

예비 국어교사의 성별 및 학생 성별 인식에 따른 평가 차이 분석

박 영 민(한국교원대학교 부교수)*

최 속 기(한국교원대학교 강 사)

《 요 약 》

이 연구는 평가자의 성별과 평가자의 학생 성별 인식이 쓰기 평가에 영향을 미치는지를 확인하는 데 목적이 있다. 이를 위하여 예비 국어교사 69명을 표본으로 선정하여 학생 설명문 35편과 학생 논설문 35편을 각각 평가하게 한 후, 그 평가 자료를 분석하였다. 이 연구를 통해 얻은 결과는 다음과 같다. 첫째, 예비 국어교사들은 학생 설명문과 논설문 평가에서 각각 평균 14.27점, 14.02점을 부여하였는데, 이는 척도의 실질 평균인 15점보다는 낮은 것으로 나타났다. 이러한 결과는 예비 국어교사들이 학생 글을 엄격하게 평가하는 경향이 있다는 사실을 지지해 준다. 둘째, 예비 국어교사의 성별에 따른 쓰기 평가의 차이는 설명문이든 논설문이든 모두 발견되지 않았다. 그러나 논설문을 평가한 남자 예비 국어교사들은 평가 결과에서 상대적으로 큰 편차를 보여 일치도 확보를 위한 교육적 노력이 필요하다는 점을 알 수 있었다. 셋째, 예비 국어교사의 학생 성별 인식에 따른 평가 결과의 차이도 발견되지 않았으며, 동성 평가 절하 현상도 발견되지 않았다. 그러나 남자 예비 국어교사들은 학생 성별을 여학생으로 인식할 때 표준편차가 크게 나타남으로써 평가자간 신뢰도를 확보할 수 있는 방안을 강구해야 한다는 점을 알 수 있었다. 이 연구의 결과를 일반화하기 위해서는 현직 국어교사를 대상으로 하는 후속적인 연구가 이루어져야 하며, 남자 예비 국어교사의 쓰기 평가 전문성 신장을 위한 노력이 뒷받침되어야 한다.

주제어 : 쓰기 평가, 예비 국어교사, 평가자 성별, 학생 성별 인식

* 제1저자 및 교신저자, enrapture@knue.ac.kr

I . 서론

쓰기 평가는 신뢰도가 낮다는 본질적인 한계를 가지고 있다. 높은 수준의 평가자내 신뢰도를 확보하기도 어렵고, 평가자간 신뢰도를 유지하기도 어렵다. 그래서 쓰기 연구자들은 신뢰도 확보를 위하여 쓰기 평가 방법과 절차를 체계화하는 데 노력을 기울여 왔다. 쓰기 평가에서 신뢰도를 확보하고 유지하는 것이 어려운 이유는, 다른 주관적 평가가 그러하듯 글을 평가하는 여러 가지 관련 요인이 평가자에게 영향을 미치기 때문이다. 평가자가 가지고 있는 배경지식, 인지 양식, 쓰기 신념, 쓰기 동기 등에 따라 평가 결과가 달라지기도 하고, 학생들이 제출한 글의 글씨 모양이나 문체, 이름이나 성적 등 학생들에 관한 정보로부터 영향을 받아 평가 결과가 달라지기도 한다(Freedman, 1979; Huot, 1990; Hake & Williams, 1981; White & Brunning, 2005; 박영목, 1999; 박영민, 2009a).

쓰기 평가를 체계적으로 관리하기 위해서는 어떠한 요인이 평가자에게 영향을 미치는지를 구체적으로 파악해야 한다. 그래야 그 영향 요인을 효과적으로 통제할 수 있으며, 이를 통해서 적절한 수준의 신뢰도를 확보하고 유지할 수 있기 때문이다. 적절한 수준의 쓰기 평가 신뢰도는 학교에서 쓰기 평가를 더욱 널리 활용할 수 있는 근거를 제공한다. 그러므로 쓰기 평가에 영향을 미치는 평가자 관련 요인을 탐색하는 것은 중요한 의미가 있다.

그러나 우리나라에서 학생 글을 평가하는 평가자들이 어떠한 요인으로부터 영향을 받고 있는지를 탐색한 연구는 찾기 어려운 실정이다. 쓰기 평가는 주관식 평가로서, 평가자의 개인적 성향이나 특성이 영향을 미칠 수밖에 없음에도 불구하고, 이를 분석한 연구는 거의 이루어지지 않았다. 쓰기 평가와 관련한 논의는 일반적으로 쓰기 평가 기준에 집중되어 온 경향이 있다. 평가자간 신뢰도를 확보하기 위해서는 세밀한 평가 기준의 설정이 중요하다고 보고 이를 주요 연구 주제로 다루어 왔기 때문이다(박영목, 1999, 2008a, 2008b). 최근에는 평가 기준을 넘어 평가 예시문을 활용한 평가자간 신뢰도 확보 방안을 탐색하는 논의가 전개되기도 했다(박영민, 2009b; 박영민 · 최숙기, 2010a, 2010b).

이러한 상황을 고려하여 이 연구에서는 쓰기 평가 영향 요인으로서 ‘평가자의 성별’과 ‘평가자의 학생 성별 인식’을 설정하고, 이러한 변인이 학생 글 평가에 어떠한 영향을 미치는지를 분석하고자 한다. 평가자 관련 요인 중에서 ‘평가자 성별’과 ‘학생 성별 인식’에 주목하는 이유는 외국의 선행 연구에서는 이 둘의 영향력이 매우 뚜렷하다고 밝히고 있음에도 불구하고 우리나라에서는 이에 대한 관심이 상대적으로 적었기 때문이다. 외국의 연구 결과와 같이, 우리나라의 경우에도 평가자들이 성별에 따라, 그리고 학생 성별 인식에 따라 평가 차이를 보이는지를 확인할 필요가 있다.

이 연구에서는 평가자 성별 및 학생 성별 인식에 따른 쓰기 평가의 차이를 파악하기 위하

여 예비 국어교사 69명을 평가자로 선정한 후, 이들로부터 학생 글 평가 자료를 수집·분석하였다. 학생 글의 유형은 설명문과 논설문 두 가지로 구분하였는데, 이를 바탕으로 하여 글의 유형에 따라 예비 국어교사의 쓰기 평가가 차이가 있는지를 부수적으로 탐색하고자 한다. 평가자를 예비 국어교사로 설정함으로써 이 연구의 결과는 예비 국어교사의 쓰기 평가와 관련한 정보를 얻을 수 있을 것으로 기대된다. 이는 예비 국어교사의 쓰기 평가 전문성 신장을 위한 프로그램 구성에도 긍정적으로 기여할 수 있을 것이다.

이 연구는 일부 대학의 예비 국어교사를 연구 대상으로 선정하였다는 점에서 결과의 일반화에는 한계가 있다. 또한, 남자 예비 국어교사의 수가 상대적으로 적어 성별 차이의 분석 결과를 일반화하는 데에도 한계가 따른다. 그러나 지금까지 관심 영역에서 배제되어 왔던 평가자 성별 및 학생 성별 인식에 따른 쓰기 평가의 차이를 분석했다는 점, 글의 유형에 따른 평가 차이를 고려하여 분석했다는 점, 예비 국어교사의 평가 전문성 신장에 필요한 정보를 제공한다는 점에서는 의의를 찾을 수 있다.

