

서답형 문항의 인터넷 기반 채점시스템 설계 연구

조지민[†] · 김경훈^{††}

요 약

대규모 집단을 대상으로 시행하는 서답형 문항을 수기 채점 방식으로 채점하는 경우 채점 과정에서 발생하는 채점 오차를 줄이기 위해서는 많은 노력과 시간을 필요로 한다. 이러한 문제점을 해결하기 위해 세계 각국에서는 서답형 문항 채점 시스템을 개발하여 운영하고 있지만, 국내에서는 채점 시스템의 기초적인 연구도 미흡한 실정이다. 본 연구의 목적은 '서답형 문항의 인터넷 기반 채점 시스템' 설계의 기반을 마련하는데 있다. 시스템 설계의 기반 마련을 위한 시스템의 개요와 기본 설계도, 채점자 배정 방법, 채점자 신뢰도 추정에 의한 채점자 오류 검색 알고리즘을 제안하였다. 본 시스템은 실시간으로 채점 신뢰도를 추정하여 채점 오차를 보정할 수 있는 장점이 있다.

키워드 : 서답형 문항, 인터넷 기반 채점 시스템, 채점자 오류 검사

A Study on design of The Internet-based scoring system for constructed responses

Ji-Min Cho[†] · Kyung-Hoon Kim^{††}

ABSTRACT

Scoring the constructed responses in large-scale assessments needs great efforts and time to reduce the various types of error in Paper-based training and scoring. For the purpose of eliminating the complexities and problems in Paper and pencil based training and scoring, many of countries including U.S.A and England already have applied online scoring system. There, however, has been few studies to develop the scoring system for the constructed responses items in Korea. The purpose of this study is to develop the basic design of the Internet-based scoring system for the constructed responses. This study suggested the algorithms for assigning scorers to constructed responses, employing methods for monitoring reliability, etc. This system can ensure reliable, quick scoring such as monitor scorer consistency through ongoing reliability checks and assess the quality of scorer decision making through frequent various checking procedures.

Keywords : constructed response, Internet-based scoring system, reliability checks.

1. 서 론

† 정 회 원: 한국교육과정평가원 연구위원
 †† 종신회원: 한국교육과정평가원 선임연구위원(교신저자)
 * 논문 접수: 2007년 3월 13일, 심사완료: 2007년 3월 22일

컴퓨터를 이용한 정보통신 기술의 비약적인 발전과 이에 관련된 인프라의 구축이 확충됨에 따

라 인터넷이나 웹을 이용하지 않고서는 일상적인 생활이 어려울 정도로 다양한 분야에서 인터넷 이용하는 사례가 급증하였다.

교육 분야에서도 웹이나 인터넷을 교육에 접목하려는 연구는 공학자, 교육학자, 현장 교사 등 많은 연구자들에 의해서 끊임없이 이루어져 왔다. 인터넷이나 웹은 엄청난 잠재력을 가지고 있고, 사용하기에 따라서 교육의 질을 높일 수 있으며 많은 교육적 문제를 해결할 수 있는 가능성을 가지고 있다[2]. 최근에 들어서는 교육평가 분야에서 인터넷이나 웹을 이용한 평가 시스템의 개발이 활성화되고 있는 추세이다.

일반적으로 평가를 한다는 것은 어떤 대상의 가치나 질에 대해 판단을 내리는 것을 의미한다. 이러한 평가는 인간이 삶을 이루어 가는 가운데 의식적으로든 무의식적으로든 끊임없이 행하게 되는 기본적인 행동 중에 하나다. 우리는 살아가면서 무엇인가를 선택해야 할 때, 혹은 어떤 의사 결정을 내려야 할 때, 평가의 과정을 거쳐 그 답을 얻는다[5].

ICT를 이용한 평가 시스템은 VoiceXML 기반 영어 교육평가시스템설계 및 구현[10], 부분점수를 고려한 개별적용 평가 시스템[3] 등 많은 연구가 진행되어 왔으며, 대부분의 선행 연구들은 평가 문항의 출제, 평가 시행, 평가 결과 분석 등에 이르기까지 평가의 전 과정을 인터넷이나 웹을 기반으로 구현하였다는 점에서는 큰 의의가 있다고 할 수 있다. 그러나 이미 개발된 대부분의 평가 시스템은 검사를 시행하는 과정에서 응시자가 키보드 또는 마우스를 이용하여 답안을 작성해야 하므로 컴퓨터 활용 능력이 부족한 응시자의 검사 신뢰도에 영향을 주게 될 뿐만 아니라 동시에 대규모 집단을 대상으로 검사를 수행해야 하는 국가 또는 시·도 단위 평가에서는 현실적으로 적용하기 어렵다. 이러한 문제점을 극복하기 위해 지필 검사를 실시한 후, 지필 검사 답안지를 스캔하여 인터넷이나 웹 또는 LAN을 기반으로 채점 업무를 수행하는 채점 시스템의 개발과 활용이 점차 확대되고 있는 추세이다. 이러한 방식은 응시자가 컴퓨터 활용 방법이 미숙함으로 인해 발생하는 측정 오차를 최소화하면서 대규모 집단을 대상으로 동시에 검사를 실시해야 하는 평가에서

매우 유용한 방법이 될 수 있다.

대규모 집단검사에서 서답형 문항의 채점은 다수의 채점자가 서로 다른 응시자의 답안을 채점해야 하고, 응시자 개인의 답안은 2~3명의 서로 다른 채점자에 의해 채점이 이루어진다. 이 과정에서 채점자 오류(채점자내 오류와 채점자간 오류)를 최소화하여 검사의 신뢰도를 높이기 위한 노력으로 미국, 영국, 칠레 등과 같은 국가에서는 대규모 집단을 대상으로 한 평가에서 선다형이나 서답형 문항의 채점을 LAN, 인터넷 또는 웹을 기반으로 한 채점 시스템을 개발·운영하고 있다.

