 구축이 확충됨에 따라 다양한 분야에서 인터넷 이용하는 사례가 급증하였다. 특히 인터넷이나 웹은 정보 활용 교육의 일환으로 지속적으로 교수학습 과정에 활용되어 왔으며, 최근에 들어서는 인터넷이나 웹을 이용한 평가 시행과 결과 보고가 활성화되고 있는 추세이다[1][2].

종전과 같은 지필평가 방식으로 대규모 집단을 대상으로 실시하는 고부담 시험에 있어서는 서답형 문항의 답안지를 채점하는 과정에서 많은 문제점이 발생한다. 채점자 전원이 동시에 동일한 장소에 집결하여 피험자가 작성한 응답지를 일일이 채점하게 되므로, 대규모의 채점 공간 확보 및 비용이 증가하고, 많은 시간과 노력이 필요하다.

이에 대한 대안으로 인터넷이나 웹을 이용한 컴퓨터 기반 검사(CBT, WBT 등) 방식을 이용하기도 하지만 대규모의 집단을 대상으로 동시에 실시하는 경우에는 이를 뒷받침 할 수 있는 방대한 인프라 구축비용이 소요될 뿐만 아니라 검사를 실시하는 시간 동안 모든 시스템이 정상적으로 작동하는 것이 담보되어야 한다.

이러한 문제점을 극복하기 위해 지필 검사를 실시한 후, 지필 검사 응답지를 스캔하여 인터넷이나 웹 또는 LAN을 기반으로 채점 업무를 수행하는 채점 시스템의 개발과 활용이 점차 확대되고 있는 추세이다[3][4][5].

대규모 집단이 동시에 실시하는 검사에서 인터넷이나 웹을 이용하여 채점 업무를 수행하는 온라인 채점 방식은 앞서 언급한 바와 같이 모든 채점자가 특정 장소에 물리적으로 모일 필요가 없다는 장점을 갖는 반면, 얼마나 효율적으로 피험자의 응답지를 분배하고, 채점 과정과 결과의 질을 체계적으로 관리할 수 있는 시스템을 확보하는가가 중요한 문제점이라 할 수 있다.

서답형 문항을 채점하는 과정에서 채점자 신뢰도를 높이기 위하여 채점자 연수를 시행하거나, 답지 별 복수 채점이나 재 채점 실시 등의 타당한 평가 기법을 활용함으로써 채점의 공정성과 신뢰성을 확보해야하는 것이 매우 중요하다. 이를

위하여 채점 과정과 채점 결과에는 채점자의 개인적인 주관, 편견, 선입견, 종교, 지역, 도덕적·정치적 성향 등이 작용되지 않도록 해야 하는 것이다. 서답형 문항의 채점에 대한 채점자의 영향을 최소화하기 위한 방법의 하나로 복수 채점을 실시하게 되는데, 이때 두 채점자의 채점 결과에 대한 일관성을 유지하는 것이 매우 중요하다.

본 연구의 목적은 대규모 집단을 대상으로 실시하는 지필평가에서 적용 가능한 서답형 문항의 온라인 채점시스템을 설계·개발하고, 이의 시범 적용을 통하여 채점시스템의 효율성을 검증하는데 있다.

2. 이론적 배경

서답형 문항의 경우에 채점자 신뢰도를 높이기 위하여 채점자 연수를 시행하거나, 답지 별 복수 채점이나 재 채점 실시 등의 타당한 평가 기법을 활용함으로써 채점의 공정성과 신뢰성을 확보해야하는 것이 매우 중요하다. 이를 위하여 채점 과정과 결과에는 채점자의 개인적인 주관, 편견, 선입견, 종교, 지역, 도덕적·정치적 성향 등이 작용되지 않도록 해야 하는 것이다. 구체적으로 서답형 문항의 점수에 대한 채점자의 영향을 배제하기 위하여 복수 채점을 실시하며, 이때 복수 채점 결과의 일관성을 채점자 신뢰도로 표현된다.

채점자 신뢰도 추정 방법은 점수가 연속변수인가 비연속변수인가에 따라 구분한다. 비연속변수일 경우는 채점자 혹은 평정자간의 분류 일치도를 분석하는 일치도 통계(agreement statistics), 일치도 통계에서 우연에 의한 확률을 제거하고 평정자간의 일치도를 추정하는 Cohen의 Kappa 계수 등을 사용할 수 있다. 반면 연속변수로 채점하는 경우, 주로 사용되는 방법으로 상관계수를 산출하는 방법과 일반화가능도 이론 등을 적용한다. 각각의 신뢰도 추정 방법들은 각각의 특징과 더불어 장점과 단점을 가지고 있고 서답형 문항 특성 등에 따라 선택적으로 사용될 수 있다[6].

2.1 일치도 통계와 Kappa 계수

일치도 통계와 Kappa 계수는 채점자가 관찰대

상의 행위나 수행결과에 점수를 부여하기 보다는 어떤 유목이나 범주로 분류할 때, 채점자간의 분류 일치도를 추정하는 방법이다. 예로 A, B 채점자가 채점한 N명의 채점 결과를 j 개의 범주로 분류하여 평정한 결과는 <표 1>과 같다고 하자.

<표 1> 채점자(A, B)에 의한 채점 결과 예시

A \ B	1	2	3			J-1	J
1	N_{11}						
2	N_{21}	N_{22}					
3	N_{31}		N_{33}				
J-1							
J							N_{jj}

두 채점자에 의하여 일치되게 평정된 응시자 수는 J×J 분할표에서 대각선에 위치한 N_{11} , N_{22} , N_{33} , N_{44} , ..., N_{jj} 로서 두 채점자간의 일치도 통계 (agreement statistics)는 <공식 (1)>에 의한다[7].

$$P_A = \frac{N_{11} + N_{22} + \dots + N_{jj}}{N} \quad \text{공식 (1)}$$

일치도 통계는 두 채점자가 일치하게 평정하는 대각선 부분에 우연에 의하여 채점된 응시자가 포함되어 있어 두 채점자간의 일치도가 과대 추정되는 문제점 해결하게 위하여 Cohen(1960)은 우연에 의한 확률을 제거한 Kappa 계수를 제안하였다[8]. 우연에 의하여 J × J 분할표의 대각선에 있는 칸에 분류되는 사례 수는 <공식 (2)>에 의하여 계산된다.

$$N_{jc} = \frac{N_j \times N_j}{N} \quad \text{공식 (2)}$$

전체 N명의 응시자 중 우연히 두 채점자에 의하여 일치된 채점을 받은 응시자의 확률은 <공식

(3)>에 의한다.

$$P_c = \frac{N_c}{N} \quad \text{공식 (3)}$$

이때 N_c 는 J×J 분할표에서 두 채점자에 의하여 동일하게 분류되는 사례수를 더한 값을 말한다.

