

## 현직 국어교사와 예비 국어교사의 쓰기 평가 비교 연구

박 영 민(한국교원대학교 조교수)

최 속 기(한국교원대학교 강 사)

---

### 《 요 약 》

---

이 연구는 쓰기 평가에 대한 현직 국어교사와 예비 국어교사의 차이를 비교·분석함으로써, 쓰기 평가의 전문성, 쓰기 평가의 결과, 쓰기 평가의 신뢰도에 영향을 미치는 요인을 탐색하는 데 목적이 있다. 쓰기 평가에 대한 이 두 집단의 차이를 비교·분석하기 위하여, 현직 국어교사는 49명, 예비 국어교사는 24명에게 중학생 글 39편을 평가하게 하고 그 결과를 통계 처리하였다. 평가자 협의의 영향력을 조사하기 위해 평가 과정에서 평가자 협의는 진행하지 않았다. 이 연구를 통해 얻은 결과는 다음과 같다.

첫째, 중학생 글을 평가한 결과는, 평가자 협의를 거치지 않았음에도 불구하고 현직 국어교사와 예비 국어교사 사이에 매우 높은 상관도를 보였다. 둘째, 현직 국어교사와 예비 국어교사는 형식 어법 요인에서 통계적으로 유의한 차이를 보였다. 또한 전반적으로 예비 국어교사들이 현직 국어교사보다 쓰기 평가 점수를 낮게 주는 현상이 확인되었다. 셋째, 현직 국어교사와 예비 국어교사의 평가 총점을 비교한 결과, 현직 국어교사는 글의 내용에 평가의 중점을 두는 경향이 발견되었다. 그러나 예비 국어교사는 내용 및 형식 요소를 균등한 관점에서 평가하고 있음을 알 수 있었다.

위 연구 결과는 현직 국어교사의 쓰기 평가 전문성에 대한 반성적 성찰의 계기를 제공한다. 또한, 쓰기 평가에서 현직 국어교사의 평가 전문성이 드러나지 않은 원인과 현 쓰기 평가 방식의 문제점을 재고하도록 한다. 더불어 평가자 협의의 효과도 확증할 수 없는 결과가 나타났다. 평가자 협의와 쓰기 평가의 상관 등에 대한 논의가 추후 보완되어야 하며, 또한 평가자인 국어교사의 쓰기 신념 요인 등에 대한 영향 분석도 후속 연구로 이루어져야 할 것이다.

주제어 : 쓰기 평가, 쓰기 평가자 요인, 쓰기 평가 비교, 국어교사의 쓰기 평가

---

## I. 서론

이 글은 현직 국어교사와 예비 국어교사의 쓰기 평가 결과가 어떻게 차이가 있는지를 비교·분석하는 데 목적이 있다. 이를 통해서 교과 전문가인 국어교사의 쓰기 평가가 얼마나 일관성이 있는지를 파악하고자 한다. 분석 결과는 국어교사의 평가 전문성을 탐색하는 데 중요한 정보를 제공하고, 쓰기 평가에 관한 교육적 함의를 이끌어내는 데 기여할 수 있을 것으로 기대된다. 현직 국어교사는 학생 글의 평가 경험이 많으므로, 평가 경험이 없는 예비교사에 비해 어떤 차이가 있는지를 밝힘으로써 평가 경험의 효과도 파악할 수 있다.

쓰기 평가는 쓰기 교수·학습에 관한 학생 정보를 수집하는 체계적인 활동이다. 이 정보를 수집하기 위해서 교사는 일반적으로 학생이 작성한 글을 직접 읽고 수량화된 점수를 얻는 방법을 사용한다. 이 방법은 학생의 쓰기 능력을 파악하는 효과적인 방법이지만, 평가자의 일치도가 떨어지는 문제가 있다. 그래서 쓰기 결과를 평가할 때는 평가자의 일치도가 떨어지는 문제를 방지하기 위하여 구체적인 평가 기준 작성과 평가자 간 협의를 권장한다.

이 글에서는 현직 국어교사와 예비 국어교사 모두 평가자간 협의 없이 학생 글을 직접 평가하도록 조건화하였다. 협의 없이 글을 평가하였을 때, 두 집단의 평가 결과의 차이가 어떤 양상으로 나타나는지를 파악하기 위한 것이다. 방법적으로 평가자간 협의를 권장하지만, 실제로 학교에서는 평가자간 협의를 거의 하지 않는다. 그러므로 이 조건은 현실을 반영한 것으로 이해할 수 있다. 이 조건을 통해서 기대하는 바는 현직 국어교사의 평가 결과가 예비 국어교사의 평가 결과보다 더 높은 일치도를 보이는 것이다. 즉, 현직 국어교사의 평가 전문성 지표가 더 높다는 점을 기대하는 것이다.

이 글에서 주목하고 있는 쓰기 평가 요인은 평가자, 즉 교사 요인이다. 쓰기 평가에 영향을 미치는 여러 요인 중에서 평가자 요인에 주목하는 이유는 평가자인 교사가 바로 평가의 주체이기 때문이다. 국어교사가 쓰기 평가를 올바르게 수행할 수 있어야 학생들의 쓰기 능력 신장을 효과적으로 계획할 수 있다. 이러한 논리에 따르면 쓰기 평가자를 탐색하는 연구는 매우 중요하다고 할 수 있는데, 그럼에도 불구하고 평가자 요인을 구체적으로 탐색한 연구는 그리 많지 않다. 이 글은 거의 연구되지 않은 평가자 요인을 다루었다는 점, 쓰기 평가에 영향을 미치는 국어교사 요인에 관한 정보를 제공한다는 점에서 의의를 발견할 수 있다.

## Ⅱ. 이론적 배경

### 1. 쓰기 평가와 국어교사의 내적 요인

쓰기 평가에 영향을 미치는 교사의 내적 요인은 교사의 인지적 요인과 정의적 요인으로 분류할 수 있다. 교사의 인지적 요인은 국어교사가 지닌 쓰기 과정 및 쓰기 평가에 대한 지식을 반영한다. 쓰기 평가에서 교사의 지식 요인을 다룬 연구로 이성영(2005)을 꼽을 수 있다. 이성영(2005)은 국어교사의 쓰기 평가 전문성 기준과 모형을 다루면서 국어교사의 평가 전문성 요인으로 교사의 지식, 수행, 태도 등을 제시하였다. 이러한 요인들은 국어교사가 학생 글 평가할 때 복합적이면서 메타적으로 영향을 미친다고 하였다. 특히, 국어교사의 지식은 좋은 글이 무엇인지 등에 관한 쓰기관, 문장관을 형성하는 데 큰 영향을 미친다고 하였다. 쓰기 평가자의 전문성에 바탕을 둔 평가 차이에 관한 실제적 연구는 김라연(2007)에서도 실시되었다. 김라연(2007)은 현직 국어교사, 예비 국어교사 및 고등학생 총 72명을 대상으로 한 쓰기 평가 결과 비교를 통해 평가 전문성의 차이가 평가 결과의 차이를 이끌어 낸다는 점을 논의한 바 있다.