Ⅱ . 이론적 배경

쓰기 평가 과정과 결과에 영향을 미치는 평가자 관련 요인은 매우 다양하다. 배경지식, 쓰기 동기, 쓰기 신념, 인지 양식 등이 이러한 예에 속하는데, 이러한 요인들은 평가자를 매개로 하여 쓰기 평가의 결과를 달라지게 함으로써 신뢰도에 영향을 미친다. 그래서 쓰기 교육 전문가들은 이러한 평가자 관련 요인이 얼마나 통제되었는가를, 평가 결과를 얼마나 믿을 수 있는지를 판단하는 지표로 삼기도 한다. 학생 글을 평가하는 평가자는 국어교사이므로, 학생이나 학교를 배경으로 한 쓰기 평가 연구에서는 국어교사 요인으로 초점화하여 이 주제를 다루어왔다.

쓰기 평가에 영향을 미치는 평가자 요인으로 주목을 끌었던 것은 평가자의 쓰기 지식이다. 국어교사가 가지고 있는 쓰기 지식, 예를 들면 평가자가 가지고 있는 쓰기의 본질에 대한 지식, 문장에 대한 관점 및 지식, 좋은 글의 요건에 대한 지식, 화제나 주제에 대한 배경 지식은 학생이 쓴 글을 평가할 때 큰 영향을 미친다. 그래서 Runder(1992), 이성영(2005) 등은 평가자인 국어교사의 쓰기 지식이 학생 글을 평가할 때 복합적이면서도 메타적으로 영향을 미친다는 점을 지적하면서 국어교사의 지식 요인을 쓰기 평가 전문성 신장을 위한 주요 변인으로 다루어야 한다는 점을 강조하였다.

학생 글을 평가하는 평가자가 어떠한 유형의 인지 양식과 쓰기 신념을 가지고 있는가도 쓰기 평가 결과에 영향을 미치는 것으로 알려져 있다. 민태범(1991)은 장(場) 독립적 인지 양

식을 지닌 평가자들이 장 의존적 인지 양식을 지닌 평가자보다 학생 글을 더 엄격하게 평가한다는 점을 실증적으로 확인하였다. 교사 130명이 채점한 논술 답안을 분석한 결과, 장 독립적인 집단은 의존적인 집단보다 평균 점수가 더 낮았다. 이러한 결과는 평가자의 인지 양식이 쓰기 평가에 영향을 미치는 요인임을 알려준다.

평가자인 국어교사의 쓰기 신념은 교류 신념(transactional writing belief)과 전달 신념(transmissional writing belief)으로 구분되는데, 이 요인도 쓰기 평가에 영향을 미치는 것으로 알려져 있다(White & Brunning, 2005; 가은아, 2008). 현직 국어교사 56명을 대상으로 하여 쓰기 신념에 따라 국어교사의 쓰기 평가가 차이가 있는지를 분석한 가은아(2008)에 따르면, 총점 수준의 전체적인 차이는 발견되지 않았으나 전달 신념을 가지고 있는 국어교사들이 글 조직 요인에서 교류 신념을 지닌 국어교사들보다 더 높은 점수를 부여하는 것으로 나타났다. 이러한 결과를 통해 쓰기 신념 요인이 쓰기 평가에 영향을 미친다는 점을 확인할 수 있다.

평가자가 어떠한 성(性)을 지니고 있는가도 학생 글을 평가하는 데 영향을 미친다. 이 주제를 조사하고 분석한 Peterson & Kennedy(2006)는 남녀 교사 108명에게 학생들이 쓴 서사문과 논설문을 서술형으로 평가하게 한 후 어떤 차이가 있는지를 성별에 따라 분석하였다. 분석 결과, 교사 성별에 따라 학생 글에 대한 평가 결과가 차이가 있는 것으로 나타났다. 여교사는 학생 글에서 규칙과 조직에 대해 중점을 두고 평가했지만, 남교사는 학생 글에서 미적인 문체에 초점을 두고 평가하는 경향이 뚜렷했다. 성별에 따른 평가의 차이는 Barnes(1990)와 Roulis(1995)에서도 확인할 수 있다. 이 두 연구에 따르면, 학생 글을 서술형으로 평가할 때 여교사들은 일반적으로 남교사들보다 더 많은 양의 논평을 쓰는 것으로 나타났다.

이 연구에서 관심을 두고 있는 평가자의 학생 성별 인식도 쓰기 평가에 영향을 끼치는 것으로 알려져 있다. 평가자의 성별 인식이란, 평가자가 학생 글을 평가할 때 그 글을 작성한 학생을 남학생으로 인식하는가, 아니면 여학생으로 인식하는가를 말하는데, 이것이 쓰기 평가에 영향을 미친다는 것이다. Peterson(1998)은 교사가 학생 글을 평가할 때 여학생이 쓴 것으로 인식한 글보다는 남학생이 쓴 것으로 인식한 글에 대해 수정 사항을 더 많이 지적한다는 사실을 분석적으로 밝혀냈다. 교사들은 여학생들보다 남학생들에게 쓰기 평가를 통한 지도나 안내가 더 많이 요구된다고 생각하고, 남학생이 쓴 것으로 인식한 글에 대해 설명이나 제안과 같은 지도 내용을 더 많이 남긴다고 지적하였다.

Haswell & Haswell(1995)에서도 평가자들이 학생 성별을 아는가의 여부가 쓰기 평가에 영향을 미친다는 점이 확인된 바 있다. Peterson(1998)과 달리, Haswell & Haswell(1995)은 수량적 방법을 활용하여 평가자의 성별 인식뿐만 아니라 성별 지식도 함께 분석하였다. 평가자의 성별 지식은 평가자가 학생 성별이 남자인지 여자인지를 이미 알고 있는 것을 일컫는다. 분석 결과에 따르면, 평가자의 성별 인식과 성별 지식 모두가 쓰기 평가에 영향이 있는 것으로 나타났다. 5점 척도의 점수와, 긍정적 비평의 비율이 모두 변화하였다.

그런데 Haswell & Haswell(1995)에서 흥미로운 점은 평가자의 성별이 학생 성별과 상호작용 효과가 있음을 보고하고 있다는 점이다. 여자 평가자들은 글을 쓴 학생이 여학생이라는 것을 몰랐을 때보다 알았을 때 더 낮은 점수를 부여하였지만, 남자 평가자들은 여학생이라는 것을 알았을 때 더 높은 점수를 부여하였다. 글을 쓴 학생이 남학생일 경우에는 반대되는 현상을 보였다. 이러한 현상은 Peterson(1998)에서도 확인되었다. 교사는 자신의 성과 동일한 성을 지닌 학생이 쓴 것으로 인식한 글에 대해서 더 많은 비평을 남기는 경향이 강했다.