미국의 The College Board와 ETS에서 주관하는 SAT 에세이 시험의 채점은 PEM(Pearson Educational Measure)이라는 채점 전문회사에서 웹기반 방식으로 채점을 하고 있으며, 이 시스템의 공식적인 명칭은 '분배 채점(Distributed Scoring)'이다[17]. 또한 미국 내 국가 수준 학업 성취도 평가인 NAEP(National Assessment of Educational Process)의 채점을 담당하는 NCES(National Center for Educational Statistics)의 The Image-Based System[18], 영국의 QCA(Qualifications and Curriculum Authority)의 On-line Marking System[20] 등은 대규모 집단이 동시에 실시하는 검사에서 LAN, 인터넷이나 웹을 이용하여 채점 업무를 수행하는 대표적인 사례로 손꼽을 수 있다.

국내에서는 서답형 문항의 채점 과정에서 발생하는 채점자 신뢰도를 고려한 체계적인 채점 시스템 개발은 물론 기초적인 연구도 미흡한 실정이다. 또한 서답형 문항의 인터넷 기반 채점 시스템은 고가의 하드웨어 시스템(고속 이미지 스캐너, 다수의 서버 등)을 바탕으로 수기 답안의 스캔, 채점 수행, 오류 검증, 결과 분석에 관한 소프트웨어의 개발이 이루어져야 한다.

본 연구에서는 동시에 대규모 집단을 대상으로 실시하는 평가에서 적용 가능한 서답형 문항의 인터넷 기반 채점시스템 모형을 제안하고, 서답형 문항의 채점 과정에서 채점의 공정성과 타당도를 높일 수 있는 방안을 탐색하고자 한다.

2. 서답형 문항의 채점

2.1 서답형 문항의 개념

서답형 문항이란 응시자가 주어진 답지에 따라 정답을 선택하는 것이 아니고 직접 정답을 구성하여 작성하는 문항 형태를 의미한다[13]. 서답형 문항의 유형은 형태에 따라 단답형, 괄호형, 완성형, 서술형(논술형)으로 나누어 볼 수 있다.

단답형은 간단한 단어, 구, 절 혹은 수나 기호로 응답하는 문항 유형으로, 용어의 정의나 의미를 물을 때나 계산 문제에 자주 사용한다.

괄호형 문항은 질문을 위한 문장에 여백을 두어 응답을 유도하는 문항 형태를 의미하며, 완성형 문항은 질문을 위한 문장의 끝부분에 응답을 하게 하는 문항 형태로 괄호형 문항의 특수한 형태를 의미한다.

서술형은 문장 형태로 답을 작성하지만 객관적인 정답이 존재하는 형태이고 논술형은 논리성, 창의성 등을 평가하고자 할 때 사용하는 것으로 서술형과 구분하기도 하나 서술형과 논술형은 문장 형태로 답을 작성하는 동일한 의미의 문항 형식으로 사용되고 있다.

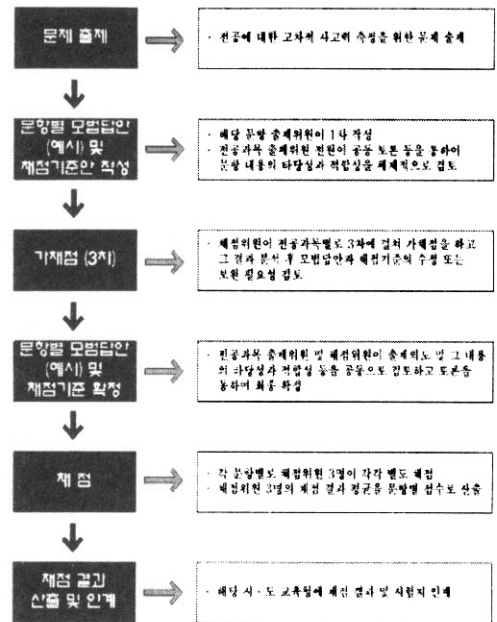
논술형은 응시생 반응의 허용정도에 따라 응답자유형과 응답 제한형으로 구분하는데, 응답자유형은 응답의 내용범위와 서술양식에 큰 제한이 가해지지 않는 문항 형태를 말한다. 응답 제한형의 경우 수험자의 반응 허용 정도가 상당히 제한된 것으로, 반응의 내용과 형식을 구체적으로 지시하여 제한된 방향으로 답을 구성하도록 요구하는 문항 형태이다.

응답 제한 방식에 따른 세부 유형으로는 분량 제한형, 내용범위 제한형, 서술양식 제한형으로 분류한다. 분량 제한형은 진술 요소의 수, 답안의 길이(문장 수, 글자 수, 또는 행수) 등에서 물리적 제한이 가해지는 문항 유형이며, 내용 범위 제한형은 응답할 내용 범위에 제한을 가하는 문항 유형이다. 서술양식 제한형은 응답할 서술양식에 제한이 가해지는 문항 유형을 말한다[6].

‘서답형’ 대신 ‘수행평가’라는 용어를 사용하기는 하지만 본 연구에서는 여러 가지 의미로 혼용되어 사용되는 혼란을 막기 위해 ‘서답형’이라는 용어로 통일하여 사용한다.

2.2 서답형 문항의 수기 채점 방식

수기 채점 방식(Paper and Pencil Scoring)은 서답형 답안지를 채점자가 직접 보면서 점수를 수기로 기록하는 방식을 말하며, 소규모 집단의 검사에서 일반적으로 많이 활용하는 방법이다.