교사의 인지 양식에 따라 평가자의 평가 양상이 달라질 수 있다는 점에 주목한 연구도 있다. 민태범(1991)은 장(場) 의존적 인지 양식을 보이는 교사와 장 독립적 인지 양식을 보이는 교사를 구분하여 이 두 집단에서 발견되는 쓰기 평가의 차이를 분석하였다. 장 의존적인 평가자는 사회 지향적이고 타인 지향적이어서 평가 대상에 대해 교정적 태도를 보이는 예가 드물다. 그러나 장 독립적인 평가자들은 학생들의 글에 대해 비판적 비평을 하는 교정적 피드백 양상을 보이고, 학생 글이 성취 수준이 낮을 때에는 심지어 불쾌감을 드러내는 것으로 나타났다. 이에 따라 교사 130명이 채점한 논술 답안을 분석한 결과, 대부분의 답안에서 장 의존적인 집단이 장 독립적 집단보다 평균 점수가 높았다. 이와 관련하여 Runder(1992)와 Engelhard(1994)는 평가자의 엄격함과 관대함에 대한 인지 양식이 평가에 영향을 미친다고 보았다.

쓰기 평가에 영향을 미치는 교사의 정의적 요인에는 성별 요인과 쓰기 신념 등이 있다. 먼저 성별에 따른 국어교사의 평가 경향은 두 가지로 구분하여 살펴 볼 수 있다. 첫 번째는 교사 자신의 성별 변인에 기반을 둔 쓰기 평가 차이이고, 두 번째는 학생들의 성별 경향성에 대한 교사의 성별 경향성이다. 전자와 관련한 연구로는 Peterson(2006)이 있다. Peterson(2006)은 서사적인 글과 설득적인 글에 대한 남녀 교사 108명의 쓰기 평가 결과를 분석하면서 쓰기 평가에 영향을 미치는 교사의 성별 변인과 성 인식 경향을 탐색하였다. 분석 결과, 교사 성별 변인에 따라 평가 결과의 차이가 뚜렷하게 나타났다. 여교사와 남교사는 쓰기에 대한

논평(comment)에서 서로 다른 양식으로 평가하였다. 여교사는 글의 규칙과 조직에 대해 중점을 두었으나, 남교사는 미적인 문체(style)에 초점을 두었다. 이러한 성별 변인에 따른 평가 양상의 차이는 Barnes(1990)와 Roulis(1995)에서도 확인할 수 있다. 이들의 연구에 따르면, 평가의 양적 수준에 있어서도 여교사들은 일반적으로 남교사들보다 더 많은 논평을 남겼다.

후자에 대한 연구로는 Peterson(1998)이 주목된다. Peterson(1998)에 따르면, 교사들은 남학생이 쓴 것으로 인식한 글에 대해서 여학생이 쓴 글로 인식한 것보다 더 많은 수정을 가했고 더 많은 비평을 제시했다. 그 결과, 남학생이 쓴 것으로 인식된 학생 글에 나타난 수정 지시와 비평의 양은 여학생으로 인식된 글에 나타난 것보다 훨씬 더 많은 양을 보였다. 교사들은 여학생들보다는 남학생들에게 쓰기 평가를 통한 도움이 더 많이 필요하다고 생각했기 때문에, 남학생으로 인식된 글에 더 많은 설명, 제안, 지도 내용을 작성한 것으로 나타났다.

Peterson(1998)에서 관심을 끄는 것은 쓰기 평가에서 동성(同性)에 대한 평가 절하 현상이 발견된다는 점이다. 즉, 남자 교사는 남학생이 쓴 것으로 인식한 글에 대해서 비판을 더 많이 하고, 여자 교사도 여학생이 쓴 것으로 인식한 글에 대해서 비판을 더 많이 한다는 것이다. 평가자의 성이 학생의 성과 같으면 더 부정적으로 평가하는 경향이 바로 쓰기 평가의 동성 평가 절하 현상이다. 이러한 현상은 다른 연구에서도 이미 확인된 바 있다(Etaugh, Houtler & Ptasnik, 1988; Haswell & Haswell, 1996). 선행 연구에 따르면 남교사들은 여학생의 글보다 남학생의 글에서 더욱 많은 수정 의견을 제시하고 비판을 했다. 남교사들은 남학생의 것으로 보이는 글보다 여학생의 것으로 보이는 글을 더 칭찬하는 경향이 있었다. 대조적으로 여교사들은 여학생이 쓴 것처럼 보이는 글에 칭찬을 하기보다는 남학생이 쓴 것으로 보이는 글을 더욱 칭찬했다. 이러한 연구들은 교사의 학생 성별에 대한 인식이 쓰기 평가에 영향을 미친다는 점을 잘 보여준다.

쓰기 평가에 영향을 미치는 교사의 정의적 요인 중의 하나는 쓰기 신념이다(박영민, 2007; 박영민, 2008). 교사의 쓰기 신념은 크게 두 가지 양상으로 나타나는데, 하나는 교류적 쓰기 신념(transactional writing belief)이고, 다른 하나는 전달적 신념(transmissional writing belief)이다. 전달적 신념은 쓰기를 권위 있는 자료로부터 독자에게 정보를 전달하기 위한 하나의 수단으로 인식하는 관점이고, 교류적 신념은 쓰기를 개인적, 비판적 텍스트 구성 방식으로 인식하고 필자의 쓰기 수행 과정에 대해 자신의 생각을 적극적으로 통합시키고자 하는 관점이다(White & Brunning, 2005). 이러한 쓰기 신념은 학생 글을 대하는 교사의 인식과 태도에 영향을 미친다. 교사가 어떤 신념을 가졌는지에 따라 학생들의 글에 대한 평가를 다르게 내릴 수 있다. 가온아(2008)는 56명의 현직 국어교사의 쓰기 효능감과 쓰기 신념에 따른 학생 글에 대한 평가를 분석하였는데, 이에 따르면 글의 조직 요인에서 전달적 쓰기 신념을 지닌 국어교사 집단이 학생의 글에 대해 더 높은 점수를 부여한 것으로 밝혀졌다.

## 2. 쓰기 평가와 국어교사의 외적 요인

쓰기 평가에 영향을 미치는 교사의 외적 요인은 학생 요인과 글 요인으로 구분할 수 있다. 쓰기 평가에 영향을 미치는 학생 요인에는 학생의 성(性), 인종, 가정환경, 외모 등이 보고되어 왔다. 대부분 쓰기 평가에서 후광 효과로 작용하는 요인들이다. 예를 들어, 외모가 뛰어난 학생이 평가에서 교사로부터 유리한 결과를 받는다는 결과는 Dion(1972)이나 Rich(1975)의 연구를 통해 밝혀진 바 있다. 심지어 외모가 지능을 예측할 것이라는 교사의 신념에 대한 연구도 있다(Fleson, 1980). 외모가 매력적인 학생들은 학업 성취 뿐 아니라 미래 성공이 예측되기도 한다(Adams & Cohen, 1976). 그러므로 평가의 객관성과 신뢰성을 확보하기 위해서는 이러한 요인들을 적절하게 통제해야 한다.

쓰기 평가에 영향을 미치는 것으로는 글 요인이 있다. 글 요인에는 학생들의 필체, 답안의 길이 정도, 맞춤법 등 학생이 작성한 글의 특성과, 평가 상황에서 어떠한 방식으로 글이 배열되었는지에 대한 것 등이 포함된다. 일반적으로 국어교사들은 글의 내용, 조직, 표현을 중심으로 학생 글을 평가한다. 그러나 필체, 배열 방식, 길이 등에 의해서도 교사는 영향을 받기도 한다.