채점자 간에 우연하게 일치된 채점결과를 제거하고 순수하게 채점이 일치한 두 채점자 간의 일치도를 계산하기 위하여 Cohen(1960)이 제안한 Kappa 계수 계산공식은 <공식 (4)>와 같다.

$$K = \frac{P_A - P_c}{1 - P_c} \quad \text{공식(4)}$$

K계수는 우연에 의하여 두 채점자의 평정결과가 일치하는 확률을 제거하여서 일치도 통계보다 항상 낮은 수치로 산출되기 때문에 채점자간 신뢰도 추정에서 일치도 통계와 수치가 동일한 경우 채점 자료에 대한 신뢰성을 더 인정받을 수 있다. 하지만 두 채점자가 일치하게 평정하는 대각선 부분과 더불어 대각선에 인접한 cell을 오차 허용 범위로 허락하여 두 채점자에 의하여 일치된 채점을 받은 응시자의 확률을 계산하는 것으로 채점의 pass/fail rate를 보고하였다.

2.2 상관 계수

상관계수로 채점자 간 신뢰도를 산출하는 경우 서로 다른 채점자가 동일한 수험생의 서답형 문항 답안에 얼마나 유사하게 점수를 부여하였나를 분석하기 위해 Pearson 단순적률상관계수 <공식 (5)>을 산출한다[9].

$$r_{xy} = \frac{S_{xy}}{S_x S_y} \quad \text{공식 (5)}$$

S_{xy} 는 채점자 두 사람의 채점 점수 공분산

S_x , S_y 는 각 채점자의 채점 점수의 표준편차

상관계수가 수치 1에 가까울수록 각기 다른 두 채점자가 동일한 채점 결과를 보이는 것으로 해석할 수 있으며, 이와는 반대로 상관계수가 0에 가까울수록 채점자간 상이한 결과를 보이는 것으로 해석할 수 있다. 결과적으로 채점자간에 문항 배점의 일정 기준치 이상이 차이가 나게 되면 그

문항은 재검토되어 수정해야 하고 검사의 규모나 목적에 따라 일정 기준치에 대한 의사결정이 있어야 한다.

단 상관계수의 경우 두 채점자 점수간의 함께 변화하는 정도만을 산출하기에 두 채점자의 점수간 차이를 반영하지 않는 한계를 가지고 있다. 예를 들어 한 문항의 채점 점수가 1점씩 증분 되는 부분점수 10개를 갖는 10점 만점의 문항의 경우, 5개 각각의 답안에 한 채점자가 1에서 5까지의 연속 점수를 주고 다른 채점자는 6에서 10까지의 연속 점수를 줄 경우에 두 점수 간의 상관계수는 1로 동일한 채점 결과를 보이는 것으로 잘못 해석할 수 있는 경우가 생기게 된다. 이때 채점자들에 의해 부여된 점수 간의 차이를 반드시 함께 고려해야 한다. 이를 위하여 동일한 답지에 대한 채점자들 점수 간의 차이를 계산하고 이에 대한 평균을 산출하여 비교 분석하여야 한다. 이때 평균은 차이점수의 절댓값을 계산하여 산출하지 않으면 차이점수가 갖는 +와 -값이 서로 상쇄되어 정확한 차이에 대한 평균을 계산하기 어렵다. 하지만 표준편차와 평균의 오차를 계산하여 그 차이 점수들에 대한 분포를 함께 비교하여 설명하는 것이 필요하다.

3. 온라인 채점 시스템의 개요

검사의 일반적인 절차는 검사 도구(문항) 제작 → 검사지 배부 → 검사 실시 → 검사 문답지 회수 → 채점 → 결과 분석 및 활용 등으로 이루어진다. 본 연구에서 설계하는 온라인 채점 시스템은 종래의 지필 검사와 동일한 절차와 방법으로 시행되는 검사에서 서답형 문항의 채점을 온라인상에서 수행하기 위한 시스템이다. 채점과정을 온라인으로 수행하기 위해서는 답안지 이미지 스캐닝, 이미지 분할, 답안지 배분, 온라인 채점 수행, 채점 결과 검증과 같은 일련의 과정이 필요하다. 본 연구에서는 서답형 문항의 온라인 채점 과정에서 답안지의 배분, 채점 수행, 채점 결과 검증, 채점자 신뢰도 점검에 주안점을 두었다.

3.1 시스템의 목표

조건 1. 지필 검사(paper and pencil test) 방식으로 검사가 실시된다는 것을 전제로 하되 서답형 문항의 채점은 채점자가 온라인(인터넷, LAN, WEB)상에서 채점을 수행 할 수 있어야 한다.

조건 2. 모든 답안은 서로 다른 2명의 채점자에 의해 채점(all responses double scored)하며, 모든 채점자 쌍에게 배당하는 답안은 고르게 분배되 채점자간의 신뢰도를 추정할 수 있어야 한다.

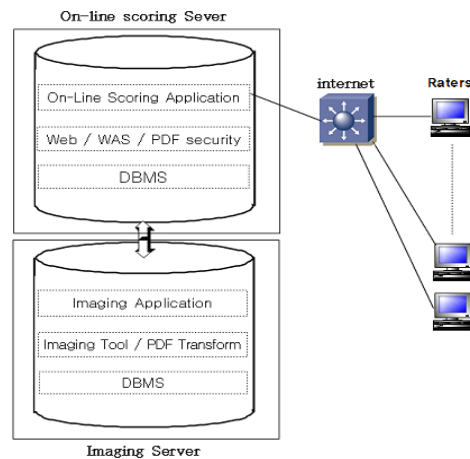
조건 3. 모든 문항에 대하여 채점자간 채점 신뢰도(상관계수, 일치도, 차이점수)는 일정 수준이상으로 조정할 수 있는 기능이 있어야 한다.

조건 4. 채점자 간 점수가 일정 범위를 벗어난 답안은 재 채점이 이루어져야 한다. 재 채점한 두 점수가 또 다시 일정 범위를 벗어나면 관리자에 의해 채점이 이루어진다.

조건 5. 재 채점 후에도 채점 신뢰도에 문제가 있는 답안은 Manager 또는 Supervisor에 의해 최종 채점이 이루어져야 한다.

3.2 시스템의 개요

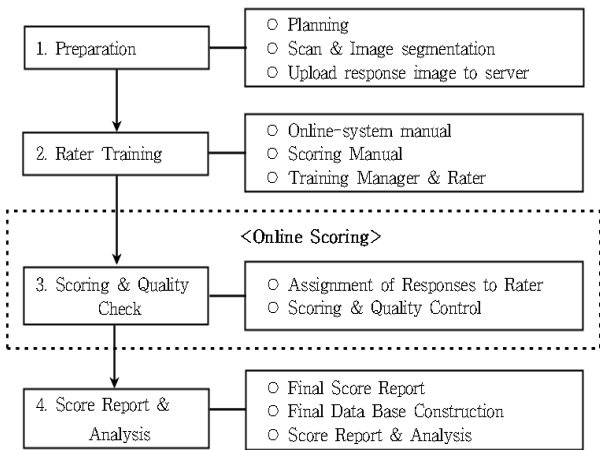
하드웨어 구성은 크게 답안지를 스캔하여 이미지를 변환하고 분할하는 이미지 처리서버, 답안지 이미지(PDF)와 채점결과를 저장, 관리, 채점자에 대한 작업을 관리하는 온라인 채점서버 등으로 구성하였다([그림 1]).