<그림 1> 중등교사 임용 시험 채점 절차

대규모 집단검사에서 서답형 문항의 수기 채점 방식은 답안지의 분배, 채점 장소, 채점 결과 수합 및 검증, 결과 분석이나 활용에 어려움이 많다. 국내에서 서답형 문항을 수기 채점 방식으로 채점하는 ‘중등 교사 임용시험’ 전공분야의 채점 절차는 <그림 1>과 같이 문항별 모범 답안(예시) 및 채점 기준안 작성 → 기재점(3차) → 문항별 모범답안(예시) 및 채점기준 확정 → 채점 → 채점 결과 산출 및 인계 등의 순서로 이루어지며, 서답형 문항의 채점 신뢰도를 높이기 위한 체계적인 절차를 거치고 있음을 알 수 있다[19]. 그러나 채점 과정에서 채점자 간의 개인차 문제, 채점의 일관성 유지 등에 관한 검증 결과를 채점자에게 Off-line으로 제공하기 때문에 서답형 문항의 채점 시간과 비용이 증가하게 된다.

2.3 채점자 신뢰도

채점 과정과 결과에는 채점자의 개인적인 주관, 편견, 선입견, 종교, 지역, 도덕적·정치적 성향 등이 작용되지 않도록 해야 한다. 채점자 신뢰도는 어떤 채점자가 채점을 하더라도 동일한 채점 결과가 나올 수 있는 정도를 말하며, 채점자간 신뢰도와 채점자 내 신뢰도로 구분된다. 채점자간 신뢰도는 채점자들 간 얼마나 유사하게 채점을 하느냐를 측정할 값을 말하며, 채점자내 신뢰도는 한 채점자가 많은 평가대상에 대하여 일관성 있게 평가하였느냐를 의미한다.

채점자간 신뢰도 추정 방법은 점수가 연속변수인가 비연속변수인가에 따라 구분한다. 연속변수로 채점하는 경우, 상관계수를 산출하는 방법과 일반화가능도 이론을 적용한다. 비연속변수일 경우는 채점자 혹은 평정자간의 분류 일치도를 분석하는 일치도 통계(agreement statistics), 일치도 통계에서 우연에 의한 확률을 제거하고 평정자간의 일치도를 추정하는 Cohen의 Kappa 계수를 사용할 수 있다.

2.3.1 상관계수

상관계수로 채점자 간 신뢰도를 산출하는 경우 서로 다른 채점자가 동일한 수험생의 서답형 문항 답안에 얼마나 유사하게 점수를 부여하였느냐를 분석하기 위해 Pearson 단순적률상관계수 <공식 (1)>을 산출한다.

$$r_{xy} = \frac{s_{xy}}{s_x s_y} \quad \text{공식 (1)}$$

S_{xy} 는 채점자 두 사람의 채점 점수 공분산

S_x, S_y 는 각 채점자의 채점 점수의 표준편차[16]

상관계수가 수치 1에 가까울수록 각각의 채점자가 동일한 채점 결과를 보이는 것으로 해석할 수 있으며, 이와는 반대로 상관계수가 수치 1에서 멀어질수록 채점자간 서로 다른 채점 결과를 보이는 것으로 해석할 수 있다. 국가 수준의 대규모

평가에서 채점자 간 상관계수는 0.8 이상의 수치를 기대하지만 절대적인 기준은 없고 검사의 규모와 성격 등에 따라 그 보다 낮은 상관계수로도 채점자간 신뢰성을 인정할 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.

2.4 일치도 통계와 Kappa 계수

일치도 통계와 Kappa 계수는 채점자가 관찰대상의 행위나 수행결과에 점수를 부여하기 보다는 어떤 유목이나 범주로 분류할 때, 채점자간의 분류 일치도를 추정하는 방법이다[8]. 예로 A, B 채점자가 채점한 N명의 채점 결과를 j 개의 범주로 분류하여 평정한 결과는 <표 1>과 같다고 하자.

<표 1> 두 채점자에 의한 채점 결과

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 $J \times J$ 분할표에서 대각선에 위치한 $N_{11}, N_{22}, N_{33}, \dots, N_{JJ}$ 로서 두 채점자간의 일치도 통계(agreement statistics)는 <공식 (2)>에 의한다.

$$P_A = \frac{N_{11} + N_{22} + N_{33} + \dots + N_{JJ}}{N} \quad \text{공식 (2)}$$

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[15]. 우연에 의하여 J × J 분할표의 대각선에 있는 칸에 분류되는 사례 수는 <공식 (3)>에 의하여 계산된다.

$$N_{jc} = \frac{N_j \times N_c}{N} \quad \text{공식 (3)}$$

전체 N명의 응시자 중 우연히 두 채점자에 의하여 일치된 채점을 받은 응시자의 확률은 <공식 (4)>에 의한다.

$$P_c = \frac{N_c}{N} \quad \text{공식 (4)}$$

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 (5)>와 같다.

$$K = \frac{P_A - P_c}{1 - P_c} \quad \text{공식 (5)}$$

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[6].

2.4.1 일반화가능도 이론

선택형 문항으로만 이루어진 검사와 달리 '채점자'를 포함하고 있는 서답형 문항으로 구성된 검사에서 일반화가능도 이론을 이용한 연구가 필수적인 요소로 인정되고 있고, 연구결과를 활용한 다양한 측정 모형을 적용한 연구가 점차 활발하게 진행되고 있다[1].