Marshall & Power(1969)의 예비 교사를 대상으로 한 쓰기 평가 영향 연구에 따르면, 쓰기 평가를 수행할 때 내용에 초점을 두라는 지시에도 불구하고 글을 평가하는 평가자인 예비 교사들은 학생들의 답안지에서 나타난 필체, 작문 과정에서 나타난 혹은 문법적 측면에서 드러난 오류 등에 의해 영향을 받았다는 사실이 보고되었다. 또한 Lee(2004)는 글을 쓰는 방식에 따라 평가자의 평가 결과가 달라질 수 있다는 점을 보고하였다. 학생들이 손으로 직접 작성한 글과 워드 프로세서를 활용하여 작성한 글에 대한 총체적 평가와 분석적 평가를 실시한 결과, 총체적 평가에서는 통계적으로 유의한 차이를 발견할 수 없었지만, 분석적 평가에서는 각 평가 요소 간에 통계적으로 유의한 차이가 발견되었다. 손으로 쓴 글에 비해 워드 프로세서를 활용한 경우, 평가자간 신뢰도가 더 높았다.

글의 분량도 쓰기 평가에 영향을 미치는 것으로 나타났다. 일반적으로 필자가 글을 많이 쓰는 경우 아이디어가 풍부하다는 생각을 갖도록 하거나 혹은 문장에서의 자신감이 나타나므로 인상적인 측면에서 좋은 글이라는 생각을 갖도록 만든다(Tollefson & Tracy, 1979). 글의 배치도 쓰기 평가에 유의한 차이를 가져오는 요인이다(Daly & Dickson-Markham, 1982). 예를 들면, 잘 쓴 글 사이에 상대적으로 낮은 수준의 글이 포함되었을 경우, 성취 수준에 대한 인상적 대비 현상이 발생하여 낮은 수준의 글을 더 낮게 평가하는 경향이 나타난다. 성취 수준이 낮은 글 사이에 잘 쓴 글이 포함되었을 경우에는 반대 현상이 일어난다. 이러한 대비 효과는 객관적인 쓰기 평가를 방해하므로 적절하게 통제해야 한다.

### Ⅲ. 연구 방법

#### 1. 연구 대상

현직 국어교사와 예비 국어교사의 쓰기 평가를 분석하기 위해 여러 지역에 재직 중인 국어교사 49명, 한국교원대학교에 재학 중인 3학년 예비 국어교사 24명을 대상으로 조사를 실시하였다. 이때 교사 변인으로 성별 변인을 포함하여 현직 교사의 경우 남교사 9명, 여교사 40명으로부터 자료를 수집하였다. 예비 교사는 남학생 14명, 여학생 10명을 선정하였다. 현직 교사는 경력별 변인에 대한 분석을 고려하여 1~5년, 6~10년, 11~20년, 20년 이상으로 구분하여 자료를 수집하였다. 연구 대상의 규모 및 분포는 <표 1>과 같다.

<표 1> 연구 대상의 규모 및 분포

		성별		총계
		남	여	
현직 교사	1~5년	1	9	10
	6~10년	7	18	25
	11~20년	1	11	12
	20년 이상	0	2	2
	총계	9	40	49
예비 교사		14	10	24
총계		23	50	73

예비 교사들은 양성 과정 중에 있기 때문에 경력 변인을 설정하지 않았다. 예비 교사들은 교육 실습의 경험이 없어 쓰기 지도 및 쓰기 평가의 경험이 없었다.

#### 2. 검사 도구

이 글에서는 “선의의 거짓말은 필요한가?”라는 논제에 따라 작성된 중학생 논술문 총 39편을 평가 대상으로 삼았다. 학생 글은 논술 대회에서 작성된 것으로, 논술 대회 평가에서 양호한 글로 분류되었던 것들이다. 평가 과정에서 필체 등 주관적 요인에 의한 영향을 통제하기 위하여 학생 글을 워드 프로세스로 입력 처리하였다. 띄어쓰기와 같은 부분적인 맞춤법 오류는 수정하였다. 학생 글과 더불어 쓰기 평가 기준표와 채점표를 함께 제시하였다. 쓰기 평가 기준표는 Spendel & Culham(1996)이 제시한 총체적 평가 기준을 <표 2>와 같이 중

학생 쓰기 수준에 적합하도록 수정하여 사용하였다. 평가 범주는 내용, 조직, 표현, 단어선택, 형식 어법으로 구성되었다.

〈표 2〉 Spendel & Culham(1996)의 평가 기준표

평가 범주	평가 기준
내용	<ul style="list-style-type: none"> <li>5점: 글의 중심 내용(주제)가 명료하며 독자의 주의를 끈다. 세부적인 내용들은 전체적인 중심 내용(주제)과 부합한다.</li> <li>3점: 글의 중심 내용(주제)이 다소 명확하지 못하다. 글 전체의 중심 내용(주제)과 부합하지 않은 세부 내용이 들어 있다.</li> <li>1점: 글의 중심 내용(주제)이 잘 드러나 있지 않다. 세부 내용은 글의 전체적인 중심 내용과 잘 어울리지 않는다.</li> </ul>
조직	<ul style="list-style-type: none"> <li>5점: 설명 대상의 특성을 고려하여 조직되었으며, 중심 내용이 잘 드러나도록 조직되었다. 내용의 순서나 구조가 독자가 이해하기 쉽도록 되어 있다.</li> <li>3점: 설명 대상의 특성은 고려하였으나, 글의 구조가 명확하게 드러나지 않는다. 독자가 이해하는데 어려움이 따른다.</li> <li>1점: 설명 대상의 특성이 고려되지 않았으며, 글의 구조가 명확하게 드러나지 않는다. 독자가 이해하는데 어려움이 따른다.</li> </ul>
표현 (어조 및 태도)	<ul style="list-style-type: none"> <li>5점: 독창적이며 흥미롭게 표현되어 있으며, 독자가 쉽고 정확하게 이해할 수 있도록 표현되었다. 필자의 주체적인 목소리도 드러난다.</li> <li>3점: 독자가 쉽게 이해할 수 있도록 표현되었으나, 독창성이나 흥미는 다소 떨어진다. 필자의 주체적인 목소리도 잘 드러나지 않는다.</li> <li>1점: 내용만을 기계적으로 나열하여 글의 생동감이 떨어지며 흥미를 주지 못한다. 독자가 쉽게 이해할 수 있도록 표현되지 않았다.</li> </ul>
단어 선택	<ul style="list-style-type: none"> <li>5점: 내용을 정확히, 흥미롭게, 자연스럽게 전달할 수 있는 단어가 선택되었다.</li> <li>3점: 대체적으로 단어 선택이 내용 전달에 무리가 없으나 부적절한 단어들이 포함되어 있다.</li> <li>1점: 내용을 전달하는 단어가 매우 제한적이어서 단어의 선택이 풍부하지 못하다.</li> </ul>
형식과 어법	<ul style="list-style-type: none"> <li>5점: 표준적이며 모범적인 쓰기 형식이 잘 드러나 있다.(어법, 구두점, 철자, 단락 구분 등).</li> <li>3점: 제한된 범위에서만 글의 표준적 형식이 확인된다.</li> <li>1점: 철자, 구두점, 문법에서 잘못된 것이 많아 내용을 파악하며 읽는 것이 어렵다.</li> </ul>

### 3. 연구 절차

쓰기 평가는 현직 국어교사와 예비 국어교사를 구분하여 실시하였다. 39편의 학생 글과 평가 기준, 채점표, 개인 기록표를 포함한 쓰기 평가 검사지를 배포하고, 각 교사들은 평가자 협의 없이 평가를 개별적으로 진행하였다. 현직 국어교사들은 2008년 8월 10일부터 8월

17일까지 일주일 동안 개별적으로 평가를 실시하였다. 평가 기준인 내용, 조직, 표현, 단어선택, 형식 어법의 점수를 각각 채점하였다. 예비 교사의 경우도 2008년 9월 2일부터 9월 9일까지 일주일 동안 평가자 협의 없이 개별적으로 평가를 실시하였다. 개인별 평가가 이루어졌기 때문에 평가 과정에서의 평가자 협의도 없었다. 이를 통해 얻은 평가 결과를 활용하여 비교·분석하였다. 평가자 신뢰도는 Chronbach  $\alpha$ 로 추정하였으며, 그 값은 현직 국어교사의 경우 .942, 예비 국어교사는 .934로 나타났다. 구체적인 결과는 <표 3>과 같다.