[그림 1] 온라인 채점 시스템 개요

3.3 시스템의 실행 절차

서답형 문항의 답안지를 이미지로 스캐닝한 후 각 문항별로 답안 이미지를 분할하여 온라인 채점 시스템에서 활용한다. 온라인 채점과정에서 개인 정보는 답안 내역 정보와 분리하여 관리하고 각 답안은 별도의 관리번호를 부여하여 개인 정보의 유출을 방지하도록 하였다. 서답형 답안지의 스캐닝에서 온라인 채점 완료시까지의 주요 단계는 [그림 2]와 같이 설정하였다.

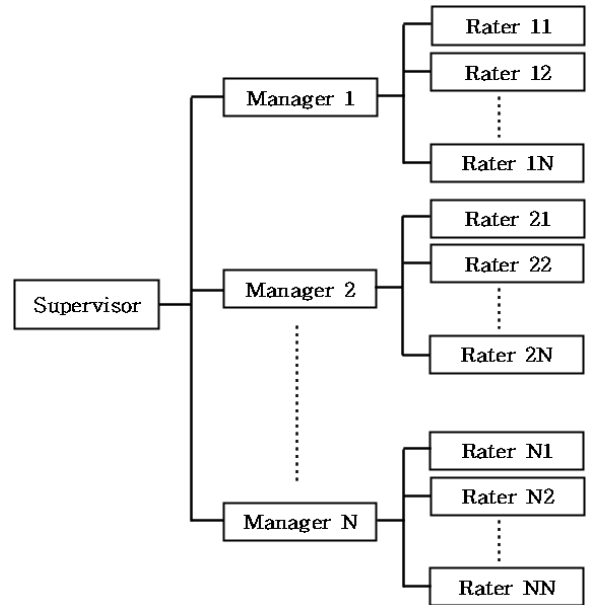


[그림 2] 온라인 채점의 실행 절차

3.3.1 채점 준비

채점 준비 단계에서 채점자 자격과 채점자 수는 검사의 특성, 채점의 난이도, 응시자 수, 채점 방식, 채점 기간 등을 종합적으로 고려하여 선정해야 한다. 본 연구에서 온라인 채점을 위한 조직은 [그림 3]과 같이 구성하였다.

서답형 답안지를 답안 이미지 전체를 스캐닝한 후, 문항별로 이미지를 분할한다. 응시자가 답을 작성하는 구역 범위를 벗어난 위치에 답안을 작성한 경우에는 전체 답안 이미지를 참고할 필요가 있기 때문에 전체 이미지와 분할된 이미지는 구분하여 저장할 필요가 있다. 문항별로 분할된 이미지는 온라인 채점을 위해 온라인 채점 서버로 이동하여 저장하였다.



[그림 3] 온라인 채점의 채점 조직도

3.3.2 채점자 선정과 연수

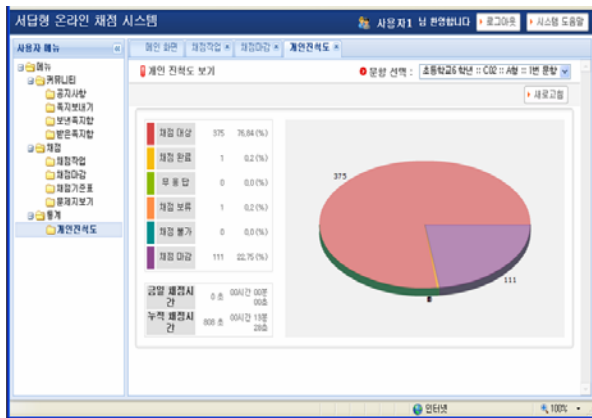
검사의 목적과 결과 활용 등에 적절한 채점자의 선정과 구성은 매우 중요하다. 본 연구에서는 응시 학생들을 직접 가르치는 학교 현장 교사를 중심으로 채점자를 선정하였다. 채점 결과의 공정성과 신뢰성을 높이는데 직접적인 효과를 기대할 수 있는 중요한 과정의 하나는 채점자 연수이다. 채점자 연수의 목적은 채점자들 간의 채점 기준에 대한 공통적인 이해와 공유된 해석의 기준들을 공유하도록 하기 위함이다. 이는 궁극적으로 신뢰도와 타당도가 높은 채점 결과를 확보할 수 있게 한다.

3.3.3 채점과 질 관리

서답형 문항의 온라인 채점을 수행하기 위해서는 채점자 별 문항 배당 전략, 답안 파일 제공을 위한 전략과 이를 뒷받침할 수 있는 시스템이 준비되어야 한다.

한 채점자의 결과가 다른 채점자와 다르거나 일관성이 없는 채점을 한다면 그 평가는 평가의 질을 갖지 못하기 때문에 채점자간 차이 점수를 검토하여 각 문항별로 채점자간 문항 배점의 일

정 기준치 이상의 차이가 나게 되면 그 문항은 재채점 되어야 한다. 검사의 규모나 목적에 따라 일정 기준치에 대한 의사결정이 있어야 한다. 채점자는 자신의 채점 진행 상황을 수시로 체크할 수 있고, 채점 관리자는 채점자들의 채점 진행 상황을 수시로 모니터링 할 수 있도록 하였다([그림 4],[그림 5]).



[그림 4] 개별 채점자 채점 진척도 화면



[그림 5] 채점자별 채점 진척도 모니터링 화면(관리자용)

3.3.4 응답지 배정

응답지의 배정은 모든 채점자에게 고르게 배정 하되, 모든 답안은 서로 다른 채점자에 의해 중복 채점이 이루어진다. 또한 모든 채점자는 여러 개의 서답형 문항 중에서 1개의 문항 답란만을 채점한다. 특정 문항의 모든 답안을 채점자들에게 배분하는 경우, 랜덤 추출 배정 방식의 효과를 유

지하면서 데이터베이스 설계 및 활용이 편리해야 한다. 본 연구에서는 답안 ID를 기준으로 채점자를 배정하였으며, [그림 6]의 Phase 1부터 Phase 4까지의 점수를 데이터베이스에 저장함으로써 채점 과정 또는 채점 종료 후에 채점자간의 신뢰도 추정이 가능하도록 배분하였다. 국가 수준의 평가에서 특정 지역, 특정학교 학생 답안지가 연속적으로 특정 채점자에게 집중하여 배정되는 것을 시스템적으로 방지하였다.

일반적으로 복수 채점은 서답형 채점 결과의 신뢰성을 확보하기 위해서는 매우 중요한 의미를 갖는다. 종전 채점방식에서의 복수 채점은 답지 분배 및 결과 정리의 어려움이 많으나 온라인 채점방식에는 답안지 분배와 채점 결과 처리가 온라인 전송을 통해 이루어지기 때문에 복수 채점을 편리하게 수행할 수 있다. 또한 한 채점자의 답지를 다른 한사람의 채점자와 겹치는 수준을 넘어 모든 채점자와 겹치게 함으로써 채점자간 신뢰도뿐만 아니라 채점자내 신뢰도를 평가할 수 있었다.

본 연구에서 적용한 답안 배분 방식은 <표 2>와 같다(예, 10명의 채점자, 45개 응답지를 배정하는 경우).