일반화가능도(generalizability) 이론은 고전검사 이론에서의 신뢰도 산출과 달리 오차점수의 분산

을 여러 개의 구성요소로 세분하여 구성요소의 크기를 규명하고 각 오차의 원인을 밝히는 특징을 지니고 있다. 일반화가능도 이론의 목적은 오차점수 분산의 원인을 밝힐 뿐 아니라, 측정방법, 절차, 그리고 목적에 따라 어떤 요인이 어떤 오차점수와 관계가 있는지 밝힌다. 그러므로 일반화가능도 이론이란 고전검사이론의 연장으로 중다 오차원인을 포함하는 측정모형에 특정 분산분석 절차를 적용하는 것으로 간주할 수 있다[7][9][14].

이 이론은 응시자를 평가하는데 적절한 측정조건, 예로 문항 유형과 채점자 유형 등을 결정한 후 먼저 G 연구(generalizability study)를 통해 이들 조건과 관련된 분산 성분을 추정한 다음에 다음 단계로 D 연구(Decision study)에서는 이러한 분산 성분 추정치를 활용하여 효율적인 측정절차나 응시자에 대한 계획을 세우게 된다[4] [11]. 예를 들어 채점자 i명이 p명의 서답형 문항 답안 채점을 한 단일국면 교차설계의 경우 <표 2> 일원 분산분석의 반복측정과 동일한 결과를 얻게 된다.

<표 2> 단일국면 교차설계의 분산분석표

효과	df	SS	MS	분산값
채점자 (i)	i - 1	SS _i	SS _i /(i-1)	σ _i ²
응시자 (p)	p - 1	SS _p	SS _p /(p-1)	σ _p ²
상호작용 (ip)	(i-1)x(p-1)	SS _{ip}	SS _{ip} /(i-1)(p-1)	σ _{ip} ²

$$\sigma_i^2 = [MS_i - MS_{ip}]/p$$

$$\sigma_p^2 = [MS_p - MS_{ip}]/i$$

$$\sigma_{ip}^2 = MS_{ip}$$

채점자 효과에 해당하는 σ_i²이 크면 채점자간의 채점기준이 상이하여 채점자들의 주관이 채점에 영향을 주었음을 뜻한다. 채점자에 의한 분산값이 작을수록 채점의 주관이 배제되고 동일한 채점기준에 의하여 응시자의 수행결과에 대한 채점이 이루어졌다 할 수 있다. 또한 채점자와 응시자의 상호작용에 해당하는 σ_{ip}²이 크면 채점자가 채점의 일관성을 유지하지 못한 것으로 분석할 수 있다.

일반화가능도 계수는 응시자 분산(σ_p^2)에 채점자와 응시자의 상호작용 분산(σ_{ip}^2)을 더한 값으로 응시자의 분산(σ_p^2)을 나누어 추정된다. 이처럼 일반화 가능도 이론은 오차분산의 요인을 분석하여 채점자 간 분산의 요인을 최소화할 수 있는 자료를 제공하는 등 채점 모형의 설계에 도움을 주게 된다.

2.4.2 채점자 연수

최종 채점된 검사 결과가 응시자들에게 영향을 미칠 수 있는 중요한 자료로 활용된다. 그러므로 채점 결과에 근거하여 내리게 되는 각종 해석 및 판단, 더 나아가 의사결정이 얼마나 타당한지를 점검해 보는 노력이 계속해서 이루어져야 한다. 세부적인 사항으로 첫째, 채점기준이나 예시가 특정 수험생에게만 유리하게 개발된 것은 아닌지 검토되어야 한다. 채점기준의 명료성, 특히 부분점수의 구분은 명확한지, 성별이나 지역 등 특정 집단의 수험생에게만 특정 답안의 형식(유형)이 유리하게 개발된 문항은 아닌지가 검토되어야 한다.

둘째, 채점자 선정과 채점자 연수가 제대로 이루어진 것인지에 대한 점검이 이루어져야 한다.

셋째, 채점의 전체 과정 속에서 채점자 내 및 채점자 간 신뢰도 산출 등의 노력으로 공정하고 타당한 채점 결과의 산출이 이루어지도록 해야 한다.

마지막으로는 최종 채점결과에 대한 채점자 간 신뢰도의 분석 결과 등 사후 보고를 통해 더 개선된 서답형 문항 출제 등에 기여할 수 있도록 해야 한다.

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 기준틀을 공유하도록 하기 위함이고 이는 궁극적으로 신뢰롭고 타당한 채점 결과를 확보할 수 있게 한다. 우선 채점자의 선정에서부터 검사의 목적과 결과 활용 등에 적합한 구성 및 선정이 우선시 되어야 한다.

채점자 선정에 있어서 무엇보다 서답형 문항 내용 분야의 전문적인 지식과 교육 경험을 갖춘 채점자 선정이 중요하고, 또한 최연희(2002)의 연구결과에서 보여준 채점 경험이 없는 초보자를 위한 '채점자 교육'과 숙련된 채점자를 위한 장기적인 '채점자 조정과정'의 단계적 훈련과정을 통해 각각의 채점자들에게 적합한 훈련 내용을 제공하여 채점자 신뢰도를 높일 수 있는 기회를 제공할 필요성도 고려가 되어져야 한다[12].

3. 서답형 문항의 Online 채점방식

현재 개발되어 운영되고 있는 서답형 문항 On-line 채점 시스템은 Distributed Scoring, Image Based System, Online Marking System, Online Scoring Network, Internet-Based OSN(Online Scoring Network) System 등과 같이 각 시스템의 특성 또는 운영 기관에 따라 다양한 명칭을 사용하고 있다. 그러나 이들 시스템의 구현과 운영 방식에 따라 LAN 기반 채점 방식, 인터넷 또는 웹 기반 채점 방식으로 구분할 수 있다.