<표 3> 평가자 신뢰도

평가 항목	현직 국어교사	예비 국어교사
총점	.942	.934
내용	.908	.864
조직	.901	.872
표현	.895	.860
형식어법	.916	.896
단어선택	.930	.910

#### 4. 분석 도구

위와 같은 과정을 통하여 수집된 현직 국어교사와 예비 국어교사들의 평가 자료는 SPSS 12.0 for windows 한글판을 이용하여 분석하였다. 먼저 현직 교사와 예비 교사의 평가 결과에 대해 독립표본 t 검증을 실시하였다. 이때 쓰기 평가의 총점과 더불어, 내용, 조직, 표현, 단어 선택, 형식 어법에 대한 세부 점수별 차이도 분석하였다. 다음으로 세부 변인으로서의 성별 차이를 살펴보기 위해 독립표본 t 검증을 실시하였으며, 현직 교사의 경력별 차이를 보기 위해 ANOVA를 실시하였다. 또한 두 집단 평가자 사이의 상관성을 탐색하기 위한 상관 분석을 실시하였다.

### IV. 결과 및 논의

#### 1. 쓰기 평가의 집단별 평균 차이 비교

쓰기 평가에 대한 현직 국어교사와 예비 국어교사의 평균 차이를 비교하기 위해 독립표본 t 검증을 실시하였다. 결과는 <표 4>와 같다.



〈표 4〉 쓰기 평가의 평균 차이

평가변인	교사변인	표본크기	평균	표준편차	최댓값	최솟값	t
총점	현직 교사	39	16.0727	2.56352	21.14	10.14	1.541 (p=.127)
	예비 교사	39	15.1344	2.80834	20.84	8.44	
내용	현직 교사	39	3.1863	.62847	4.33	1.61	.857 (p=.394)
	예비 교사	39	3.0564	.70712	4.44	1.68	
조직	현직 교사	39	3.1884	.67795	4.55	1.69	1.567 (p=.121)
	예비 교사	39	2.9467	.68397	4.28	1.32	
표현	현직 교사	39	3.2266	.49505	4.18	2.12	1.486 (p=.142)
	예비 교사	39	3.0513	.54583	4.24	1.84	
형식어법	현직 교사	39	3.1753	.49039	4.12	1.90	2.013 (p=.048)
	예비 교사	39	2.9241	.60585	4.12	1.64	
단어선택	현직 교사	39	3.2967	.44712	4.16	2.14	1.267 (p=.209)
	예비 교사	39	3.1559	.53112	4.16	1.60	

분석 결과에 따르면, 논술문 평가에서 총점, 내용, 조직, 표현, 형식 어법, 단어 선택에 대한 평가의 평균은 현직 국어교사가 예비 국어교사보다 더 높은 것으로 나타났다. 전반적으로 예비 국어교사들이 현직 국어교사들에 비해 학생 글에 대해 더 낮은 점수를 부여하는 현상을 보였다. 그러나 이러한 평균 차이는 쓰기 평가의 총점, 내용, 조직, 표현, 단어 선택의 측면에서는 통계적으로 유의하지 않았다. 단, 형식 어법에서는 t 값은 2.013이고 유의확률은 .048( $p < .05$ )로 두 집단 사이에 통계적으로 유의한 차이가 발견되었다. 예비 국어교사들이 현직 국어교사들에 비해 형식 어법에 대한 평가를 더욱 엄격히 적용하는 것으로 볼 수 있다.

전반적으로 현직 국어교사에 비해 예비 국어교사들의 점수가 낮았는데, 이는 예비 국어교사들이 평가 기준을 엄격하게 적용하였기 때문이다. 이러한 결과는 예비 국어교사의 지식 효과(knowledge effect)에 의한 것으로 해석된다. 지식 효과란 어떤 사람이 어떤 문제를 잘 알고 있는 경우, 다른 사람들도 그 문제를 잘 알고 있으리라고 믿는 현상을 말한다. 그래서 쓰기 평가에 지식 효과가 나타나면 자기 자신만큼 알고 있을 것으로 생각하면서 평가 점수가 낮게 매기는 경향이 나타난다. 초임 교사가 학생들의 수준을 잘 이해하지 못하고 시험 문제를 어렵게 출제하는 경향도 지식 효과에 따른 것이다. Hayes et al.(2006)은 서로 다른 난이도를 보이는 텍스트에 대해 88명의 대학원생들을 대상으로 하여 지식 효과를 검증한 결과, 자신의 이해 수준을 바탕으로 하여 다른 독자들의 이해 수준을 평가한다는 사실을 발견한 바 있다.

예비 국어교사의 쓰기 평가 상황에서도 지식 효과가 동일하게 작동한 것으로 보인다. 예

비 국어교사들은 자신이 상정한 이상적인 수준을 바탕으로 하여 중학생들의 글을 평가했던 것이다. 그러나 현직 국어교사의 경우에는 지식 효과의 영향이 적었다. 현직 국어교사가 상정한 쓰기 수준은 실제 중학생의 쓰기 능력 수준을 반영하고 있다고 볼 수 있다. 이에 따라 현직 국어교사는 상대적으로 높은 점수를 부여하였다. 예비 국어교사들의 평가 점수가 낮은 것은 학생 글에 대해 평가해 본 경험이 없다는 점도 관련되어 있다. 학생들의 수준을 잘 이해하지 못하는 지식 효과는 대체적으로 경험이 축적되면서 해소된다. 중학생에게 쓰기를 지도한 경험이 많은 국어교사는 지식 효과에 빠지지 않는다. 경험으로써 지식 효과를 방지하고, 중학생들의 실제적인 쓰기 능력을 예측할 수 있게 된다. 그러나 예비 국어교사들은 교육 실습의 경험도 없고 쓰기 평가 및 쓰기 지도의 경험도 없어 자신이 설정한 수준에 따라 학생 글을 평가할 수밖에 없었다. 그 결과, 쓰기 평가 점수가 낮았다.

다음은 현직 국어교사의 성별 변인과 경력별 변인에 따라 독립표본 t 검증과 ANOVA를 적용하여 분석하였다. 먼저 현직 교사의 성별 및 경력별 변인에 따른 쓰기 평가 차이의 분석 결과를 보면 다음과 같다. <표 5>는 현직 성별에 따른 국어교사의 쓰기 평가 결과를 분석한 것이다. 쓰기 평가의 총점, 내용, 조직, 표현, 형식 어법, 단어 선택 등에서 전체적으로 여교사가 남교사에 비해 쓰기 평가가 더 엄격한 것으로 나타났다. 특히 내용에 대한 평가에서 여교사는 남교사보다 가장 낮은 점수를 부여하였다. 그러나 쓰기 평가의 총점, 조직, 표현, 단어 선택, 형식 및 어법에서의 차이는 모두 통계적으로 유의하지 않았다.