이러한 상호작용 효과는 동성(同性) 평가 절하(same-sex depreciation) 현상으로 불린다(Peterson, 1998). 동성 평가 절하 현상이란, 평가자는 자신의 성과 동일한 성을 가진 학생이 쓴 것으로 인식하는(알고 있는) 글에 대해 더 엄격하게, 더 부정적으로 평가하는 것을 일컫는다. 즉, 학교에서 쓰기 평가를 할 경우, 남자 교사는 남학생이 쓴 것으로 인식한 글을 더 엄격하게 평가하고, 여자 교사는 여학생이 쓴 것으로 인식한 글에 대해서 더 엄격하게 평가하는 현상을 말한다. 동성 평가 절하의 경향은 다른 선행 연구에서도 이미 확인된 바 있다. Etaugh, Houtler, & Ptasnik(1988), Haswell & Haswell(1996)에 따르면, 남자 교사들은 여학생이 쓴 글보다 남학생 글에서 더 많은 수정 의견을 제시하고 비판을 하였으며, 남학생이 썼다고 생각하는 글보다는 여학생이 쓴 것으로 보이는 글을 더 칭찬하는 경향이 있었다. 이와 달리, 여교사들은 여학생이 쓴 것으로 인식된 글보다는 남학생이 쓴 것으로 보이는 글을 더 많이 칭찬했다. 이러한 선행 연구들은 교사가 가지고 있는 학생 성별 인식이 쓰기 평가에 영향을 미친다는 점을 잘 보여준다.

Ⅲ. 연구 방법

1. 연구 대상

예비 국어교사의 성별 및 학생 성별 인식에 따른 쓰기 평가의 차이 분석에 필요한 자료를 얻기 위하여 사범대학 국어교육과 및 대학원 국어교육과에서 양성 교육을 받는 학생 69명을 연구 대상으로 선정하였다. 예비 국어교사 69명에게 중학생이 쓴 설명문과 논설문을 평가하게 하고 그 결과를 수집하였다. 평가 결과를 수집할 때, 평가자로 참여한 예비 국어교사들의 성별과 학생 성별 인식을 조사하였다. 자료를 조사한 연구 대상의 규모 및 분포를 정리하면 <표 1>과 같다.

〈표 1〉 연구 대상의 규모 및 분포

성별 표집 집단	남	여	계
A대학 사범대 국어교육과	10	12	22
B대학 사범대 국어교육과	11	18	29
C대학 대학원 국어교육과	3	15	18
계	24	45	69

전체 69명 중 남학생은 24명(34.8%), 여학생 45명(65.2%)이었다. A대학이나 B대학과는 달리 C대학은 남학생의 비율이 상대적으로 더 낮았다.

2. 검사 도구

평가자인 예비 국어교사들에게 제시한 글은 중학생이 작성한 설명문 35편과 논설문 35편이었다. 설명문은 ‘자신이 관심 있거나 좋아하는 책, 영화, 운동(스포츠), 취미 활동 등을 학급 친구들에게 소개하고 이를 좋아하게 된 이유를 설명하라.’는 과제에 따라 작성되었으며, 논설문은 ‘학교에서 학생들의 복장 및 두발 규제에 대한 자신의 주장과 이에 대한 근거를 서술하라.’는 과제에 따라 작성되었다.

필체 등 평가에 영향을 미치는 주관적 요인을 통제하기 위하여 학생 글은 워드 프로세서로 입력하여 제시하였다. 학생 글을 워드 프로세서로 입력할 때, 띄어 쓴 것인지 아닌지가 모호한 경우에는 띄어 쓴 것으로 입력하였으며, 띄어쓰기가 명백한 경우에는 오류라 하더라도 원문대로 입력하였다. 그 밖의 맞춤법도 오류가 있더라도 원문대로 입력하였다. 그 이유는 이와 관련된 평가 기준이 설정되어 있기 때문이다.

예비 국어교사들에게는 학생 글과 함께 쓰기 평가 기준표 및 채점표를 함께 제공하였다. 쓰기 평가 기준표는 Spandel & Culham(1996)이 제시한 총체적 평가 기준을 중학생의 쓰기 평가에 적합하도록 일부 수정하여 사용하였다. 평가 요인은 내용, 조직, 표현, 형식 및 어법, 단어 선택으로 구성되었다. 평가 기준표에는 각 평가 요인별로 1, 3, 5점에 해당하는 평가 기준을 제시하였으나, 채점을 할 때에는 1점에서 5점 사이에서 자유롭게 점수를 부여할 수 있도록 하였다.

예비 국어교사의 성별은 자기 자신의 성별을 표시하게 하여 정보를 수집하였으며, 학생 성별 인식은 설명문과 논설문을 평가할 때 어떤 성별을 가진 학생이 쓴 글로 생각되는지를 표시하게 하여 정보를 수집하였다.

3. 연구 절차 및 분석 도구

쓰기 평가 자료를 수집하기 위해 예비 국어교사들에게 학생 설명문 35편과 학생 논설문 35편, 평가 기준표 및 채점표를 제공하였다. 평가자로 선정된 예비 국어교사들은 다른 예비 국어교사들과 평가자 협의를 거치지 않고 개별적으로 평가하였다. 평가자 협의를 실시하지 않은 이유는 예비 국어교사들의 학생 성별 인식의 영향을 다른 요인의 개입 없이 이끌어내려는 목적 때문이었다. 평가자 협의를 거칠 경우, 예비 국어교사들이 처음 형성한 학생 성별 인식이 소거되거나 변화될 수 있다.

예비 국어교사들의 평가 결과는, A대학과 B대학은 2009년 5월 1일부터 5월 20일까지, C대학은 4월 1일부터 4월 20일까지 수집하였다. 예비 국어교사들은 평가 기준에 따라 5점 척도로 내용, 조직, 표현, 형식 및 어법, 단어 선택의 점수를 각각 채점하여 총점을 산출하였으며, 예비 국어교사 69명이 채점한 점수의 평균을 학생 글 각각의 개별 점수로 새롭게 지정한 후 통계적으로 처리하였다. 설명문 총점에 대한 평가자간 신뢰도는 Cronbach α .962로 나타났으며, 논설문에 대한 평가자간 신뢰도는 Cronbach α .986으로 나타났다.

예비 국어교사의 성별 및 학생 성별 인식에 따른 쓰기 평가 차이 분석에는 독립 집단 t 검증, 일원 분산 분석의 방법이 쓰였다. 이러한 분석에는 SPSS for Windows 15.0 판을 활용하였다.

IV. 연구 결과 및 논의

1. 평가 결과에 대한 기술 통계

이 연구에서는 예비 국어교사가 평가한 점수의 평균을 각 학생 글의 개별 점수로 새롭게 지정하여 분석하였다. 예를 들면, 학생 설명문 01번을 69명의 예비 국어교사가 평가하였는데, 이 69명이 채점한 점수의 평균을 학생 설명문 01번의 점수로 새롭게 지정하여 분석하였다. 학생 글 분석 자료에 소수점 이하의 점수가 나타나는 것은 이러한 이유 때문이다. 학생들이 작성한 설명문과 논설문에 새롭게 지정된 점수를 제시하면 <표 2>와 같다.