3.1 LAN 기반 채점 방식

'LAN 기반 채점 방식'이란 LAN의 정의에서 알 수 있듯이 인터넷이나 웹과 같은 개방된 네트워크 체제가 아닌 제한된 범위의 일정 지역 내에서 다수의 컴퓨터를 속도가 빠른 통신 선로로 연결하여 기기간의 통신을 가능하도록 하는 근거리 통신망을 이용한 채점 방식을 말한다. 개방적인 인터넷을 이용하지 않고 제한적인 소규모의 독립된 네트워크를 기반으로 운용되므로 극도의 보안을 유지해야 하는 서답형 문항의 채점에 적합한 방식이다. 앞서 언급했던 NAEP의 1994년 평가에서 LAN 기반 채점 방식을 도입하였으며, NCES의 공식적인 명칭은 이미지 기반 시스템(The Image-Based System)이다. 칠레의 국가 수준 학업성취도 평가의 채점 업무도 LAN 기반 채점 방식으로 운영하고 있다[21].

3.2 인터넷과 웹 기반 채점 방식

현재 인터넷과 웹을 거의 구분하지 않고 사용하는 경향이 있는데, 이는 인터넷과 웹을 이용하는 사용자 입장에서는 둘 사이에 존재하는 차이점을 느끼지 못하는데서 기인한다고 볼 수 있다.

인터넷(Internet)은 전 세계의 컴퓨터를 연결하는 거대한 컴퓨터 통신망을 말하며, 1980년대 말부터 TCP/IP 프로토콜을 바탕으로 한 세계 각국의 컴퓨터 통신망들이 연결되면서 수많은 컴퓨터가 연결된 형태의 인터넷이 시작되었다.

‘인터넷 기반 채점 방식’은 인터넷 상에서 서버와 채점자 PC 간에 답안지와 채점 결과에 관한 데이터를 주고, 받는 방법으로 채점을 수행해 나가는 채점 방식을 지칭한다.

웹(World Wide Web)은 하이퍼텍스트의 원리를 응용하여 인터넷을 활용하는 하나의 방법적 체제로, 기술적으로 말하면 TCP/IP라는 통신 규약을 바탕으로 성립하는 HTTP(Hyper Text Transfer Protocol) 규약을 지키는 통합적인 자료 교환을 목적으로 구안된 자료 전달 체제를 말하며, 인터넷이 없다면 웹은 존재할 수 없다.

‘웹기반 채점 시스템’은 텍스트, 이미지, 동영상, 사운드 등이 하이퍼텍스트의 원리 하에 한꺼번에 종합적으로 통합된 정보교환 방식을 통하여 채점 업무를 수행하는 것을 말하며, ‘인터넷 기반 채점’ 방식에 비해 사용자의 편의성이 향상된 시스템을 지칭한다.

본 연구에서는 웹보다는 좀 더 포괄적인 의미의 ‘인터넷 기반 채점 시스템’이라는 용어를 사용한다.

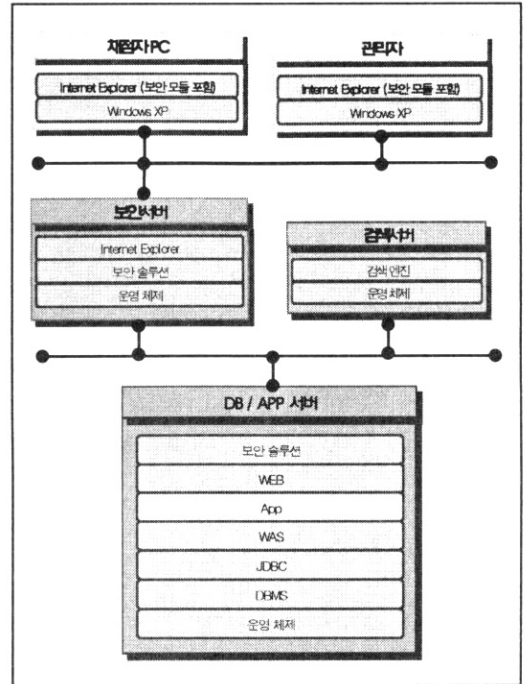
4. 서답형 문항의 인터넷 기반 채점 시스템 설계

4.1 시스템의 개요와 특징

4.1.1 시스템의 개요

본 연구에서 설계하는 서답형 문항의 인터넷 기반 채점 시스템의 목표는 대규모 집단을 대상으로 동시에 지필 검사로 실시하는 서답형 문항을 효율적으로 채점하고, 채점과정에서 발생하는

채점자(채점자 내, 채점자 간)오류를 최소화하는데 있다.



<그림 2> 채점 시스템 구성도

본 시스템의 어플리케이션 모듈(<그림2>의 App)은 서답형 답안의 이미지 파일을 생성하고 저장·관리하는 입력 모듈, 임의의 답안 이미지 파일을 주어진 조건에 따라 추출하여 복수의 채점자에게 분배하는 채점 모듈, 채점자 오류 점검과 채점 결과를 점검하는 검증 모듈, 채점 결과를 분석하고 활용하는 분석 및 활용 모듈 등과 같은 하위 모듈로 구성된다.

검사 실시 이전에 수행하는 서답형 문항의 설계와 검사지 개발 등에 관련된 내용은 본 연구 범위에 포함하지 않았다.

4.1.2 시스템의 특징

본 연구에서 설계하는 시스템의 주요 특징은 첫째, 국가 수준(대규모 집단을 대상으로 검사 결과의 공신력을 최우선으로 고려하는 평가)에서 시행 하는 서답형 문항을 채점하기 위한 시스템

이지만, 집단의 크기 또는 문항의 유형에 관계없이 소규모 집단 평가나 선다형 평가에도 적용 가능한 시스템을 지향한다.

둘째, 수기 채점 방식에 비해 서답형 문항 채점의 효율성과 채점 결과에 대한 신뢰성을 획기적으로 향상시킬 수 있는 시스템이다.