〈표 5〉 성별에 따른 현직 국어교사의 쓰기 평가 비교

변인	성별	표본크기	평균	표준편차	t
총점	남	9	17.0199	2.47727	1.549 (p=.128)
	여	40	15.8596	1.92593	
내용	남	9	3.4558	.52121	2.004 (p=.051)
	여	40	3.1256	.42983	
조직	남	9	3.3761	.50377	1.427 (p=.160)
	여	40	3.1462	.42145	
표현	남	9	3.4274	.49155	1.685 (p=.099)
	여	40	3.1814	.37277	
형식 어법	남	9	3.3761	.55173	1.486 (p=.144)
	여	40	3.1301	.42421	
단어 선택	남	9	3.4274	.51982	.899 (p=.414)
	여	40	3.2673	.47475	

현직 국어교사의 또 다른 변인인 경력에 따라 쓰기 평가의 차이를 살펴보기 위해 우선 기술 통계량을 구하였다. 그 결과는 <표 6>과 같다.

<표 6> 경력에 따른 현직 국어교사 쓰기 평가의 기술통계

변인	표본크기	평균	표준편차	최솟값	최댓값
1~5년	10	15.8846	2.04494	13.41	19.67
6~10년	25	16.2421	2.36881	11.13	21.03
11~20년	12	16.0662	1.52333	13.69	19.44
20년 이상	2	14.9359	1.46861	13.90	15.97
합계	49	16.0727	2.05976	11.13	21.03

경력을 기준으로 하여 살펴보면, 전체 교사의 쓰기 평가에 대한 평균 점수는 16.0727이며, 이때 25점 만점에서 평가 최솟값의 평균은 11.13점, 최댓값의 평균은 21.03점이다. 이에 따르면 경력 20년 이상 집단과, 경력 1~5년의 집단이 다른 경력 수준의 교사들에 비해 상대적으로 낮은 수준으로 쓰기 평가 점수를 부여했다. 경력 분포에서 가장 높은 경력 수준을 보이는 집단과 가장 낮은 경력 수준을 보이는 집단이 학생들의 쓰기 평가에서 가장 낮은 점수를 부여 했다는 점은 매우 주목할 만한 사실이다. 이와 대조적으로 경력이 6~10년인 교사 집단이 가장 높은 점수를 학생들에게 부여한 것으로 나타났고 이들 점수간의 표준편차도 가장 높은 수준으로 나타났다.

경력별로 표본크기가 다르지만 Levene의 등분산 검증에서 유의확률이  $p=.227$ 로 등분산이 가정되어 차이 검증의 가능성을 확보하였다. <표 7>은 경력에 따른 현직 국어교사의 평가 차이를 분석하기 위한 일원분산분석의 결과를 나타낸 것이다. 4개 집단으로 구분된 현직 국어교사의 평균 차이 분석에서 F 값이 .274, 유의확률이 .884로 나타나 각 집단 간의 차이는 통계적으로 유의하지 않은 것으로 나타났다.

<표 7> 현직 국어교사의 경력별 쓰기 평가 차이에 대한 일원분산분석 결과

	제곱합	자유도	평균제곱	F	유의확률
집단-간	3.656	3	1.219	.274	.844
집단-내	199.990	45	4.444		
합계	203.645	48			

다음으로 성별에 따라 예비 국어교사의 쓰기 평가 차이를 분석하였다. <표 8>의 결과를 살펴보면, 조직과 단어 선택을 제외한 나머지 평가 점수에서 여자 예비 교사가 남자에 비해

더욱 낮은 수준으로 평가 점수를 부여하였다. 특히 내용에 대한 평가에서 남학생에 비해 여학생이 더 낮은 점수를 부여하였다는 점은 현직 교사의 여교사와 동일한 양상을 보였다. 그러나 내용 측면에서 여자 평가자의 쓰기 평가가 더욱 엄격히 적용되는 원인에 대한 실제적 탐색은 더 탐색할 필요가 있다. 그러나 예비 교사들의 성별 변인에 따른 차이는 모두 통계적으로 유의하지 않은 것으로 나타났다.

〈표 8〉 성별에 따른 예비 교사 쓰기 평가 차이

변인	성별	표본크기	평균	표준편차	t
총점	남	14	15.0879	1.97147	-.010 (p=.992 )
	여	10	15.0974	2.58919	
내용	남	14	3.0806	.46512	.288 (p=.776 )
	여	10	3.0256	.48393	
조직	남	14	2.9432	.52454	-.039 (p=.969 )
	여	10	2.9510	.46189	
표현	남	14	3.0806	.39797	.369 (p=.715 )
	여	10	3.0140	.50447	
형식 어법	남	14	2.9304	.45452	.068 (p=.947 )
	여	10	2.9161	.60451	
단어 선택	남	14	3.0531	.47795	-1.093 (p=.300 )
	여	10	3.2867	.59184	

〈표 5〉와 〈표 8〉에 따르면, 현직 교사이든 예비 교사이든 쓰기 평가의 성별 차이는 통계적으로 유의하지 않다고 결론지을 수 있다. 교사의 성이 무엇인가에 따라 쓰기 평가의 양상이 달랐다고 한 Peterson(1998) 등의 연구와는 차이가 있다. 이렇게 결과가 다른 것은 이 글에서 다른 평가는 점수를 매기는 수량적인 결과 평가였다는 점 때문일 것이다. 교사의 자유반응이 허용하지 않는 평가 방법이었기 때문에 성별에 따른 교사의 특징이 잘 드러나지 않은 것으로 보인다.

## 2. 쓰기 평가 결과의 상관 분석

현직 국어교사와 예비 국어교사의 평가 결과에 대해 상관성을 분석하여 두 집단의 쓰기 평가가 어떠한 관련성을 보이는지 살펴보고자 한다. 두 집단 사이의 상관 분석에 앞서 두 집단 내의 상관도를 분석하였다. 즉, 쓰기 평가의 하위 요인이 쓰기 평가 총점과 어떤 수준의

상관이 있는지를 현직 국어교사와 예비 국어교사로 구분하여 분석하였다. 그 결과는 <표 9>와 같다. 현직 교사와 예비 교사의 쓰기 평가 총점과 쓰기 평가 요인별 상관은 예비 국어교사의 단어 선택의 상관계수 .888을 제외하고 모두 .9이상의 매우 높은 상관관계를 보여준다.

<표 9> 현직 교사 총점 및 하위 요인 간 상관

변인		총점	내용	조직	표현	형식 어법	단어 선택
총점	현직 국어교사	1	.929**	.914**	.955**	.955**	.952**
	예비 국어교사	1	.921**	.918**	.927**	.911**	.888**

\*\*  $p < 0.01$

39명의 학생 논술에 대한 현직 국어교사의 평가와 예비 국어교사의 평가에서 평균 점수를 비교하여 두 집단 사이의 상관을 분석하였다. 상관 분석은 쓰기 평가의 총점과, 쓰기 평가의 하위 요소 모두에 적용하였다. 그 결과는 <표 10>과 같다.

<표 10> 현직 국어교사와 예비 국어교사 간 쓰기 평가 상관도

총점 상관도	내용 상관도	조직 상관도	표현 상관도	형식 어법 상관도	단어 선택 상관도
.941	.903	.920	.880	.880	.887

<표 10>에 따르면 총점 상관은 .941이고 이때 유의확률은 .000으로  $p < .01$ 로 통계적으로 유의한 수준에서 매우 높은 상관을 보였다. 또 개별 쓰기 평가 요소별 상관은 조직, 내용, 단어 선택, 형식 어법, 표현의 순서로 상관관계가 높게 나타났으며 모두 유의확률 .000으로  $p < .01$  수준에서 통계적으로 유의하게 나타났다. 특히 조직, 내용의 경우는 상관계수가 .920, .903으로 매우 높은 수준의 상관을 보였다. 그러나 이에 반해 단어 선택의 상관계수는 .887이고 형식 어법과 표현의 상관계수는 .880으로 상대적으로 낮은 수준의 상관을 보였다.