〈표 2〉 새롭게 지정된 학생 글의 점수 현황

설명문 글 번호	지정 점수	설명문 글 번호	지정 점수	논설문 글 번호	지정 점수	논설문 글 번호	지정 점수
01	12.83	19	13.36	01	15.09	19	14.83
02	10.23	20	19.38	02	17.17	20	14.94
03	18.14	21	17.59	03	14.36	21	17.93
04	8.29	22	15.10	04	16.41	22	10.61
05	14.23	23	15.77	05	13.42	23	15.43
06	12.03	24	10.12	06	10.28	24	15.00
07	11.65	25	15.88	07	10.80	25	11.75
08	15.28	26	11.28	08	13.90	26	13.84
09	19.74	27	16.80	09	7.57	27	12.94
10	12.29	28	14.55	10	17.17	28	12.83
11	13.57	29	15.80	11	18.41	29	10.51
12	11.45	30	14.38	12	20.62	30	10.61
13	12.19	31	12.65	13	13.17	31	10.67
14	16.35	32	14.94	14	12.91	32	14.65
15	18.49	33	14.91	15	19.67	33	15.68
16	14.88	34	15.75	16	12.49	34	11.25
17	13.42	35	13.29	17	15.77	35	14.83
18	13.16	-	-	18	11.16	-	-

〈표 2〉의 점수를 바탕으로 하여 학생 설명문 및 학생 논설문 평가 결과에 대한 기술 통계를 구할 수 있다. 결과는 〈표 3〉과 같다. 기술 통계는 설명문과 논설문 평가에 드러난 예비 국어교사의 일반적인 특성을 파악하는 데 도움을 준다.

〈표 3〉 설명문과 논설문 평가에 대한 기술 통계

글 유형	최솟값	최댓값	평균	표준편차
설명문	8.13	20.08	14.27	2.64
논설문	7.36	21.29	14.02	2.95

〈표 3〉에 따르면, 예비 국어교사들은 중학생들이 작성한 설명문에 평균 14.27점을 부여하였고, 논설문에는 평균 14.02점을 부여하였다. 이 연구에서 적용한 평가 척도는 최저 5점, 최고 25점이므로 척도 평균은 15.0점이라고 할 수 있는데, 〈표 3〉의 두 평균은 이 점수를 다소 하회한다. 이는 학생 설명문과 논설문이 평가 하위 요인에서 평균적으로 3점을 얻지 못했음을 뜻한다. 그러므로 예비 국어교사들은 학생 글을 척도 평균보다 낮게, 다소 엄격하게 평가하고 있다고 할 수 있다.

이러한 결과는 평가 경험이 없거나 적은 평가자들이 엄격성이 높다는 점을 지적해 온 선

행 연구 결과와 부합한다. Weigle(1998)은 Rasch 모형을 활용하여 쓰기 평가 경험이 없는 평가자들이 경험이 많은 평가자보다 훨씬 더 엄격하게 평가한다는 점을 분석한 바 있고, 박영민·최숙기(2009)도 평균 비교를 통해 평가 경험이 적은 예비 국어교사가 경험이 많은 현직 국어교사보다 더 엄격하게 평가하는 경향이 있음을 보고한 바 있다. 영어 평가에서도 평가자가 원어민인가 아닌가에 관계없이, 평가 경험이 적을 때 훨씬 더 엄격하게 평가한다는 사실이 분석된 바 있다(신동일, 2001; 최연희, 2002).

또한, <표 2>에 따르면 예비 국어교사들은 평가 척도의 범위를 평가 기준에서 제시한 것보다 좁게 활용하고 있다. 평가 척도는 최저 5점에서 최대 25점으로 구성되어 있지만(점수 구간 20점), 예비 국어교사들은 약 7점에서 21점의 범위로 척도를 적용하고 있다(점수 구간 14점). 이러한 경향도 앞에서 밝힌 선행 연구의 결과와 유사하다. 박영민·최숙기(2009)에서 조사된 예비 국어교사들은 최저 8.44점에서 최고 20.84점의 범위로 평가하였다.

2. 성별에 따른 평가 차이 분석

선행 연구에 따르면 평가자 성별은 학생 글 평가에 영향을 미친다. 학생 글을 질적으로 평가할 때 여자 국어교사는 미적 문체에 중점을 두는 남자 국어교사와 달리, 규칙과 조직을 강조하는 경향이 있으며, 학생 글에 대해 평가한 양도 더 많았다(Peterson, 1998; Peterson & Kennedy, 2006). 이러한 영향이 예비 국어교사의 쓰기 평가에도 존재하는지를 파악하기 위하여 예비 국어교사를 남녀 두 집단으로 구분한 후 그 차이를 독립 집단 t 검증을 적용하여 분석하였다. Levene의 등분산 검증에서 설명문은 $F=.047$, $p=.829$ 로, 논설문은 $F=.009$, $p=.927$ 로 나타나 설명문과 논설문 모두 등분산이 가정되었다. 분석 결과는 <표 4>와 같다.

<표 4> 예비 국어교사의 성별에 따른 평가 결과의 차이 분석

글 유형	성별	사례수	최솟값	최댓값	평균	표준편차	df	t	p
설명문	남	35	8.13	20.08	14.24	2.63	68	-.074	.941
	여	35	8.38	19.56	14.29	2.68			
논설문	남	35	7.96	21.29	14.24	3.02	68	.624	.535
	여	35	7.36	20.27	13.80	2.90			

<표 4>에 따르면, 설명문이든 논설문이든 예비 국어교사의 성별에 따른 평가 결과의 유의한 차이는 발견되지 않았다. 남자 예비 국어교사이든 여자 예비 국어교사이든 학생 설명문과 논설문을 유사한 수준으로 평가하고 있다고 할 수 있다. 성별에 따른 쓰기 평가 차이가 발견되지 않는 것은 주요 선행 연구의 결과인 Peterson(1998), Peterson & Kennedy(2006) 등의

결과와 일치하지 않는다.

이러한 상반된 결과를 얻게 된 이유는 특히 Peterson(1998), Peterson & Kennedy(2006) 등 국외 선행 연구에서 적용한 연구 방법이 이 연구에서 적용한 연구 방법과 다르기 때문인 것으로 해석된다. 이들 선행 연구에서는 비평이나 논평과 같은 국어교사의 ‘질적 평가’를 바탕으로 하여 평가자 성별이 쓰기 평가에 영향을 미친다고 보고한 것이다. 선행 연구처럼 국어교사가 학생 글을 세밀하게 읽고 비평이나 반응을 기록하는 질적 평가를 적용할 때에는 읽기나 쓰기와 관련된 성별 특성이 영향을 미칠 수 있다. 여성은 남성에 비해 읽기 능력과 쓰기 능력이 월등히 우수한데(Berninger & Fuller, 1992; Engelhard et al., 1994; Pajares & Valiante, 1997), 이러한 특성이 질적 평가 장면에서도 영향을 미치는 것으로 보인다.