셋째, 인터넷 기반 채점 방식을 취하고 있지만, LAN 기반 채점 방식으로도 적용할 수 있는 시스템이다.

넷째, 종전의 지필 검사(paper and pencil test) 방식으로 검사가 실시된다는 것을 전제로 하되 인터넷이나 웹기반 검사에서도 적용할 수 있는 시스템을 지향한다.

다섯째, 채점은 2인 이상의 채점자의 이루어지며, 특정 문항의 채점은 복수의 채점자 집단으로 분산하여 채점하는 것을 전제로 한다.

4.2 답안지 스캔

4.2.1 답안지 설계시 고려할 점

일반적인 지필 평가에서 문답지는 평가 문항의 유형, 응시자의 편의성, 채점 관리의 용이성, 응시 집단의 규모에 따라 그 형식을 달리한다.

문제지와 답안지의 분리 여부에 따라 문답지를 통합하는 방식과 문답지를 분리하는 방식을 들 수 있다. 학생의 응답 편의성을 고려한다면 평가 문항의 유형(선다형 또는 서답형)에 관계없이 문제지와 답안지를 함께 배열하는 방식이 좋은 방법이지만 채점 관리의 용이성을 고려한다면 문지와 답지를 별도로 분리하여 검사 도구를 제작하는 것이 유리하다.

문답지의 구성이나 유형은 절대적인 기준이 존재하는 것이 아니라 응시자의 연령, 문항의 특성, 집단의 크기 등에 따라 의사결정이 이루어진다.

4.2.2 답안 스캔 방법 설계

대규모 집단 평가에서 채점의 정확성과 신속성을 위하여 일반적으로 컴퓨터를 이용한 채점 방식을 많이 이용하고 있다. 이러한 채점 방식은 응시자가 작성한 수기로 작성한 답안을 컴퓨터 파

일로 변환하는 스캔 과정이 필수적이다.

현재 활용되는 스캔 방식은 OMR(Optical Mark Recognition) 방식, ICR(Intelligent Character Recognition) 방식, OCR(Optical Character Recognition) 방식 등이다. OMR 방식은 선다형 평가에서 매우 유용하지만 서답형 평가에는 적용하기 어렵다. OCR 방식과 ICR 방식은 OMR 방식에 비해 진보된 방식으로 답안의 문자나 표기를 인식(잘 인쇄된 영문자의 경우 99.5%의 정확도)할 수 있으나 응시자가 수기로 작성한 답안 정보를 완벽하게 판독해야 하는 경우, 현재의 기술 수준으로는 한계가 있다.

본 연구에서 개발하는 시스템은 이러한 어려움을 극복하고자 Image Scan 방식을 병행하여 적용하고자 한다. 응시자 정보는 스캐너에서 지원 가능한 OMR, OCR, ICR 등의 방식으로 스캔하고, 동시에 답안 내용은 이미지 형태로 스캔하여 이들 정보를 답안 Database에 저장한다.

일반적인 검사에서 채점 과정에 필요한 답안의 스캔 정보는 응시자 정보(문자와 숫자), 선다형 답안 내역(숫자), 답안 이미지의 조합으로 구성된다<표 4>.

<표 4> 답안지 스캔 정보 예시

스캔 내용	형식	설명
답안지 정보	문자, 수	과목명, 답안 면수 등
응시자 정보	수, 문자	수험번호, 성명 등
선다형 문항	수	선다형 답란
서답형 문항	이미지	서답형 답란

4.3 채점 설계

4.3.1 채점 기준 설계

서답형 문항의 출제과정에서는 정답과 채점 기준표를 작성하는 것이 매우 중요하다. 이때 정답은 모범/예시 답안의 형식으로 작성하게 되는데 가장 먼저 모범/예시 답안을 작성한 후 모범/예시 답안을 구성하는 채점 요소 각각에 대하여 배점을 명시하는 방법으로 채점 기준표를 작성하게

된다.

예를 들어, 2006년 국가수준 학업성취도 평가 고등학교 1학년 과학 교과의 배점 3점인 서답형 문항의 채점 기준표 예시는 <표 5>와 같다.

<표 5> 서답형 문항의 채점 기준표 예시

3점	NaCl 아래와 같이 쓴 경우 $Na^+ + Cl^- \rightarrow NaCl$ $NaCl \rightarrow Na^+ + Cl^-$ $NaCl \ Na^+, Cl^-$ $NaCl + Na^+ + Cl^-$ ※ Na^+, Cl^- 오답으로 처리함. NaCl 화학식을 정확히 쓴 경우
2점	아래와 같이 쓴 경우 Na^+Cl^- , $NaCl^-$, Na^-Cl , Cl^-Na^+
1점	아래와 같이 쓴 경우 $ClNa$, $Na^+ + Cl^-$, $Cl^- + Na^+$, $ClNa^+$, $Cl + Na^+$

이 과정에서 중요한 것은 예로 문항 배점을 3점으로 하였을 경우 채점 기준을 0, 1, 2, 3점으로 구성하고 검사 시행 후 실제 수험생들의 최종 점수는 각각의 배점인 0, 1, 2, 3점에 모두 분포가 될 수 있도록 채점 기준표를 명확히 하고 상세화하여야 한다. 이는 궁극적으로 문항의 양호도를 갖춘 서답형 문항의 개발과 직결되는 문제가 되기 때문에 채점 기준의 상세화는 매우 중요한 과정이다. 이를 위해 각 배점에 해당되는 채점 기준의 상세한 설명과 더불어 상세한 사례의 예시를 구체적으로 제시하거나 나열하는 것이 바람직하다. 이때 배점의 구분에 있어서 전체 문항의 배점이 4점이 경우 배점을 0점에서 4점까지의 각각의 정수 부분점수를 주거나, 0점 2점 4점 이렇게 동일한 간격의 부분점수 배점을 부여해야만 한다.