쓰기 평가 결과에 대한 현직 국어교사와 예비 국어교사의 상관분석에서 이들의 상관은 매우 높은 수준으로 나타났다. 두 집단의 평가 결과는 매우 높은 정적 상관을 보였다. 단, 내용 및 조직에 대한 상관관계에 비해 상대적으로 단어 선택, 형식 어법, 표현 등의 쓰기 평가 요소에 대해서는 상대적으로 낮은 수준의 상관을 보였다. 그러나 전체적으로는 매우 높은 상관도를 보여주고 있다. 이러한 결과는 두 집단의 쓰기 평가의 차이가 크지 않음을 뜻한다. 쓰기 평가 경험의 유무에 따라 두 집단 사이에 차이가 있을 것으로 예상하였으나 상관 수준이나 차이 분석 결과에서 통계적으로 유의한 차이를 발견할 수 없었다. 평가자들은 평가에 대한 협의 없이 개별적으로 평가하였음에도 불구하고 두 집단의 차이는 거의 나타나지 않았

다. 평가자 협의가 없으면 일치된 결과에 도달할 수 없다는 선행 연구의 주장과는 크게 다른 결과이다. 두 집단의 평가 결과에서 차이가 발견되지 않은 이유는 다음과 같이 분석할 수 있다.

첫째, 평가 대상으로 삼은 39편의 중학생 글이 논술 대회에서 양호한 글로 평가된, 일종의 선발된 높은 수준의 글이라는 특성에서 비롯된 것으로 보인다. 평가 대상인 학생 글이 이미 일정한 수준에 도달해 있었기 때문에 평가자들 사이의 관점 차이나 해석 차이가 적게 나타난 것으로 해석된다. 둘째, 평가 기준의 구체성에서 비롯된 것으로 해석된다. 쓰기 평가 이론에서는 평가자의 일치도를 높이려면 구체적인 평가 기준을 제시해야 한다는 점을 강조한다. 기준이 구체적이어야 평가 기준 해석의 일치가 가능하기 때문이다. 그런데 이 연구에서 적용한 평가 기준은 이미 구체적으로 작성되었기 때문에 평가자 일치도가 높게 유지된 것으로 보인다. 일반적인 수준의 평가 기준을 제시하였다면 두 집단 간 평가 차이는 분명하게 나타났을 것이다. 셋째, 예비 교사들의 쓰기 평가에 대한 지식 수준이 기대 이상으로 높았던 데에서 비롯된 현상으로 보인다. 이 평가에 참여한 예비 교사들은 학부 과정 3학년 학생들로서 쓰기 평가에 요구되는 지식 수준이 어느 정도 확보되어 있는 상태였던 것으로 예상된다. 이에 따라 실제적인 평가 경험은 없었지만, 이들이 가지고 있는 쓰기 평가 지식이 평가에 반영됨으로써 경험 부족에서 나타날 수 있는 평가 주관성이 해소될 수 있었다고 해석할 수 있다.

이러한 결과는 두 가지 점에서 중요한 시사를 제공한다. 첫째는 학생이 작성한 글이 일정한 수준에 도달해 있다면 평가자 훈련이 없더라도 매우 근사한 평가 결과에 도달하게 된다는 점이고, 둘째는 학생이 작성한 글이 일정한 수준에 도달해 있다면 현직 국어교사의 평가와 예비 국어교사의 평가가 크게 다르지 않다는 점이다. 평가자 협의는 평가의 객관성을 위해 매우 강조되는 절차임에도 불구하고 이 연구에서는 큰 영향력이 발견되지 않았다. 그러므로 평가자 협의가 필요한 상황과 조건을 더 구체적으로 탐색하고 논의할 필요가 있다. 또한, 현직 국어교사와 예비 국어교사의 평가 결과가 큰 차이가 없었던 이유에 대해서도 반복된 대비 연구가 이루어져야 할 것이다. 전체적으로 평가에 차이가 없었지만, 특징적인 것은 표현, 형식 어법, 단어 선택에서 예비 교사들이 엄격한 평가를 하는 경향성이 포착되었다는 점이다. 이러한 차이가 통계적으로 유의한 것은 아니나 예비 교사들은 현직 교사들에 비해 쓰기 평가 측면에서 표현이나 단어 선택과 같은 표현적인 요소와 형식 및 어법과 같은 문법적 요소에서 더욱 엄격한 성향을 보이고 있음을 뜻한다. 예비 국어교사들이 이러한 요소들에 대해 더 엄격함을 보이는 이유에 대해서도 후속적인 논의가 이루어질 필요가 있다.

### 3. 쓰기 평가 결과의 학생별 비교

현직 국어교사와 예비 국어교사의 평가 결과 비교를 보다 구체화하기 위해 학생별로 평가 점수의 차이가 있는지를 검증하고자 한다. 학생 각각에 대해 현직 국어교사 집단과 예비 국어교사 집단이 평가 차이를 보이고 있는지를 검토하는 것이다. 학생별로 두 집단의 평가가 실제 어떤 양상을 보이는지를 살펴보는 것은 의미 있는 일이다. 두 집단이 공통적으로 지향하고 있는 평가관이 어떠한지 분석할 수 있기 때문이다. 또한 평가에서 차이를 보이는 학생들의 글을 분석한다면, 현직 국어교사의 쓰기 평가와 예비 국어교사의 쓰기 평가에 있어 핵심적인 특징들을 추출할 수도 있을 것이다.

학생별로 두 집단의 쓰기 평가의 차이를 분석하기 위해 독립표본 t 검증을 실시하였다. 평가 결과는 쓰기 평가의 총점 및 쓰기 평가 하위 다섯 개 요소로 나누어 비교하였다. <표 11>은 쓰기 평가의 총점에서 차이가 있는 학생들의 결과를 모은 것이다.

<표 11> 총점에 대한 학생별 쓰기 평가 차이

변인	교사	표본크기	평균	표준편차	t
학생1	현직 교사	49	18.3265	2.69558	2.626 (p=.011)
	예비 교사	25	16.2800	3.95306	
학생6	현직 교사	49	19.3878	3.86661	3.003 (p=.004)
	예비 교사	25	16.4000	4.38748	
학생13	현직 교사	49	10.3469	3.19864	2.552 (p=.013)
	예비 교사	25	8.4400	2.69382	
학생16	현직 교사	49	14.0612	3.00283	2.696 (p=.009)
	예비 교사	25	12.0400	3.14219	
학생33	현직 교사	49	14.8163	3.88197	2.795 (p=.007)
	예비 교사	25	12.3200	3.07842	
학생37	현직 교사	49	15.3469	3.57393	2.071 (p=.042)
	예비 교사	25	13.5200	3.61847	

<표 11>에서 확인할 수 있는 바와 같이 총점에서 두 집단 간에 통계적으로 유의미한 차이를 보인 학생 글은 1, 6, 13, 16, 33, 37번이다. 현직 국어교사가 예비 국어교사에 비해 통계적으로 유의한 수준의 높은 점수를 부여하였다.