그러나 이 연구처럼, 학생 글에 대해 ‘질적 판단’이 아니라 ‘양적 판단’을 적용할 경우에는 결과가 다소 달라질 수 있다. 성별 차이가 발견되지 않은 이 연구 결과는, 수량 평가를 적용하여 성별 차이를 살폈던 박영민·최숙기(2009)의 결과와 유사하기 때문이다. 박영민·최숙기(2009)에서는 현직 국어교사 49명과 예비 국어교사 24명을 대상으로 하여 쓰기 평가 결과의 성별 차이를 분석하였는데, 현직 국어교사와 예비 국어교사 모두 통계적으로 유의한 성별 차이를 보이지 않았다. 선행 연구가 충분하지 않아 제한적이기는 하지만, 현재까지의 결과를 토대로 하면 학생 글을 양적으로 평가할 때에는 평가자의 성별 차이가 잘 드러나지 않는다고 할 수 있다.

한편, <표 4>에서 주목을 끄는 것 중 하나는 논설문에서 남자 예비 국어교사들의 표준편차가 크게 나타나고 있다는 점이다. 표준편차가 크다는 것은 평가 점수의 산포도가 크다는 뜻이므로, 학생 논설문에 대한 남자 예비 국어교사들의 평가 결과의 일치도가 상대적으로 낮다고 볼 수 있기 때문이다. 설명문에서 남자 예비국어교사와 여자 국어교사의 표준편차가 거의 일치하고 있는 것과는 대조적이다. 이러한 결과는 남자 예비 국어교사들이 학생 논설문을 평가할 때에는 평가자간 신뢰도가 떨어질 수 있음을 보여준다.

설명문 평가에서는 설명 내용을 확인하기가 쉽고 설명 내용에 따른 조직을 파악하기가 쉬워 비교적 평가 결과가 일치할 가능성이 높지만, 논설문 평가에서는 학생들이 주장하는 내용이 산발적이고 근거도 다양해서 평가 결과가 일치할 가능성이 낮다. 남자 예비 국어교사들은 논설문을 평가할 때 이러한 유형적 특징에 약점을 보이고 있으므로 이를 극복할 수 있도록 돕는 노력을 강구해야 한다. 논설문을 중심으로 하여 평가 전문가들의 평가 관점 및 평가 결과를 비교·대조하는 평가자 훈련, 논설문에서 주요 요소와 주요 내용을 발견하는 능력을 기르기 위한 협의 활동과 교육 활동, 논설문 평가의 기준을 설정 및 내면화를 위한 교육 활동 등을 예비 국어교사들에게 적용할 수 있을 것이다.

3. 학생 성별 인식에 따른 쓰기 평가 차이

앞에서는 예비 국어교사의 성별이 학생 글 평가에 영향을 미치지 않는다는 점을 확인하였다. 그렇다면 예비 국어교사가 학생 성별을 남학생으로 인식하는지, 여학생으로 인식하는지에 따라 쓰기 평가 결과에 차이가 있는지를 분석할 필요가 있다. Haswell & Haswell(1995), Peterson(1998)의 주장과 같이, 예비 국어교사들이 글을 쓴 학생의 성별을 어떻게 인식하는가에 따라 쓰기 평가가 달라진다면 예비 국어교사의 학생 성별 인식을 쓰기 평가에 영향을 미치는 요인으로 확정할 수 있기 때문이다. 학생 성별 인식에 따른 쓰기 평가의 차이를 분석하기 위하여 예비 국어교사들이 설명문과 논설문을 쓴 학생의 성별을 각각 어떻게 인식하였는지를 정리하였다. 정리한 결과는 <표 5>와 같다.

<표 5> 학생 글에 대한 성별 인식 빈도

설명문						논설문					
글 번 호	남학생 글로 인식	여학생 글로 인식	글 번 호	남학생 글로 인식	여학생 글로 인식	글 번 호	남학생 글로 인식	여학생 글로 인식	글 번 호	남학생 글로 인식	여학생 글로 인식
01	53	16	19	65	4	01	12	57	19	59	10
02	68	1	20	66	3	02	10	59	20	49	20
03	68	1	21	67	2	03	59	10	21	50	19
04	66	3	22	50	19	04	37	32	22	42	27
05	2	67	23	31	38	05	31	38	23	31	38
06	66	3	24	65	4	06	44	25	24	44	25
07	52	17	25	45	24	07	56	13	25	52	17
08	28	41	26	61	8	08	33	36	26	33	36
09	23	46	27	4	65	09	35	34	27	41	28
10	8	61	28	50	19	10	29	40	28	51	18
11	55	14	29	41	28	11	39	30	29	44	25
12	13	56	30	35	34	12	40	29	30	56	13
13	27	42	31	5	64	13	23	46	31	47	22
14	5	64	32	40	29	14	56	13	32	43	26
15	8	61	33	67	2	15	37	32	33	19	50
16	11	57	34	36	33	16	30	39	34	60	9
17	26	43	35	58	11	17	48	21	35	27	42
18	42	27	-	-	-	18	60	9	-	-	-

〈표 5〉를 바탕으로 하여, 종다수의 방법에 따라 남학생이 쓴 것으로 인식된 글과 여학생이 쓴 것으로 인식된 글을 구분할 수 있다. 이에 따르면, 설명문에서 남학생이 쓴 것으로 인식된 글은 22건이고 여학생이 쓴 것으로 인식된 글은 13건이다. 논설문에서는 24건이 남학생 쓴 것으로 인식되었고, 11건이 여학생이 쓴 것으로 인식되었다. 따라서 전체적으로는 남학생 글로 인식된 것이 46건, 여학생 글로 인식된 것이 24건이다. 전체적으로 볼 때, 남학생 글로 인식된 예가 더 많다고 할 수 있다.

〈표 5〉에서 02, 03, 04, 05, 06번 설명문은 성별 인식이 매우 명료하다. 남학생이든 여학생이든, 예비 국어교사들이 한 쪽 성으로 인식하는 경향이 절대적으로 우세하다. 이에 비해 30번 설명문은 성별 인식이 모호하다. 남학생 글로 인식한 수와 여학생 글로 인식한 수가 매우 유사하기 때문이다. 또한, 04, 05, 08, 09번 논설문도 성별 인식이 모호하다. 남학생이 쓴 것으로 인식하는 예비 국어교사의 수와, 여학생이 쓴 것으로 인식하는 예비 국어교사의 수가 비슷하다. 이러한 글은 성별 특성이 잘 드러나지 않는 중성적인 글로 분류할 수 있다.

〈표 5〉의 성별 인식 분포를 살펴보면, 논설문보다는 설명문에서 예비 국어교사의 성별 인식이 비교적 명확하게 드러나고 있음을 발견할 수 있다. 성별 인식에 대한 차이가 설명문에서 더 뚜렷하다. 그러므로 글의 유형에 따르면, 학생 글에 대한 성별 인식은 논설문보다 설명문에서 뚜렷하게 파악된다고 할 수 있다. 그 이유는, 설명문의 경우, 설명 대상 선정에서 성별적 특징, 예를 들면 남학생들은 운동이나 컴퓨터 게임과 같은 취미를 선택하고, 여학생들은 영화나 책을 선택하는 것과 같은 특징이 선명하게 드러나기 때문이다.