<표 2> 채점자별 응답지 배분 방식

	U01	U02	U03	U04	U05	U06	U07	U08	U09	U10
U01		1	10	18	25	31	36	40	43	45
U02			2	11	19	26	32	37	41	44
U03				3	12	20	27	33	38	42
U04					4	13	21	28	34	39
U05						5	14	22	29	35
U06							6	15	23	30
U07								7	16	24
U08									8	17
U09										9
U10										

n 번째 응시자의 답안, U(m) : 채점자 m

<표 2>에서 11번째 학생의 답안은 U02와 U04 채점자가, 33번째 학생의 답안은 U03과 U08 채점자에 의해 각각 채점이 이루어진다. 여기서 채점자 신뢰도는 채점자간 신뢰도와 채점자내 신뢰도

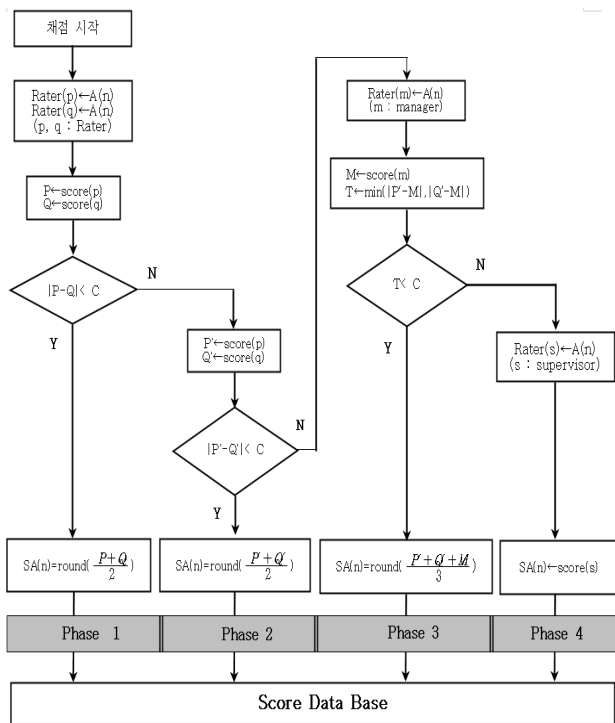
로 평가될 수 있다. 먼저 채점자간 신뢰도는 45개 복수 채점 결과(차이 점수 혹은 상관)를 평가함으로써 특정 문항에 대한 전체 10명 채점자의 신뢰도를 파악할 수 있다. 둘째, 각 채점자는 9개 복수 채점 결과를 산출하여 비교함으로써 채점자의 상대적 비교가 이루어질 수 있다.

모든 답안에 대하여 채점자 배정이 완료될 때까지 <표 2>와 같은 순서를 반복하여 배정하였다.

3.3.5 점수 산출 과정

본 연구에서는 모든 학생의 모든 응답지는 서로 다른 두 명의 채점자에 의한 복수 채점을 전제로 한다. 특정 답안에 대하여 서로 다른 두 채점자에 의해 부여된 두 점수가 주어진 범위를 벗어나게 되면, 재 채점을 하게 된다. 재 채점 후에도 두 채점자간의 점수 차이가 일정 범위를 벗어나면 채점 관리자가 채점하게 되며, 채점 관리자가 개입한 후에도 일정 범위를 벗어나는 답안은 책임자(supervisor)에 의해 최종 채점이 이루어지도록 하였다.

본 시스템에서 점수 산출 과정은 [그림 6]과 같다.



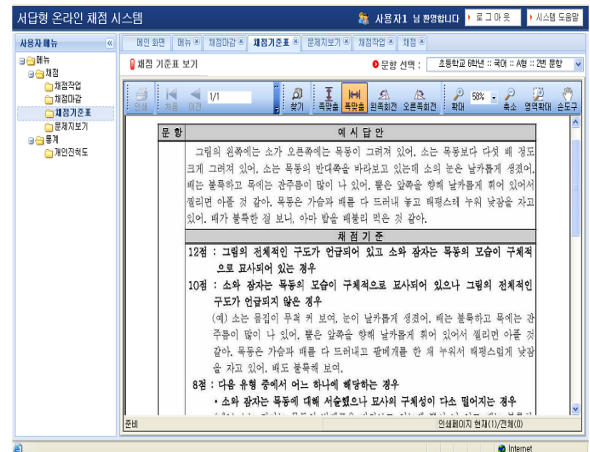
[그림 6] 채점 시스템의 점수 산출과정

1) Phase 1 :

채점 전 채점자 연수를 통해 서답형 문항의 내용과 채점기준 등에 대해 채점자간의 동일한 이해를 도출하고, 채점 시 채점자 개인의 주관적 판단의 개입을 최소화하는 훈련을 거친다. 이러한 훈련을 받았다 하더라도 채점 결과에 대한 공정성과 신뢰성을 최대화하기 위한 기본 방향과 원칙은 모든 답지는 <표 2>의 응답지 배정 방식으로 배정된 2명의 채점자가 온라인상에서 답지를 채점하도록 한다.

이때 책임자가 채점 전에 결정한 기준치로서의 최소 차이점수(이하 기준치 C값으로 명명)와 두 채점자가 결정한 점수간의 차이점수를 비교하도록 하여 차이점수가 기준치 C값보다 작은 경우에는 두 채점자가 결정한 점수의 평균을 반올림한 점수를 최종 점수로 확정한다. 반면 기준치 C값보다 크거나 같은 경우에는 그 해당 답안은 점수를 부여하였던 두 채점자에게 동일하게 다시 분배되어 재 채점을 진행하게 된다.

점수 산출 과정에서 가장 중요한 의사결정이 두 채점자들 간의 차이점수에 대한 기준치 값을 어떻게 결정할 지는 검사의 방향과 성격 문항의 난이도 등 여러 가지를 고려하여 결정할 수 있다.

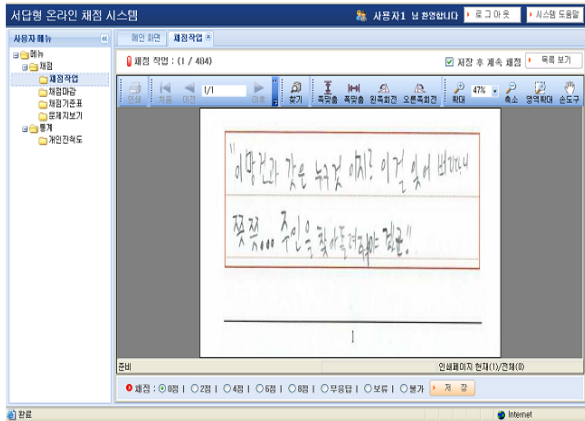


[그림 7] 문항 채점 기준표 화면(채점자용)

2) Phase 2 :

두 채점자의 점수 차이가 기준치 이상이 되어 재 분배받은 답안지를 다시 채점 하는 절차로써 재채점 점수간의 차이가 기준치 C값보다 작으면

두 채점자의 재채점 점수를 평균하여 반올림한 점수를 최종 점수로 확정하는 단계이다. 반면 기준치 C값보다 크거나 같은 경우에 그 답지는 채점자들을 관리하는 채점관리자가 채점을 시행하게 되는 Phase 3을 진행하게 된다.