쓰기 평가의 하위 요인 중 내용 요인에서는 다섯 건의 학생 글에서 차이가 있었다. 학생 1, 6, 16, 29, 33이 이에 해당한다. 내용 요인에서도 현직 국어교사의 평균 점수가 통계적으로

유의하게 높았다. 조직 요인 통계적으로 유의한 차이가 나타난 학생 글은 일곱 건이었다. 학생 6, 10, 13, 17, 19, 33, 37이 이에 속한다. 이 경우에도 현직 국어교사의 평가 점수가 더 의미 있게 높았다. 표현 요인에서 차이가 나타난 학생 글은 네 건이었다. 6번, 13번, 26번, 33번 학생 글이 차이를 보였으며, 이 요인에서도 현직 국어교사가 예비 국어교사보다 통계적으로 더 높은 점수를 부여하였다. 형식 어법 요인에서 통계적으로 유의차이가 나타난 차이를 보인 학생 글은 여덟 건이었다. 1번, 6번, 8번, 16번, 18번, 29번, 34번, 37번이 이에 해당한다. 형식 어법은 총점 평균 차이가 통계적으로 유의하게 나타났기 때문에 이를 고려하여 차이를 살펴볼 필요가 있다. 형식 어법 요인에서 평가 간 차이가 나타난 이유를 학생 글의 탐색 과정에서 파악할 수 있기 때문이다. 다음으로 단어 선택에서의 차이를 보인 학생 글은 1, 4, 13, 19번이다. 단어 선택에서도 현직 국어교사들의 평가 점수가 더 높았다.

총점 및 평가 요소별로 두 집단 사이의 차이를 보인 학생 글을 정리하면 <표 12>와 같다. <표 12>에 따르면 평가 집단에 따라 점수 차이가 발생하는 학생 글이 반복적으로 나타나는 현상을 발견할 수 있다. 학생 1, 학생 6, 학생 13, 학생 33 등은 총점뿐만 아니라 평가 하위 요인별 비교에서도 통계적으로 유의한 차이를 초래하고 있다. 그러므로 이러한 반복적으로 차이를 불러오는 학생 글을 분석함으로써 학생 글이 어떤 특징을 가지고 있을 때 평가 차이, 즉 일관성을 떨어뜨릴 수 있는지를 판단해 볼 수 있다.

<표 12> 쓰기 평가 차이의 학생별 빈도

일련번호 평가 요인	1	2	3	4	5	6	7	8
총점 (6명)	학생1 t=2.626, p=.011	학생6 t=3.003, p=.004	학생13 t=2.552, p=.013	학생16 t=2.696, p=.009	학생33 t=2.795, p=.007	학생37 t=2.071 p=.042	.	.
내용 (5명)	학생1 t=3.083, p=.003	학생6 t=2.817, p=.006	학생16 t=2.257, p=.027	학생29 t=2.263, p=.029	학생33 t=2.704, p=.009	.	.	.
조직 (7명)	학생6 t=3.081, p=.004	학생10 t=2.216, p=.034	학생13 t=2.045, p=.045	학생17 t=2.556, p=.013	학생19 t=2.152, p=.035	학생33 t=2.030, p=.046	학생37 t=2.520, p=.014	.
표현 (4명)	학생6 t=2.380, p=.020	학생13 t=2.020, p=.047	학생26 t=2.231, p=.029	학생33 t=2.951, p=.004	.	.	.	.
형식 어법 (8명)	학생1 t=2.733, p=.008	학생6 t=2.236, p=.028	학생8 t=2.496, p=.015	학생16 t=2.855, p=.006	학생18 t=2.005, p=.049	학생29 t=2.314, p=.026	학생34 t=2.754, p=.007	학생37 t=2.820, p=.006
단어 선택 (4명)	학생1 t=2.313, p=.024	학생4 t=2.728, p=.008	학생13 t=2.599, p=.011	학생19 t=2.570, p=.012	.	.	.	.



〈표 12〉에서 총점을 중심으로 검토하면, 학생 글에 대한 평가에서 다음과 같은 특징을 발견할 수 있다. 전반적으로 현직 국어교사가 예비 국어교사에 비해 더 높은 점수를 부여하였다. 학생 1은 내용, 형식 어법, 단어 선택 요인에서 현직 국어교사가 더 높은 점수를 부여하였으며, 학생 6은 내용, 조직, 표현, 형식 및 어법 요인에 대해서 더 높은 점수를 부여하였다. 학생 13의 경우는 조직, 표현, 단어 선택 요인에 대해서, 학생 16의 경우는 내용, 형식 어법 요인에 대해서 현직 국어교사의 점수가 더 높게 나타났다. 학생 33은 내용, 조직, 표현 요인에서 더 높았고, 학생 37번의 경우는 조직과 형식 요인에서 현직 교사의 점수가 더 높게 나타났다.

학생 글을 통해 각 평가 요소별 차이를 살펴보면, 현직 국어교사들은 학생 글의 형식이나 문단 조직, 단어 선택이나 맞춤법 요인보다는 학생이 자신의 주장을 뒷받침할 수 있는 근거가 풍부하거나 합리적일 때 더욱 높은 점수를 부여하는 것으로 분석되었다. 이에 비해 예비 국어교사들은 중학생들이 쓴 글임에도 불구하고 엄격하게 평가하는 경향이 있었다. 학생 글을 평가할 때 내용 요인에 대한 기준이 매우 높아, 현직 국어교사들이 풍부한 예시나 합리적인 근거로 본 경우에도 상대적으로 낮은 점수를 부여하였다. 또한 맞춤법, 띄어쓰기, 문단 배열, 단어 선택에 대해서도 평가 기준의 적용이 매우 엄격한 것으로 나타났다.

## V. 결론

이 글은 쓰기 평가 연구에서 조명되지 못했던 평가자 요인을 연구의 중점으로 삼아, 현직 국어교사와 예비 국어교사의 쓰기 평가의 차이를 비교·분석하였다. 쓰기 평가에 대한 두 집단의 차이를 비교·분석하기 위하여 중학생이 작성한 39편의 글을 직접 평가하도록 하고 그 결과를 수집하여 처리하였다. 평가 전이나 중에 평가자 협의는 시행하지 않았다. 현직 국어교사는 49명이, 예비 국어교사는 24명이 참여하였다. 이 연구 과정을 통해 얻은 결과는 다음과 같다.

첫째, 중학생의 논술문을 평가한 결과는, 평가자 협의를 거치지 않았음에도 불구하고 현직 국어교사와 예비 국어교사 사이에 매우 높은 상관도를 보였다. 평가자 협의는 평가자간 신뢰도를 높이는 데 기여하는 것으로 알려져 있는데, 이러한 결과는 평가자 협의의 효과를 재검토해야 할 필요성을 제기한다.

둘째, 쓰기 평가에 대한 현직 국어교사와 예비 국어교사의 높은 상관도에도 불구하고 두 집단 간에는 차이도 발견되었다. 전체 평균 비교에서 형식 어법 요인에서 통계적으로 유의한 차이가 발견되었다. 예비 국어교사들이 쓰기 평가 점수를 현직 국어교사보다 낮게 부여하는 현상이 확인되었다. 사실, 예비 국어교사들은 대부분의 쓰기 평가 요인에서 현직 국어

교사들보다 낮은 점수를 부여하였다. 이를 통해 예비 교사들이 더 엄격하게 평가 기준을 적용하고 있음을 알 수 있다.

셋째, 현직 국어교사와 예비 국어교사의 평가 총점을 비교한 결과, 현직 국어교사는 평가의 중점을 글의 내용에 두고 있다는 점을 확인하였다. 현직 국어교사는 내용에 높은 점수를 줄 경우, 형식 어법 및 단어 선택에서도 높은 점수를 부여하였다. 이에 비해, 예비 국어교사는 내용 요소와 형식 요소를 균등한 관점에서 평가하고 있음을 알 수 있었다. 이에 따라 맞춤법과 같은 형식 어법에서 낮은 점수를 부여하였다.