〈표 5〉에 정리된 결과를 바탕으로 하여, 예비 국어교사가 학생 성별을 무엇으로 보는지에 따라 쓰기 평가 결과가 차이가 있는지를 분석할 수 있다. 예비 국어교사의 학생 성별 인식 분석을 위해 집단을 넷으로 구분하였다. 남자 예비 국어교사 중에서 학생 글을 남학생이 쓴 것으로 인식한 집단은 1, 여학생이 쓴 것으로 인식한 집단은 2로, 여자 예비 국어교사 중에서 학생 글을 남학생이 쓴 것으로 인식한 집단은 3, 여학생이 쓴 것으로 인식한 집단은 4로 구분하였다. 이러한 집단 구분에 따라 학생 설명문과 학생 논설문에 대한 평가 결과의 기술 통계를 구하면 〈표 6〉과 같다.

〈표 6〉 성별 인식 집단별 기술통계

글 유형	집단	사례수	평균	표준 편차	평균 95% 신뢰구간		최솟값	최댓값
					하한값	상한값		
설명문	1	35	13.87	3.53	12.66	15.08	.00	19.35
	2	35	12.39	5.23	10.60	14.19	.00	20.47
	3	35	13.83	2.71	12.90	14.76	7.50	19.67
	4	35	15.18	2.73	14.24	16.12	10.25	23.00
	합계	140	13.82	3.78	13.19	14.45	.00	23.00
논설문	1	35	13.90	2.81	12.94	14.87	7.88	20.50
	2	35	14.84	3.43	13.65	16.02	8.14	22.88
	3	35	13.28	2.73	12.34	14.22	6.33	19.21
	4	35	14.54	3.19	13.44	15.63	8.04	21.48
	합계	140	14.14	3.08	13.62	14.65	6.33	22.88

〈표 6〉을 검토해 보면 두 가지 특징적인 현상을 파악할 수 있다. 하나는 설명문이든 논설문이든 집단 2, 즉 남자 예비 국어교사들이 글을 쓴 학생을 여학생으로 인식할 때 표준편차가 매우 크다는 점이고, 다른 하나는 설명문에서 여자 예비 국어교사들이 학생 성별을 여학생으로 인식한 집단에서 평균 점수가 높게 분포한다는 점이다.

학생 성별을 여학생으로 인식한 남자 예비 국어교사들의 평가 점수는 표준편차가 매우 크다. 표준편차가 크다 보니 95% 수준에서 추정한 모평균의 신뢰구간도 가장 길게 나타나고 있다. 남자 예비 국어교사들은 학생 성별을 여학생으로 인식했을 때 평가 결과의 산포도가 커지므로, 남자 예비 국어교사들에게는 평가자간 신뢰도를 확보할 수 있도록 돕는 양성 교육이 필요하다. 예를 들면, 남학생 글과 여학생 글을 대하는 평가 전문가의 관점과 태도를 익히는 평가자 훈련, 이러한 편향이 있을 수 있음을 스스로 인지하도록 돕는 협의 활동, 또한 이러한 편향 가능성을 이해하고 대책을 수립하는 협의 활동 및 토론 활동을 적용할 수 있을 것이다. 이러한 노력을 기울이지 않는다면, 남자 예비 국어교사는 쓰기 평가의 전문성을 획득하고 발휘하는 데 어려움을 겪을 수 있다. 남자 예비 국어교사는 여학생이 쓴 것으로 인식된 글을 평가할 때뿐만 아니라, 실제로 여학생이 썼다는 사실을 알고 평가할 때 편향을 보일 것으로 예상되기 때문이다.

한편, 여자 예비 국어교사들은 설명문을 쓴 학생이 여학생이라고 인식했을 때 평가 점수가 높았으므로 분산 분석을 통해 그 차이의 유의성을 확인할 필요가 있다. 이 차이는 집단 간 차이를 반영하고 있기 때문이다. 분산 분석을 통해서 예비 국어교사의 학생 성별 인식이 쓰기 평가에 영향을 미치는지를 판단할 수 있다. 일원 분산 분석을 위하여 Levene의 등분산

검증을 실시한 결과, 논설문은 등분산이 가정되었으나($F=.857$, $p=.465$), 설명문은 등분산이 가정되지 않았다($F=3.145$, $p=.027$). 분산 분석 결과는 <표 7>과 같다.

<표 7> 학생 성별 인식에 대한 분산분석

변인	제곱합	df	평균제곱	F	p
설명문	집단-간	3	45.397	3.323	.022
	집단-내	136	13.663		
	합계	139			
논설문	집단-간	3	16.796	1.793	.151
	집단-내	136	9.367		
	합계	139			

<표 7>에 따르면, 집단 간 평균이 유사했던 논설문에서는 차이가 발견되지 않았으나, 설명문에서는 집단 간 차이가 확인되었다. 이러한 차이가 어떤 집단 사이에서 나타난 것인지를 확인하기 위하여 사후 분석을 실시하였다. 사후 분석은 Tukey HSD를 적용하였다. 사후 분석 결과는 <표 8>과 같다.

<표 8> 사후 분석

글 유형	(I) 집단	(J) 집단	평균 차(I-J)	표준오차	p
설명문	1	2	1.47629	.88360	.343
		3	.04029	.88360	1.000
		4	-1.31171	.88360	.450
	2	3	-1.43600	.88360	.368
		4	-2.78800*	.88360	.011
	3	4	-1.35200	.88360	.422

사후 분석 결과, 집단 2와 집단 4에서 차이가 발견되었다. 즉, 남자 예비 국어교사 중 학생 성별을 여학생으로 인식한 집단과, 여자 예비 국어교사 중 학생 성별을 여학생으로 인식한 집단 사이에서 통계적으로 유의한 차이가 발견되었다. 여자 예비 국어교사가 설명문을 작성한 학생의 성별을 여학생으로 인식했을 때 평가 점수가 통계적으로 유의미하게 높았다.

이러한 결과는 예비 국어교사의 학생 성별 인식이 쓰기 평가에 영향을 미칠 것이라는 가정을 지지하지 않는다. 논설문에서는 학생 성별 인식의 영향이 발견되지 않았을 뿐만 아니라, 차이가 발견된 설명문의 경우에도 일부 집단에서만 차이를 보이고 있기 때문이다. 특히,

설명문을 쓴 학생을 여학생으로 인식한 남자 예비 국어교사들보다 그 학생을 여학생으로 인식한 여자 예비 국어교사들의 평가가 더 높았다는 점은 동성 평가 절하 현상을 지적한 선행 연구 결과와 상반된다는 점에서 주목된다.