4.3.2 채점자 배정 모듈 설계

응시자 수와 채점 기간, 문항 수 등을 고려한

채점자 수가 결정되면 어떤 채점자에게 어떤 응시자의 응답지를 배정해야 할지 결정해야 한다. 응답지를 무작위로 각 채점자에게 배분하되 신뢰롭고 타당한 채점 결과를 얻기 위해서는 좀 더 체계적인 배분 절차가 필요하다.

채점자의 여러 가지 인적 배경과 연수 과정에서 얻을 수 있는 정보를 근거로 응시자 응답을 어떤 채점자에게 어떻게 배정해야 하는지를 고려해야 한다. 예를 들어 p명이 응시한 서답형 1문항을 n명의 채점자에게 배정하되, 모든 응시자의 각 답안은 2명의 채점자에 의해 채점하는 경우 각 채점자에게 문항을 배정하는 알고리즘은 다음과 같다.

<채점자 배정 알고리즘>

- 1 단계: 응시 인원(p), 채점자 수(n)
- 2 단계: 응시자의 답안 P_i ($i=1$ to p)를 무작위 추출하여 n개의 Group으로 배정한다.

$$\text{Random}(P_i) \rightarrow \text{Group}(x), (x=1 \text{ to } n)$$

- 3 단계: 채점자(Marker)를 랜덤 추출하여 채점자 번호 M_x 를 부여한다.

$$\text{Random}(\text{Marker}) \rightarrow M_x, (x=1 \text{ to } n)$$

- 4 단계: 채점자(M_x)에게 2개의 채점할 답안 그룹을 배정한다.

```

For x=1 to n {
  if x ≠ n then
    Group(x)+Group(x+1) → Marking( $M_x$ )
  else if
    Group(x)+Group(1) → Marking( $M_x$ )
  end if
}
    
```

채점자 배정 알고리즘의 수행 결과는 <표 6>과 같다.

<표 6> 채점자 배정 알고리즘 수행 결과

채점자 \ 답안	1	2	3	4	...	n-1	n
Group(1)							
Group(2)							
Group(3)							
Group(4)							
...							
Group(n-1)							
Group(n)							

기존의 Online 채점 시스템에서는 채점자 간의 상관계수의 변화 추이를 통해 채점자 간 신뢰도를 측정하고 있다.

본 연구에서 설계하는 시스템에서 채점자 배정을 위와 같이 수행하는 이유는 채점자 간 신뢰도는 물론이고, 종전의 방식에서 추정하기 어려웠던 채점자 내 신뢰도(채점자의 채점 일관성)를 검출하기 위함이다. 이에 대한 구체적인 방법은 4.3.3에서 논한다.

4.3.3 채점 신뢰도 분석 모듈 설계

서답형 문항의 채점은 다수의 채점자에 의해 채점이 수행되기 때문에 채점자 신뢰도는 평가 전반의 성패를 가늠할 수 있는 중요한 척도이다. 인터넷 기반 채점 시스템의 장점 중의 하나는 채점 과정에서 발생하는 채점자 오류를 실시간으로 검증할 수 있다는 점이다.

채점과정에서 채점자 오류가 감지되면, 이에 대한 원인을 규명하고 해당 채점자가 수행한 채점 결과를 분석하여 점수를 보정할 것인지, 재 채점을 할 것인지에 대한 의사결정이 이루어진다.

본 연구에서는 채점자간 신뢰도 분석을 바탕으로 채점자 내 신뢰도 오차를 추정할 수 있는 방법을 제안한다.

예를 들어 4.3.2의 <표 6>과 같이 채점자를 배

정하여 채점을 수행하는 경우, 채점자 간 신뢰도를 분석하여 채점 오류자를 검색하는 알고리즘은 다음과 같다.

< 채점 오류자 검색 알고리즘 >

1단계: 채점자 (M_x)는 배정된 문항(Marking (M_x))을 채점(Score(P_i))한다.

$$Marking(M_x) \rightarrow Score(P_i)$$

2단계: 2.3.1.의 공식(1)(또는 공식(5))을 이용하여 두 채점자 (M_x, M_{x+1})의 채점 결과 값 $M_x(Score(P_i)), M_{x+1}(Score(P_i))$ 의 상관 계수 값($r_{x(x+1)}$)을 계산한다.

```

For x= 1 to n {
  if x ≠ n then
     $M_x(Score(P_i)), M_{x+1}(Score(P_i)) \rightarrow r_{x(x+1)}$ 
  else if
     $M_x(Score(P_i)), M_1(Score(P_i)) \rightarrow r_{x1}$ 
  end if
}
    
```

3단계: 인접한 채점자간의 신뢰도 값($r_{x(x+1)}$) 이 기준값(C) 이하이면, 인접한 다음 채점자와의 상관계수($r_{(x+1)(x+2)}$)를 바탕으로 오류 채점자 집단을 추출한다. 첫번째 오류자(First_Error)부터 마지막 오류자 이전의 오류자(Last_Error-1)까지의 채점자는 채점 오류자이다.

```

For x = 1 to n {
   $M_x \rightarrow First\_Error$ 
  if  $r_{(x)(x+1)} \leq C$  then
     $M_{x+1} \rightarrow Last\_Error$ 
    For y = (x+1) to n {
      if  $r_{(y)(y+1)} \leq C$  then
    
```

```

        My+1 → Last_Error
    else if
        return (First_Error, Last_Error)
    }
}

```

4단계: 추출한 채점 오류자를 출력한다.

```

For x = First_Error to (Last_Error-1)
printf ("Error Marker ", Mx)

```

5단계: 채점 오류자가 있는 경우 채점 결과를 보정하거나 재 채점 작업을 수행한 후 '1 단계'로 돌아간다.