이 연구의 결과는 국어교사의 쓰기 평가 방안 및 연구에 대한 성찰의 계기를 제공한다. 우선, 교과 전문가인 현직 국어교사와, 쓰기 지도 및 쓰기 평가 경험이 전혀 없는 예비 국어교사가 유사한 평가 결과를 보였기 때문이다. 또한, 평가자 협의를 거치지 않았음에도 불구하고 평가 결과가 집단 간 및 집단 내 차이를 보이지 않은 것으로 확인되었기 때문이다. 이를 통해서 볼 때, 쓰기 평가에 차이를 초래하는 조건과 환경, 원인을 여러 각도로 탐색해 볼 필요성이 제기된다. 평가 차이의 원인과 조건, 환경을 탐색함으로써 평가 결과의 차이를 줄이는 방안을 강구할 수 있다.

평가자 협의는 그간 쓰기 평가에서 평가자의 주관성을 제한하고 평가의 객관성을 높이는 주요한 장치로 인정되어 왔다. 그러나 평가자 훈련은 대단위로 평가하거나 교실에서 학생의 쓰기 수행을 평가할 때에는 효율적이지 못한 방법일 수도 있다. 이 연구의 결과를 통해 본다면, 평가자 훈련이 반드시 필요한 것은 아니라고 판단할 수도 있다. 그러므로 평가자 협의의 방법 이외에 평가자의 주관성을 극복할 수 있는 대안적 방법에 대한 필요성이 요청된다. 구체적인 평가 기준 설정 외에 주목할 수 있는 대안적 방법으로는 평가 예시문 활용을 고려해 볼 수 있다. 평가 예시문은 학생이 작성한 글로서 평가 기준을 반영하고 있는 예시적인 글을 일컫는다. 총체적 결과 평가에서는 평가자의 해석적 차이를 줄이기 위하여 평가 예시문을 선정·활용하도록 권장하고 있는데, 이를 더욱 확대함으로써 평가자의 일치도를 높이는 쓰기 평가를 기획할 수 있다. 평가 예시문은 평가 기준을 구체적으로 예시하고 있으므로 평가 예시문과 비교·대조하여 학생 글을 평가함으로써 더욱 높은 수준의 일치도에 도달할 수 있기 때문이다.

후속적인 논의에서는 쓰기 평가의 주체인 국어교사의 쓰기 신념 요인, 평가자의 평가 요소 중요도 인식 등에 대한 분석도 이루어져야 한다. 쓰기 평가에서 평가자의 주관성을 제한하고 평가 타당성을 확보하기 위해서는 쓰기 평가에 영향을 미치는 평가자 요인에 대한 연구가 이루어져야 하기 때문이다. 평가자 요인에 대한 정보를 갖추고 있어야 이를 효과적으로 통제할 수 있다. 동일한 글을 평가하였을 때 발생하는 쓰기 평가 결과의 차이가 평가자의 어떤 특성에서 비롯되는지 알 수 있다면 쓰기 평가에 관한 교사 교육 프로그램을 구성하는 데에도 도움이 될 수 있다.

## 참 고 문 헌

- 가은아(2008). 국어교사의 쓰기 효능감 및 쓰기 신념이 학생 쓰기 평가에 미치는 영향. 석사 학위 논문, 한국교원대학교.
- 김라연(2007), 총체적 쓰기 평가와 분석적 쓰기 평가의 상관 연구-고등학생의 논설문 평가를 중심으로. **이중언어학**, 35, 103-126.
- 박영민(2007), 예비 국어교사의 쓰기 동기에 영향을 미치는 요인. **새국어교육**, 75, 163-192.
- 박영민(2008), 국어교사의 쓰기 동기 구성 요인. **배달말**, 42, 305-339.
- 이성영(2005). 국어과 교사의 쓰기 영역 평가 전문성 기준과 모형. **국어교육**, 117, 363-376.
- 민태범(1991). 채점자의 인지 양식이 논술고사 채점에 미치는 영향. 석사학위 논문, 한남대학교.
- Adams, G.R., & Cohen, A.S. (1976). An examination of cumulative folder information used. by teachers in making differential judgments of children's abilities. *The Alberta Journal. of Educational Research*, 22(3), 216-225.
- Daly, J.A. & Dickson-Markham. F. (1982), Contrast effects in evaluating essays. *Journal of Educational Measurement*, 19(4), 309-316.
- Dion, K. (1972). Physical attractiveness and evaluation of children's transgressions. *Journal of Personality and Social Psychology*, 24, 207-213.
- Eckes, T. (2008). Rater types in writing performance assessments : A classification approach to rater variability. *Language testing*, 25(2), 155-185.
- Engelhard, G. Jr. (1994). Examining rater errors in the assessment of written composition with a many-faceted Rasch model. *Journal of Educational Measurement*, 31(2), 93-112.
- 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.
- Hayes, J.R. & Bajzek, D. (2008). Understanding and reducing the knowledge effect: Implication for writers. *Written communication*, 25(1), 104-118.
- Mashall Jon C. & Powers, Jerry M. (1969). Writing neatness, composition errors, and essay grades. *J Educ Meas*, 6(2), 97-101.
- Lee, H. K. (2004). A Comparative Study of ESL Writers' performance in a paper- based and a computer-. delivered writing test. *Assessing Writing*, 9, 4-26.
- 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.
- Rich, J. (1975). Effects of children's physical attractiveness on teachers' evaluations. *Journal of Educational Psychology*, 67, 599-609.
- 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*, 151-185.
- Runder, L. M. (1992). Rudner, Lawrence M. (1992). Reducing errors due to the use of judges. *Practical Assessment, Research & Evaluation*, 3(3).
- Tollefson, N. & Tracy, D. B. (1979). *Response length and quality in the grading of essay tests*. Paper presented at annual meeting of the National Council on Measurement in Education, San Francisco, CA.
- White, M. J & Brunning, R. (2005). Implicit writing beliefs and their relation to writing quality. *Contemporary Educational Psychology*, 30(2), 166-189.

• 논문 접수 : 2008년 12월 29일 / 수정본 접수 : 2009년 2월 6일 / 게재 승인 : 2009년 2월 20일

## ABSTRACT

### A Comparative Study of Writing Assessment between In-service Teacher and Pre-service Teacher of Korean

Young-Min Park(Assistant Professor, Korea National University of Education)

Sook-Ki Choi(Lecturer, Korea National University of Education)

This study was comparative analysis of writing assessment between in-service teacher and pre-service teacher of Korean based on writing rater factor. 39 middle school students' argumentative essays were assessed by all 73 teachers who made up of 49 in-service teacher and 24 pre-service teacher of Korean. They assessed 39 argumentative essays equally and did not discuss about the assessment previously. So, they did not be affected by assessor discussion. Then, the result of argumentative essays assessment were analysed to find differences between two assessor groups. The results are following.

First, in the assessment result of middle school students' argumentative essays, Although assessor did not discuss about their assessment, in-service teacher and pre-service teacher of Korean had high correlation in the assessment result of middle school students' argumentative essays.

Second, although between two teacher group founded high correlation in assessment result, their assessment difference of formal grammar were have statistical significance. Because pre-service teacher of Korean had strict writing assessment perspective, they rated the essays lower than in-service teacher of Korean generally.

Third, for comparative analysis of assessment result, it was founded that when assessed, in-service teacher focused content of essay, but pre-service teacher focused content and grammar such as spell, punctuation equally. So pre-service teacher graded lowly for grammar factor in writing assessment.

The result of this study suggested reflective perspective about in-service teachers' specificity of writing assessment and writing assessment method. And further studies needed are effect test of assessor discussion in writing assessment and affect analysis of teacher of Korean as assessor' belief, correlation of between writing assessment and assessor discussion.

Key words : writing assessment, essay assessment, writing rater, writing assessment comparison