동성 평가 절하 현상은 평가자의 성별이 학생 성별 인식과 교차하면서 상호작용 효과를 낸다는 점에서 중요한 특징이 있다. 남자 평가자는 남학생이 쓴 것으로 인식할 때, 여자 평가자는 여학생이 쓴 것으로 인식할 때, 평가 점수를 낮게 부여하는 현상을 일컫는 개념이기 때문이다. 그런데 이 연구 결과는 여자 예비 국어교사들이 여학생, 즉 동성(同性)의 학생이 쓴 것으로 인식한 글에 대해서 상대적으로 더 높은 점수를 부여함으로써, 선행 연구가 지적한 동성 평가 절하 현상과 반대되는 현상을 보였다. 이러한 결과를 통해 볼 때, 예비 국어교사에게는 동성 평가 절하 현상이 발견되지 않으며, 따라서 학생 성별 인식의 영향도 일반화하기 어렵다고 판단할 수 있다.

예비 국어교사에게서 학생 성별 인식 효과가 없는 것으로 나타난 이유는 세 가지의 원인이 있었던 것으로 해석할 수 있다. 첫째, 동성 평가 절하 현상을 보고한 선행 연구와 이 연구의 방법적 차이 때문인 것으로 보인다. 동성 평가 절하 현상을 지적한 선행 연구는 학생 글에 대한 질적 평가를 바탕으로 하였으나, 이 연구에서는 양적 평가를 바탕으로 하여 분석이 이루어졌다. 앞서서도 지적했던 것처럼, 이러한 방법의 차이로 인해 선행 연구와 다른 결과를 얻게 된 것으로 보인다.

둘째, 예비 국어교사들이 학생과 직접적으로 교류하고 상호작용한 경험이 없었기 때문인 것으로 해석된다. 예비 국어교사들은 설명문과 논설문을 쓴 학생 성별을, 경험과 분리된 채 인지적 판단으로만 구별함으로써 그 영향력이 상대적으로 미미했던 것으로 보인다. 동성 평가 절하 현상을 보인 평가자들은 학생과의 교류가 빈번했고, 학생들을 실제로 지도하는 교사들이었으며, 글쓰기 지도를 통해 학생과 상호작용하는 경험이 많은 국어교사들이었다는 점이 이러한 해석을 뒷받침한다.

셋째, 여학생이 쓴 것으로 인식된 글은 실제로 완성도가 높은, 잘 쓴 글이었기 때문에 평가 점수가 높았던 것으로 보인다. 여학생이 쓴 것으로 인식된 글은 대체로 글씨 모양이나 단락 구분 등 외형 조건이 깔끔하였을 뿐만 아니라 글 내용의 질적 수준도 우수하였다. 그러므로 여학생이 쓴 것으로 인식된 글은 본래 우수한 글이었기 때문에 여자 예비 국어교사들로부터 높은 점수를 얻을 수 있었고, 그 결과, 동성 평가 절하 현상이 나타나지 않았던 것으로 해석된다. 이러한 해석은 오히려 남자 예비 국어교사들이 잘 쓴 글을 오히려 짚어내지 못했을 수 있음을 뜻하는 것이어서 주목된다. 남자 예비 국어교사들은 여자 예비 국어교사보다 우수한 글을 읽어내는 능력이 떨어지고, 따라서 이들에게는 양성 과정에서 쓰기 평가 전문성을 기르기 위한 평가자 훈련 프로그램이 적용되어야 한다는 뜻을 함축하고 있기 때문이다.

V. 결론

외국의 여러 선행 연구에서는 평가자의 성별과 평가자의 학생 성별 인식이 쓰기 평가에 영향을 미친다는 점을 지적해 왔다. 평가자의 성별과 평가자의 학생 성별 인식은 쓰기 평가에 영향을 미침으로써 쓰기 평가의 신뢰도를 떨어뜨리는 요인으로 지적되기도 했다. 그런데 국내에서는 평가자의 성별과 학생 성별 인식이 어떠한 영향을 미치는지를 검토한 연구가 거의 없다. 이에 이 연구에서는 선행 연구의 주장이 우리나라 예비 국어교사의 경우에도 존재하는지를 밝히고자 하였다. 이 연구에서는 설명문과 논설문을 적용하여 글의 유형에 따라 쓰기 평가가 어떻게 달라지는지를 동시에 파악하고자 하였으며, 예비 국어교사를 평가자로 선정함으로써 쓰기 평가 전문성을 기르는 데 필요한 정보도 확보하고자 하였다.

이 연구에서 얻은 결과는 다음과 같다. 첫째, 예비 국어교사들은 설명문과 논설문 평가에서 각각 평균 14.27점, 14.02점을 부여하여 척도의 실질적 평균인 15점보다는 낮게 평가하였다. 평가 척도는 최저 5점, 최고 25점으로 구성되어 있으나, 예비 국어교사들은 이 범위를 변별력 있게 충분히 활용하지 않고 범위를 좁혀 학생 글을 평가하였다.

둘째, 예비 국어교사의 성별에 따른 쓰기 평가의 차이는 설명문과 논설문 모두에서 발견되지 않았다. 남자 예비 국어교사와 여자 예비 국어교사의 평가는 유사했으며, 선행 연구에서 지적된 성별 차이는 발견되지 않았다. 그러나 논설문을 평가한 남자 예비 국어교사들은 상대적으로 큰 편차를 보여 평가 관점을 일치시킬 수 있는 노력이 필요하다는 점을 일깨워 주었다. 설명문에 대한 평가에서는 남녀 예비 국어교사들 모두 표준편차가 매우 유사했다.

셋째, 예비 국어교사의 학생 성별 인식에 따른 평가 결과의 차이도 확인되지 않았다. 선행 연구에서는 학생 성별 인식에 따른 평가 차이를 지적했지만, 이 연구에서는 이에 부합하는 결과를 얻지 못했다. 동성 평가 절하 현상도 발견되지 않았다. 설명문 평가에서 여자 예비 국어교사는 학생 성별을 여학생으로 인식할 때, 오히려 더 높은 점수를 부여하는 것으로 나타났다. 그러나 남자 예비 국어교사의 경우, 학생 성별을 여학생으로 인식할 때 표준편차가 크게 나타나 평가자간 신뢰도를 확보할 수 있는 방안을 강구해야 한다는 점을 보여주었다.

외국의 선행 연구에서는 평가자의 성별과 평가자의 학생 성별 인식을 쓰기 평가의 영향 요인으로 지적해 왔지만, 이 연구에서는 이를 지지하는 결과를 얻지 못했다. 선행 연구와 달리 평가자의 성별 특성이 적게 개입하는 수량적 평가 방법을 적용했다는 점, 학생들과의 실제적인 교류나 의사소통이 없는 예비 국어교사를 평가자로 선정했다는 점이 선행 연구와 다른 결과를 얻게 된 데에 일부 영향을 미친 것으로 보인다. 학생과의 교류, 학생과의 의사소통을 경험한 현직 국어교사를 대상으로 한다면, 이 연구와는 다른 결과를 얻을 수도 있다. 현직 국어교사를 대상으로 하는 연구가 후속적으로 이어져야 할 것이다.