[그림 8] 채점 화면

3) Phase 3 :

채점 관리자의 채점 결과와 재 채점으로 이루어진 두 채점 결과와 비교를 하고 그 차이가 기준치 C값보다 작을 경우에는 총체적으로 두 명의 재채점 결과와 채점관리자 점수 모두에 동일한 가중치를 부여하기 위하여 세 명 점수의 평균을 반올림한 값을 최종 점수로 확정한다. 반면 그 차이가 기준치 C값보다 클 경우, 최종적으로 책임자가 채점하는 Phase 4를 수행하게 된다.

4) Phase 4 :

채점자와 채점관리자 사이의 차이점수가 기준치 이상이 되면 그 답안은 책임자에게 보내지고 책임자의 점수가 그 이전의 채점자 및 채점관리자 점수와 상관없이 최종 점수로 확정되게 된다.

4. 온라인 채점 시스템의 모의실험

이 장에서는 앞에서 개발한 온라인 채점 시스템을 이용하여 서답형 문항을 실제로 채점하는 모의실험을 통하여 온라인 채점 시스템의 효율성을 검증하였다.

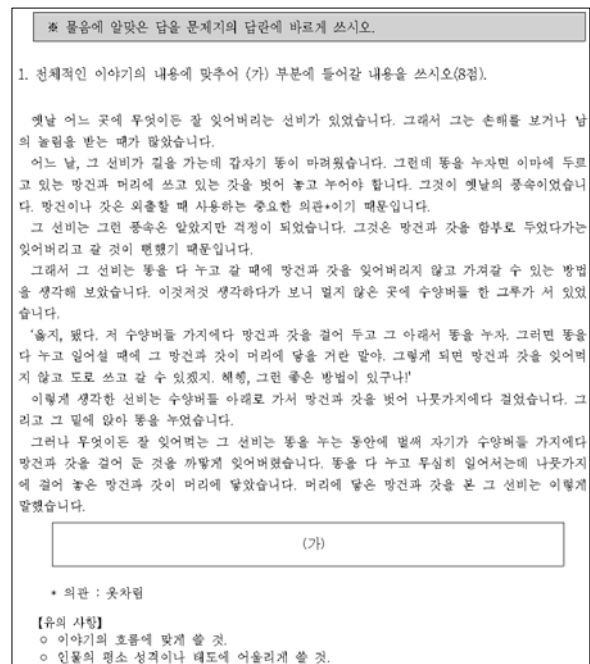
4.1 실험 설계

모의실험의 기본 원칙과 방향으로는 모든 피험자의 개별 답안 모두 2명의 채점자가 채점하도록 하는 복수채점 설계를 하였다. 또한 채점자들의 채점의 분량과 채점의 효율성 등을 고려하여 문항, 채점자, 피험자가 부분적으로 연계되도록 하는 방식으로 설계되어 답지 분배가 이루어졌다.

온라인 채점 시스템의 효율성은 채점 소요시간, 복수 채점에서 점수 일치도, 채점자 신뢰도를 근거로 이루어졌다.

실험에 참가한 피험자는 서울에 위치하고 있는 10개 초등학교의 6학년 학생 총 2443명으로 대상으로 검사를 시행하였다. 검사시간은 문항을 제작한 교과 전문가와의 협의를 거쳐 30분으로 결정하였다.

본 실험에서 사용한 검사지는 국어 과목 2개의 문항으로 구성하였다. 첫 번째 문항은 전체적인 이야기를 읽고 이야기 속에 나오는 인물의 성격이나 태도를 파악하여 주어진 이야기의 다음에 이어질 이야기를 흐름에 맞게 작성하는 8점 문항이다([그림 9]).




[그림 9] 국어 1번 검사 문항

두 번째 문항은 주어진 그림을 보고, 그림 속에

등장하는 소와 목동의 모습이 생생하게 드러나도록 그림의 내용을 친구에게 소개하는 전자 메일을 작성하는 12점 문항이다([그림 10]).

2. 철수는 그림에서 (가) 그림을 보고, 이 그림을 친구 영호에게 소개하기 위하여 전자 메일을 (나)와 같이 작성하였습니다. 그림 속에 등장하는 소와 목동의 모습이 생생하게 드러나도록 알맞은 내용을 쓰십시오(12점).

(가) <낫갈 자는 목동>



(나)

영호에게
오늘, 그림책을 뒤적거리다가 '낫갈 자는 목동'이라는 그림을 보게 되었어. 그 그림은 조선 시대에 김두팔이라는 화가가 그린 것이야. 그림에는 소와 목동이 그려져 있어. 어떻게만 이야기하면 어떤 그림인지 상상이 잘 안 되지? 내가 한 번 생생하게 그려볼게.

어때, 네 눈앞에 낫갈 자는 목동과 소의 모습이 그려지니? 다음에 널 만날 때, 내가 그림책을 가져갈게. 그때 네 상상과 그림 속의 풍경을 비교해 보렴. 다음에 만나자. 안녕.
철수가

[그림 10] 국어 2번 검사 문항

채점 기준은 2문항 모두 2점 간격으로 정하였으며 예시 답안 및 채점 기준표는 <표 3>, <표 4>와 같다.

<표 3> 문항 채점 기준표(1번 문항)

문항	예시 답안
	<p>“어떤 사람이 여기다 망건과 갓을 걸어 놓고 갔네! 그 사람 참 정신 나간 사람이군!” “아니, 누가 이곳에다 망건과 갓을 걸어 놓고 갔을까? 내가 망건과 갓이 없으니 이것을 쓰고 가면 되잖구나.”</p>
	<p>채점 기준</p> <p>조건①: 이야기의 흐름에 맞는 내용일 것 조건②: 인물의 성격이나 태도와 어울리는 내용일 것</p>
1번 (8점)	<p>8점 : 조건①, ②를 모두 충족한 경우 6점 : • 조건①을 충족하였으나 조건②가 미흡한 경우 (예) 이렇게 머리에 당게 걸어두니 걱정할 필요가 없군. • 조건①이 비흡하나 조건②를 충족한 경우 (예) 그런데 내 갓과 망건은 어디에 두었을까? 4점 : • 조건①을 충족하였지만 조건②를 충족하지 못한 경우 (예) 내가 여기 갓과 망건을 벗어 두었었지. • 조건①을 충족하지 못했지만 조건②를 충족한 경우 (예1) 그런데 여기가 어디지? (예2) 내가 어디에 가는 길이었더라? 2점 : • 조건①, ②와 관련이 있으나 매우 미흡하게 쓴 경우 (예1) 이게 뭐지? (예2) 아이고, 짬짜이야! 0점 : 조건을 모두 충족하지 못했거나, 아무 것도 쓰지 않은 경우</p>

<표 4> 문항 채점 기준표(2번 문항)