6단계: 인접한 채점자간의 신뢰도 값($r_{x(x+1)}$) 들이 모두 기준값(C) 이상이면, 두 채점자의 평균을 채점 점수로 하고, 채점을 완료한다.

이상의 알고리즘을 바탕으로 채점이 수행되었다면 기준값 이상의 신뢰도 값을 보장 받을 수 있고, 검사의 특성에 따라 기준 신뢰도 값을 증감하여 활용할 수도 있다.

5. 결론 및 제언

현재 우리나라에서 고도의 채점 신뢰도를 요구하는 대부분의 검사는 수기 방식으로 채점이 이루어지고 있기 때문에 채점 과정에서 발생하는 채점 오차를 줄이기 위해서 많은 노력과 시간을 투입하고 있다. 미국, 영국, 칠레 등의 국가에서는 채점자 오류를 최소화하여 검사의 신뢰도를 높이기 위한 노력으로 LAN, 인터넷 또는 웹을 기반으로 한 서답형 문항 채점 시스템을 개발하여 운용하고 있지만, 국내에서는 이에 대한 체계적인 채점 시스템의 개발은 물론 기초적인 연구도 미흡한 실정이다.

본 연구에서는 서답형 문항의 답안을 지필로 작성하는 대규모 집단검사에서 적용할 수 있는 '서답형 문항 인터넷 기반 채점 시스템'의 설계와

관련하여 시스템의 개요, 채점자 배정 방법과 채점자 신뢰도 추정에 의한 채점자 오류 검색 방법에 대한 알고리즘을 제안하였다.

본 연구에서 제안한 '서답형 문항의 인터넷 기반 채점 시스템'은 고가의 장비 마련을 위한 초기 투자비용이 발생하지만, 고도의 채점 신뢰도를 요구하는 평가에서 활용도가 매우 높고, 종전의 채점 방식에 비해 실시간으로 채점 신뢰도를 추정하여 채점 오차를 보정할 수 있는 장점이 있을 뿐만 아니라 네트워크의 범위를 달리하여 LAN 기반 채점시스템으로의 전환이 가능하다. 또한 인터넷 기반 검사에서 다수의 채점자가 채점을 수행해야 하는 경우에도 본 시스템에서 제안한 채점자 배정 및 채점 오류자 검색 모듈을 활용할 수 있다.

'서답형 문항의 인터넷(웹) 기반 채점 시스템'은 대입논술시험, 국가수준 학업성취도 평가, 중등교사 임용시험, 시·도 단위 평가 등과 같은 국가 또는 자치 단체 단위의 평가에 적용 가능한 시스템이다. 본 시스템의 개발과 적용은 서답형 문항 평가의 채점에 대한 신뢰도 향상은 물론 공신력을 제고하는데 기여할 것으로 기대한다.

참 고 문 헌

- [1] 강애남, 이규민(2006). 학생들의 동료평가를 활용한 수행평가 결과의 일반화가능도 분석. 교육평가연구, 19(3), 107-121.
- [2] 김정훈, 이화진, 강창동, 채선희, 양길석, 김완수(2001). KICE 교수-학습 지원센터 구축을 위한 기초 연구. 한국교육과정평가원 연구보고 PRO 2001-5.
- [3] 김소연, 홍의식(2006). 부분점수를 고려한 웹 기반 학습자 개별 적응 평가 시스템. 한국컴퓨터교육학회 논문지, 9(2).
- [4] 김양분(1991). 신뢰도 추정을 위한 새로운 접근: 일반화가능도이론 적용의 예. 교육개발 13(3), 103-110.
- [5] 김재춘, 부재율, 소경희, 채선희(2000). "교육과정과 교육평가", 교육과학사.

- [6] 박도순, 외(2007). 교육평가-이해와 적용-. 서울: 교육과학사.
- [7] 성태제(1994). 논술형 고사와 예체능계 실기고사를 위한 채점자간 신뢰도 추정. 교육평가연구, 7(1), 43-56.
- [8] 성태제(2005). 현대교육평가. 서울: 학지사.
- [9] 이종성(1988)(편저). 일반화가능도 이론. 서울:연세대학교 출판부.
- [10] 장승주(2006). VoiceXML 기반 영어교육평가시스템 설계 및 구현. 한국컴퓨터교육학회 논문지, 8(6).
- [11] 채선희(1996). 논술시험의 채점의 공정성과 효율성 확보방안. 교육평가연구, 9(1), 5-29.
- [12] 최연희(2002). 채점자 혼란이 영어 작문채점에 미치는 효과에 대한 연구. Journal of Applied Linguistics Association of Korea, 18(1), 257-291
- [13] 한국교육평가학회(편)(2004). 교육평가용어사전. 서울: 학지사.
- [14] Brennan, R. L.(1983). Elements of Generalizability Theory. Iowa: ACT Publication. p.1.
- [15] Cohen(1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 27-46.
- [16] Glass, G. V., and Hopkins, K. D.(1984). *Statistical methods in education and psychology*. New Jersey: Prentice-Hall.
- [17] <http://www.pemsolutions.com/scoring/index.htm>
- [18] <http://nces.ed.gov/nationsreportcard/pubs/guide/ques10.asp>
- [19] <http://www.kice.re.kr/kice/contents/m100330/view>
- [20] <http://www.examinerrecruitment.org/aspectsofexamining/online.htm>
- [21] <http://www.simce.cl>

조 지 민



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

김 경 훈



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