이 연구에서는 평가자의 성별 및 학생 성별 인식에 따른 평가 차이는 확인하지 못했지만, 대체적으로 남자 예비 국어교사들이 여자 예비 국어교사들에 비해 쓰기 평가의 전문성이 다소 떨어지는 경향이 있음을 발견할 수 있었다. 그러므로 예비 남자 국어교사들이 쓰기 평가 전문성을 효과적으로 획득하고 발현할 수 있도록 도움을 제공할 필요가 있다. 평가 전문가들의 평가 관점을 익히고 훈련하는 프로그램, 글의 주요 요소와 주요 내용을 발견하는 능력을 기르는 프로그램, 평가의 기준 설정 및 내면화를 위한 프로그램, 성별 인식에 따른 편향을 인지하고 통제할 수 있도록 돕는 프로그램 등을 이러한 도움의 예로 꼽을 수 있다.

참 고 문 헌

- 가은아(2008). 국어교사의 쓰기 효능감 및 쓰기 신념이 학생 쓰기 평가에 미치는 영향. 석사 학위 논문, 한국교원대학교.
- 민태범(1991). 채점자의 인지양식이 논술고사 채점에 미치는 영향. 석사학위 논문, 한남대학교.
- 박영목(1999). 작문 능력 평가 방법과 절차. **국어교육**, 99, 1-29.
- 박영목(2008a). **작문교육론**. 역락.
- 박영목(2008b). 쓰기 평가 연구의 주요 과제. **작문연구**, 6, 9-37.
- 박영민(2009a). 논술 평가자의 주관적 경험 분석. **우리어문연구**, 33, 493-523.
- 박영민(2009b). 평가 예시문을 활용한 쓰기 평가 개선 방안. **청람어문교육**, 39, 111-133.
- 박영민 · 최숙기(2009). 현직 국어교사와 예비 국어교사의 쓰기 평가 비교 연구. **교육과정평가 연구**, 12(1), 123-143.
- 박영민 · 최숙기(2010a). 국어교사의 설명문 평가에 대한 모평균 추정과 평가 예시문 선정. **우리어문연구**, 36, 293-326.
- 박영민 · 최숙기(2010b). 중학생 논설문 모평균추정과 평가예시문 선정. **국어교육**, 131, 437-461.
- 신동일(2001). 채점 경향 분석을 위한 Rasch 측정 모형 적용 연구. *Foreign Languages Education*, 8(1), 249-272.
- 이성영(2005). 국어과 교사의 쓰기 영역 평가 전문성 기준과 모형. **국어교육**, 117, 363-376.
- 최연희(2002). 채점자 훈련이 영어 작문 채점에 미치는 효과에 대한 연구: FACETS 분석을 통한 신뢰도 변화 분석. *Journal of the Applied Linguistics of Korea*, 18(1), 257-292.
- Barnes, L. L. (1990). Gender bias in teacher's written comments, In S. L. Gabriel & I. Smithson(Eds.), *Gender in the Classroom: Power and Pedagogy*, University of Illinois Press, Chicago, IL, 140-159.
- Berninger, V. & Fuller, F. (1992). Gender differences in orthographic, verbal, and compositional fluency: Implications for diagnosis of writing disabilities in primary grade children. *Journal of School Psychology*, 30, 363-382.
- Engelhard, G., Gordon, B., Siddle-Walker, E. V., & Gabrielson, S. (1994). Writing tasks and gender: Influences on writing quality of black and white Students. *Journal of Educational Research*, 87(4), 197-209.
- Etaugh, C. B., Houtler, B., & Ptasnik, P. (1988). Evaluating competence of women and men: Effects of experimenter gender and group gender composition. *Psychology of Women*

- Quarterly*, 12, 191-200.
- Freedman, S. W. (1979a). How characteristics of student essays influence teachers' evaluation. *Journal of Educational Psychology*, 71, 328-338.
- Hake, R. L., & Williams, J. M. (1981). Style and its consequences: Do as I do, not as I say. *College English*, 43, 433-451.
- Haswell, J. & Haswell, R. H. (1995). Gendership and the miswriting of students. *College Composition and Communication*, 46, 223-254.
- Haswell, R. H., & Haswell, J. T. (1996). Gender bias and critique of student writing. *Assessing Writing*, 3, 31-83.
- Huot, B. (1990). The literature of direct writing assessment: Major concerns and prevailing trends. *Review of Educational Research*, 60(2), 237-263.
- Pajares, F. & Valiante, G. (1997). Influence of self-efficacy on elementary students' writing. *The Journal of Educational Research*, 90(6), 353-360.
- Peterson, S. (1998). Evaluation and teachers' perceptions of gender in sixth-grade student writing. *Research in the Teaching of English*, 33(2), 181-208.
- Peterson, S. S. & Kennedy, K. (2006). Sixth-grade teachers' written comments on student writing: genre and gender influences. *Written Communication*, 23(1), 36-62.
- Roulis, E. (1995). Gendered voice in composing, gendered voice in evaluating: Gender and the assessment of writing quality. In D. L. Rubin(Ed.). *Composing social identity in written language*, Routledge, 151-185.
- Runder, L. M. (1992). Reducing errors due to the use of judges: Practical assessment. *Research & Evaluation*, 3(3).
- Spandel, V. & Culham, R. (1996). Writing assessment, In R. E. Blum and J. A. Alter(eds.), *A Handbook for Student Performance Assessment in an Era of Restructuring*. ASCD.
- Weigle, S. C. (1998). Using FACETS to model rater training effects. *Language Testing*, 15(2), 263-287.
- White, M. J & Brunning, R. (2005). Implicit writing beliefs and their relation to writing quality. *Contemporary Educational Psychology*, 30(2), 166-189.

• 논문 접수 : 2010년 4월 30일 / 수정본 접수 : 2010년 6월 4일 / 게재 승인 : 2010년 6월 17일

ABSTRACT

An Analysis Study of Writing Assessment Differences on Gender and Gender Perception of Pre-service Korean Teachers

Young-Min Park

(Associate Professor, Korea National University of Education)

Sook-Ki Choi

(Lecturer, Korea National University of Education)

The purpose of this paper is to analyze writing assessment differences on gender of pre-service Korean teachers and their gender perception about students who are assessed. 69 pre-service Korean teachers who are sampled test each 35 middle school students' informative and persuasive writings. The sources of their assessments are collected and analyzed for the purpose of this paper. The results are as follows.

First, the mean scores of informative writings and persuasive writing are each 14.27 and 14.02 by pre-service Korean teachers. This mean scores below the scale mean 15.0 point. This result shows pre-service Korean teachers are strict with students' writing assessment.

Second, there are no writing assessment differences according to the raters' gender in informative and persuasive writings. But it is necessary to develop writing assessment expertise of pre-service Korean male teachers because they have a big standard deviation in the assessment on persuasive writings.

Third, there are no writing assessment differences according to the raters' perception about the students' gender in informative and persuasive writings. And there is no so-called the same-sex depreciation. But pre-service Korean male teachers have not agreement on writing assessment if they perceive the students' gender as female. Pre-service Korean male teachers need programs to develop the expertise of writing assessment.

Key words : writing assessment, pre-service Korean teacher, rater's gender, perceived students' gender