문항	예시 답안
	<p>그림의 왼쪽에는 소가 오른쪽에는 목동이 그려져 있어. 소는 목동보다 다섯 배 정도 크게 그려져 있어. 소는 목동의 반대쪽을 바라보고 있는데 소의 눈은 날카롭게 생겼어. 배는 볼록하고 목에는 잔주름이 많이 나 있어. 뿔은 앞쪽을 향해 날카롭게 휘어져 있어서 찰리면 아플 것 같아. 목동은 가슴과 배를 다 드러내 놓고 태평스레 누워 낫갈을 자고 있어. 배가 볼록한 걸 보니, 아마 밥을 배불리 먹은 것 같아.</p>
	<p>채점 기준</p> <p>12점 : 그림의 전체적인 구도가 언급되어 있고 소와 잡자는 목동의 모습이 구체적으로 묘사되어 있는 경우 10점 : 소와 잡자는 목동의 모습이 구체적으로 묘사되어 있으나 그림의 전체적인 구도가 언급되지 않은 경우 (예) 소는 몸집이 무척 커 보여. 눈이 날카롭게 생겼어. 배는 볼록하고 목에는 잔주름이 많이 나 있어. 뿔은 앞쪽을 향해 날카롭게 휘어져 있어서 찰리면 아플 것 같아. 목동은 가슴과 배를 다 드러내 놓고 팔베개를 한 채 누워서 태평스레 낫갈을 자고 있어. 배도 볼록해 보여. 8점 : 다음 유형 중에서 어느 하나에 해당하는 경우 • 소와 잡자는 목동에 대해 서술했으나 묘사의 구체성이 다소 떨어지는 경우 (예) 목동은 태평스레 낫갈을 자고 있어. 소는 배가 볼록하고 몸집이 아주 커. • 소와 잡자는 목동에 대해 서술했으나 한 대상만을 구체적으로 서술한 경우 (예) 목동은 가슴과 배를 다 드러내 놓고 태평스레 누워 낫갈을 자고 있어. 팔베개를 하고 있는 것 같아. 배가 볼록한 걸 보니, 아마 점심을 배불리 먹은 것 같아. 6점 : 소와 잡자는 목동에 대해 서술했으나 묘사가 구체적이지 못한 경우 (예) 소는 뿔이 날카롭게 생겼고 목동을 등지고 있고. 목동은 배를 드러내고 낫갈을 자고 있어. 4점 : 다음 유형 중에서 어느 하나에 해당하지 않는 경우 • 소에 대해서만 서술했고, 묘사가 구체적이지 못한 경우 (예) 소는 몸집이 크고 뿔이 앞으로 휘어져 나와 있어. • 잡자는 목동에 대해서만 서술했고, 묘사가 구체적이지 못한 경우 (예) 목동이 소 옆에 누워 태평하게 잠을 자고 있어. 2점 : 소나 목동에 대해 서술했으나 그림의 내용과 관련성이 낮은 경우 (예) 소가 풀밭에서 휴식을 취하고 있어. 또는 목동이 소를 지키고 있어. 0점 : 다음 유형 중에서 어느 하나에 해당하지 않는 경우 • 문제와 무관한 서술을 하였거나 그림을 그린 경우, 아무런 서술도 하지 않은 경우</p>
2번 (12점)	

채점자는 초등학교에 재직 중인 10명의 교사로 구성하였다. 채점자를 결정한 후 실제 온라인 채점이 진행되기 전에 채점자 연수 자료를 준비하여 1일간 채점자 연수를 실시하여 채점 기준에 대한 동일한 해석을 할 수 있는 이해를 도왔고 또한 채점 시스템에 대한 설명으로 원활한 채점 진행을 위한 준비를 하였다.

온라인 채점은 총 4일에 걸쳐서 진행이 되었다. 모든 채점자에게 ID를 부여하고 동일한 날에 온라인 채점 시스템에 접속할 수 있게 하였고 4일간의 진행은 각 채점자들의 재량에 맡기어 진행을 하였다.

본 실험의 점수 산출 과정에서 두 채점자들 간의 차이점수에 대한 기준치 값은 1번 문항의 경우 배점 8점의 50% 즉 4점을 기준값으로 설정을 하였고 2번 문항의 경우는 배점 12점의 33%인 4점을 기준값으로 설정을 하여 진행하였다.

4.2 모의실험 결과

4.2.1 채점 소요시간

<표 5>에 나타난 바와 같이 1차 채점에서 1번 문항과 2번 문항의 답지는 10명의 채점자에게 개별적으로 487~490개의 답지를 배포하였으며, 개별 평균 채점시간은 1번 답지는 약 3시간, 2번 답지는 약 5시간 정도로 나타났다.

<표 5> Phase 1 수행에 걸린 시간

1번 문항 1차 채점 소요시간(phase 1)					
Rater	채점 답안수	총 채점시간 (sec)	문항당 평균시간 (sec)	최소시간 (sec)	최대시간 (sec)
U01	489	7895	15.74	5	390
U02	489	15399	31.49	5	2528
U03	490	9700	19.80	4	997
U04	490	7689	15.69	6	211
U05	489	9511	19.45	6	269
U06	489	24894	50.91	7	4223
U07	488	8332	17.07	4	342
U08	408	16959	34.75	5	2795
U09	488	11947	24.48	4	1840
U10	487	10059	20.66	6	157

2번 문항 1차 채점 소요시간(phase 1)					
Rater	채점 답안수	총 채점시간 (sec)	문항당 평균시간 (sec)	최소시간 (sec)	최대시간 (sec)
U01	488	26560	54.43	4	5263
U02	489	9701	19.84	4	647
U03	490	17913	36.56	5	857
U04	490	15603	31.84	5	878
U05	489	20821	42.58	5	2681
U06	488	45331	92.89	8	8427
U07	488	10462	21.44	6	327
U08	488	8759	17.95	5	339
U09	488	12756	26.14	4	1124
U10	487	14006	28.76	6	286

본 연구에서 수행한 것처럼, 2443명의 답지를 10명의 채점자가 이중채점 방식으로 채점하는 경우, 1번 문항 채점에 걸린 시간은 채점시간이 가장 많이 걸린 채점자를 기준으로 했을 때, 약 7.7시간 (phase 1: 24894 sec + phase 2: 1287 sec + phase 3: 1530 sec = total : 27711 sec), 2번 문항은 24.7시간(phase 1: 45331 sec + phase 2: 38816 sec + phase 3: 4923 sec = total : 89070 sec) 걸렸다. 2번 답안의 채점에 시간이 많이 걸린 이유는 1번 문항은 답의 길이가 짧고 점수의 급간이 5개인데 반해, 2번 문항은 답안의 길이가 길고, 점수의 급간이 7개이기 때문에 채점 시간이 많이 걸리는 것으로 나타났다(<표 5>,<표 6>,<표 7>).

재 채점 기준에 근거하여 2차 채점이 필요한 개별 채점자 당 답지 수는 평균 28개 정도이며 이는 채점자의 채점 답지의 약 6%에 해당하는 것이다. 즉, 약 94%의 채점 결과는 주어진 허용 기준값의 범위를 만족하였다. 주목할 점은 채점자 U07은 재 채점 문항수가 63개로 매우 많은 숫자를 보인 것을 알 수 있었고, 이러한 절차를 통하

여 채점자별 채점신뢰도를 점검하였다.

<표 6> Phase 2 수행에 걸린 시간

1번 문항 2차 채점 소요시간(phase 2)					
Rater	채점 답안수	총 채점시간 (sec)	문항당 평균시간 (sec)	최소시간 (sec)	최대시간 (sec)
U01	29	319	11.00	6	34
U02	29	777	26.79	6	251
U03	34	1287	37.85	10	164
U04	17	484	28.47	10	83
U05	22	1154	52.45	14	179
U06	25	1039	41.56	5	358
U07	63	798	12.67	7	46
U08	31	466	15.03	8	38
U09	20	564	28.20	11	76
U10	16	1176	73.50	17	225

2번 문항 2차 채점 소요시간(phase 2)					
Rater	채점 답안수	총 채점시간 (sec)	문항당 평균시간 (sec)	최소시간 (sec)	최대시간 (sec)
U01	92	5447	59.21	7	1497
U02	76	1819	23.93	7	100
U03	70	10768	153.83	7	6541
U04	100	2664	26.64	6	94
U05	71	7517	105.87	10	1552
U06	100	38816	388.16	10	9524
U07	78	1301	16.68	4	115
U08	73	1296	17.75	9	55
U09	77	3227	41.91	6	439
U10	55	2986	54.29	13	254

마지막으로 채점관리자가 관여하는 3차 채점과정에 이른 문항은 70개(3%), 112개(5%)로 전체 응시자 규모를 고려할 때 매우 작은 숫자로 1, 2차 채점과정을 통하여 전반적으로 채점결과가 안정적으로 산출되었음을 나타낸다.

<표 7> Phase 3(manager) 수행에 걸린 시간

번호	채점문항(3차)	총 시간 (sec)	평균시간 (sec)	최소시간 (sec)	최대시간 (sec)
1	70	1530	21.86	5	182
2	112	4923	43.96	5	1891

4.2.2 채점자간 점수 분포와 일치도

서답형 문항 채점의 신뢰도 분석은 국어 검사지 2개 각각의 문항 당 모든 채점자들의 공통분량인 약 54~55개 정도의 피험자 답지에 대해 수행되었고 채점자간 및 채점자내 신뢰성을 중심으로 분석하였다.

<표 8>은 10명의 채점자들이 각 문항에 부여한 채점점수 분포를 보여주고 있다. 대각선에 있는 셀 부분의 수치(굵은 글씨)는 채점자들이 누구이건 간에 동일한 답지에 대해 동일한 결과를 보

인 경우의 수치이고 대각선에서 멀리 있는 셀(음영 부분)일수록 채점자간에 서로 다른 판단을 하고 있음을 보여주고 있다. 극단적으로 <표 8>의 경우 한 채점자는 동일한 답지에 대해 0점을 주고 다른 채점자는 문항 최고 만점인 8점을 주고 있는데 그런 경우가 총 2443개 답안에서 6개가 나왔는데, 이에 대한 첫 번째 이유로는 채점자가 점수 표기를 잘못된 경우로 생각해 볼 수 있다.

<표 8> Phase 1 종료 후 채점자간 점수 분포

		score						
		0	2	4	6	8	Total	
1번 문항	score	0	65	18	9	2	3	97
		2	17	34	27	19	9	106
		4	1	14	86	107	24	232
		6	2	9	100	281	177	569
		8	3	7	55	234	1140	1439
	Total	88	82	277	643	1353	2443	

		score								
		0	2	4	6	8	10	12	Total	
2번 문항	score	0	60	28	2	3	1	1	0	95
		2	22	140	81	101	11	3	0	358
		4	2	43	77	116	46	6	0	290
		6	1	47	108	372	168	54	9	759
		8	2	9	24	158	253	115	25	586
		10	2	0	0	42	103	134	15	296
	12	0	0	0	2	9	25	23	59	
	Total	89	267	292	794	591	338	72	2443	

온라인 채점 시스템은 이러한 채점 결과를 보관·관리하고 모니터링하면서 잘못된 판단을 하거나 실수를 한 채점자들에게 재 채점의 기회를 부여하여 채점의 신뢰성을 확보할 수 있게 되는 것이다.

본 연구에서 적용한 온라인 채점 시스템에서는 동일한 답안에 두 채점자가 일치하게 평정하는 대각선 부분의 점수와 더불어 대각선에 인접한 cell을 오차 허용 범위로 설계하였다. 따라서 채점자들 간의 신뢰도를 알아보기 위하여 두 채점자에 의하여 일치 또는 유사한 채점을 받은 응시자의 수를 전체 채점 답안의 수로 나누어 이를 채점의 pass/fail rate를 산출하였다(<표 9>).

1번 문항의 채점 결과 94.15% (Kappa=.90)의 답안이 phase 1에서 pass되고, 2번 문항의 경우는 phase 1에서 83.55% (Kappa=.79)의 답안이 pass되었다. 1번 문항과 2번 문항의 모든 채점이 Phase 3에서 완료되었다.

채점이 이루어지는 횟수는 Phase1부터 Phase4까지 채점되는 모든 답안의 수를 합한 것으로 1번 답안의 경우 2656(2443+143+70)번의 채점이 수

행되고, 2번 답안의 경우는 2957번의 채점이 수행되었다(본 실험에서는 phase3에서 채점이 종료됨).

<표 9> Phase별 Pass/Fail 비율

Phase	1번 문항		2번 문항	
	total responses = 2443			
	Pass(%)	Fail(%)	Pass(%)	Fail(%)
Phase 1	2300 (94.15%)	143(5.85%)	2041(83.54%)	402(16.46%)
Phase 2	73 (51.05%)	70(48.95%)	290(72.14%)	112(27.86%)
Phase 3	70 (100%)	0(채점종료)	112(100%)	0(채점 종료)
Phase 4	-	-	-	-

4.2.3 채점자 신뢰도

1번 문항에 대한 10명의 채점자에 대한 Phase 1의 채점자 신뢰도 결과를 보면 다음과 같다. 채점자 신뢰도를 산출하기 위하여 동일한 답안을 채점한 두 채점자들 간의 점수의 적률상관계수를 산출하였다(<표 10>).

1번 문항에 대한 채점자 간의 상관관계는 대체로 상관관계수 0.7을 넘고 있어 채점자 간의 일관성 있는 채점이 이루어지고 있는 것으로 나타났다.

<표 10> Phase 1 종료 후 채점자간 상관 계수(1번 문항)

Pater.	U01	U02	U03	U04	U05	U06	U07	U08	U09	U10
U01										
U02	0.905									
U03	0.624	0.838								
U04	0.783	0.669	0.716							
U05	0.871	0.909	0.882	0.731						
U06	0.740	0.756	0.876	0.845	0.831					
U07	0.550	0.549	0.672	0.828	0.57	0.728				
U08	0.869	0.858	0.715	0.793	0.718	0.732	0.541			
U09	0.802	0.861	0.751	0.743	0.825	0.846	0.711	0.668		
U10	0.741	0.774	0.818	0.867	0.886	0.795	0.697	0.783	0.884	

현재의 채점 설계는 단순히 두 채점자가 동일한 문항을 채점하는 것이 아니라 한 채점자가 다른 9명의 채점자와 공통 답안을 순환적으로 공유하기 때문에 개별 채점자마다 9개의 상관계수를 얻을 수 있다. <표 10>에서 채점자 U07의 경우 총 9명의 채점자 중 다른 4명과의 상관관계수에 있어서 0.6이하의 수치를 보이고 있어 신뢰성이 떨어지는 채점자라고 볼 수 있다.

즉 모든 채점자간 신뢰도 정보를 활용하여 특정 채점자의 일관성, 공정성의 수준을 판단할 수

있는 자료가 제공된다.

다음으로 2번 문항에 대한 10명의 채점자에 대한 Phase 1의 채점자간 신뢰도를 산출하기 위하여 동일한 답안을 채점한 두 채점자들 간의 점수의 적률상관계수를 산출하였다(<표 11>).

<표 11> Phase 1 종료 후 채점자간 상관 계수(2번 문항)

Rater	U01	U02	U03	U04	U05	U06	U07	U08	U09	U10
U01										
U02	0.616									
U03	0.782	0.808								
U04	0.752	0.772	0.780							
U05	0.746	0.741	0.836	0.667						
U06	0.683	0.812	0.718	0.688	0.781					
U07	0.582	0.769	0.746	0.795	0.734	0.752				
U08	0.813	0.730	0.850	0.677	0.824	0.743	0.594			
U09	0.722	0.776	0.746	0.618	0.784	0.795	0.717	0.729		
U10	0.711	0.717	0.748	0.835	0.851	0.838	0.869	0.797	0.633	

2번 문항에 대한 채점자 간의 상관은 대체로 1번 문항에 비해 전반적으로 다소 낮은 상관계수를 보이고 있다. 2번 문항 역시 U07 ID를 가진 채점자의 경우 U01 채점자와 가장 낮은 상관인 .582로 나타났다. 2번 문항은 12점을 총점으로 하여 2점씩 증분 되는 총 7개의 부분점수를 갖기 때문에 5개의 부분점수 구간을 갖는 1번 문항에 비해 다소 낮은 상관을 보이고 있다. 그러나 전체적으로 모든 채점자가 유사한 수준의 일관성을 보이고 있음을 알 수 있다.

5. 결론 및 논의

채점자의 판단에 근거하여 피험자의 능력에 대한 점수를 부여하게 되는 서답형 문항의 채점과정에서 무엇보다 채점자가 피험자 집단에 대해 일관성 있게 판단을 하고 있는지, 다른 채점자와는 유사하거나 동일한 판단을 하고 있는지에 대한 평가는 매우 중요하다. 따라서 한 채점자가 피험자 집단에 대해 일관성 있게 판단하였는가와 한 채점자의 판단과 다른 채점자의 판단이 일치하는가를 검토할 필요가 있다.

온라인 채점 방식 개발 전의 서답형 답안 채점 방식은 대규모의 채점자가 특정 시기에 특정 장소에 모여 채점을 진행해야한다는 물리적 어려움

뿐만 아니라, 채점자간, 채점자내 신뢰성을 확보하기 위해서는 답지 배포 및 관리에 많은 노력이 요구되었다. 또한 개별 채점자가 채점과정에서 어느 정도의 노력을 투여하고 어떤 식으로 채점 시간을 배분하고 있는지에 대한 정보를 파악하기도 어려웠다[10].

이러한 종래 채점방식의 문제점을 개선하기 위하여 개발된 온라인 채점방식은 서답형 채점이 용이하게 이루어질 수 있게 하는 효율성뿐만 아니라 채점결과의 측정 이론적 질을 관리하는데 보다 효과적이라 할 수 있다.

먼저, 온라인 채점 방식은 채점자별 로그인, 로그아웃 시간의 기록을 통하여, 실제 개별 채점자가 채점을 위하여 얼마나 자주, 어느 정도의 시간을 소요했는지를 명확히 파악할 수 있기 때문에 전체 채점 일정과 개별 채점자의 시간 활용을 효율적으로 관리할 수 있다.

둘째, 응답지를 컴퓨터 알고리즘으로 효율적으로 배분이 가능하고, 모든 응답지에 두 사람이 채점하는 방식을 도입함으로써 개별 채점자는 다른 채점자들과 답지를 공유하기 때문에 채점자내 일관성과 신뢰성을 파악할 수 있는 정보를 제공한다. 나아가 이러한 정보를 통하여 채점 초기 혹은 진행 중 일관되지 못한 채점자의 채점을 중지시킬 수도 있다.

셋째, 채점자가 직접 컴퓨터에 개별 점수를 입력하기 때문에 paper and pencil 채점 방식에 비해 채점 결과의 관리가 용이하며 최종 결과산출을 위한 소요 시간을 단축할 수 있다.

이러한 온라인 채점의 장점에도 불구하고 최초 채점 시스템을 개발, 활용하기 위한 비용이 많이 든다는 점과 온라인상의 보안 유지의 문제는 지속적으로 해결되어야 할 과제일 것이다. 또한 채점자 연수를 온라인으로 시행하는 경우 이에 대한 효과성에 대한 검증이 필요할 것이다. 그러나 최초의 시스템 개발이 이루어진 이후에는 안정적으로 채점과정을 진행할 수 있으며, 시스템의 수정을 통하여 다양한 평가 프로그램에 유연하게 적용할 수 있다는 것은 온라인 채점 시스템의 유용한 장점일 것이다.

무엇보다도 대규모 서답형 평가에서 종전의 채점자를 집합시켜 문항을 채점하는 방식과 비교할

때, 온라인 채점은 채점자의 채점과정, 채점자내 및 채점자간 채점 일관성에 대한 다양한 정보를 제공한다는 점에서 신뢰성이 높고, 타당성 있는 채점 결과를 산출할 수 있는 유용한 방안임을 알 수 있다.

참 고 문 헌

[1] Sireci, S. G. & Zenisky, A. L. (2006). Innovative item formats in computer based testing: In pursuit of improved construct representation. In S. M. Downing & T. M. Haladyna(Eds.), Handbook of Test Development. Mahwah, New Jersey: Lawrence.

[2] Zenisky, A. L. & Sireci, S. G. (2002). Technological innovations in large-scale assessment. Applied Measurement in Education, 15, 337-362.

[3] PEM (2006). Distributed Scoring (Retrieved Nov, 2008 from www.pemsolutions.com)

[4] <http://nces.ed.gov/nationsreportcard/pubs/guide/ques10.asp>

[5] http://www.examinerrecruitment.org/aspects_of_examining/online.htm

[6] 박도순 외(2007). 교육평가-이해와 적용. 서울: 교육과학사.

[7] Allen, M. J. & Yen, W. M. (1979). Introduction to Measurement Theory. Long Grove, IL: Waveland.

[8] Cohen(1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 27-46.

[9] Glass, G. V. and Hopkins, K. D.(1984). Statistical methods in education and psychology. New Jersey: Prentice-Hall.

[10] 조지민 · 김경훈(2007). 학업성취도 평가 서답형 문항의 인터넷 기반 채점 도입 방안연구. 교육인적자원부.

조 지 민



1988 고려대학교 교육학과
(교육학사)
1990 미시간 주립대학교(석사)
1999 미시간 주립대학교(철학박사)
2002~ 현재 한국교육과정평가원 선임연구위원
관심분야: 교육측정 평가, 연구 방법과 통계
E-Mail: chojimin@kice.re.kr

김 경 훈



1980 서울교육대학교
1988 숭실대학교 전산공학과(학사)
1993 한양대학교(교육학석사)
2004 단국대학교(이학박사)
1998~현재 한국교육과정평가원 선임연구위원
관심분야: 알고리즘, 컴퓨터 교육
E-Mail: khkim@kice.re.kr