

예비 국어교사의 쓰기 평가 효능감과 쓰기 평가 수행에 대한 연구*

박 찬 흥(전주대 조교수)

< 목 차 >

- I. 서론
- II. 이론적 배경
- III. 연구 방법
- IV. 연구 결과 및 논의
- V. 결론

국문초록

본 연구에서는 예비 국어교사들의 쓰기 평가 전문성 신장을 위한 실증적인 합의점을 찾기 위해 쓰기 평가 효능감에 주목하였다. 구체적으로, 예비 국어교사들의 쓰기 평가 효능감, 쓰기 평가 과정에서 겪는 어려움과 쓰기 평가 수행 경험에 대한 의견, 쓰기 평가 결과에서 나타나는 일관성과 엄격성 등의 변인을 수집하고자 하였다. 이를 통해 쓰기 평가 효능감이 예비 국어교사들이 보여주는 쓰기 평가 변인과 어떠한 관계를 맺고 있는지를 확인하는 것이 이 연구의 1차적인 목적이다. 그리고 이를 통해 국어교사 양성과정에서 어떠한 점이 고려되어야 하는지에 대해 논의함으로써 예비 국어교사 양성과 관련한 함의를 찾고자 하였다. 집단 비교를 통해 쓰기 평가 효능감과 엄격

* 본 연구는 2019년 8월 22일 제66회 국어문화회 전국학술대회에서 발표한 내용을 수정 및 보완한 것이다.

성, 평가 범주별로 주관적으로 느끼는 어려움 및 실제 채점 수행의 적합도 등을 살펴본 결과, 쓰기 평가 효능감과 쓰기 평가에서 나타나는 엄격성(적합도)에서는 관련성이 확인되지 않았다. 또한 평가 범주별로 확인된 적합도와 예비 국어교사들이 느끼는 평가 범주별 모호함의 정도 역시 밀접한 관련이 확인되지 않았다. 이러한 연구 결과를 보다 면밀히 해석하기 위해 예비 국어교사들이 쓰기 평가 수행 과정에서 느낀 어려움을 자유기술식으로 반응하게 하였고, 이러한 반응 중 연구 주제와 관련된 요소들을 살펴보았다. 자유기술식 반응을 탐색한 결과, 현재 사범대에서 제공되고 있는 쓰기 평가 경험만으로는 쓰기 평가 효능감이나 쓰기 평가 전문성을 형성하는 것이 충분하지 않다는 점을 확인할 수 있었다.

핵심어 : 쓰기 평가 효능감, 쓰기 평가 전문성, 일관성, 엄격성, 예비 국어교사, 다국면 Rasch 모형

I. 서론

쓰기 평가 전문성은 국어교사가 갖추어야 할 다양한 전문성을 구성하는 주요 요인 중 하나이다. 하지만 쓰기 평가 전문성을 확보하는 것은 결코 쉬운 일이 아니다. 그렇기 때문에 예비교사 양성 과정에서는 쓰기 평가 전문성 신장을 위한 다양한 도움이 제공될 필요가 있다. 하지만 구체적으로 어떠한 도움을 어떠한 방식으로 제공해야 하는지에 대해서는 실증적으로 밝혀진 바가 충분하지 않다.

본 연구에서는 예비 국어교사들의 쓰기 평가 전문성 신장을 위한 실증적인 함의점을 찾기 위해 쓰기 평가 효능감에 주목하였다. 쓰기 평가 효능감은 평가자가 스스로의 평가 수행 능력에 대해 갖는 주관적인 인식을 의미한다.

선행 연구들을 통해 쓰기 평가 효능감은 쓰기 평가 전문성 신장과 밀접한 관련을 맺고 있는 요인으로 알려져 있으며, 쓰기 평가 효능감 신장을 위한 도움이 제공되어야 함이 보고된 바 있다. 하지만 구체적으로 예비 국어교사들이 갖고 있는 쓰기 평가 효능감에 대해서는 충분한 연구가 이루어져 왔다고 보기 어렵다.

이러한 맥락에서, 본 연구는 예비 국어교사들의 쓰기 평가 효능감, 쓰기 평가 과정에서 겪는 어려움과 쓰기 평가 수행 경험에 대한 의견, 쓰기 평가 결과에서 나타나는 일관성과 엄격성 등의 변인을 수집하고자 하였다. 이를 통해 쓰기 평가 효능감이 예비 국어교사들이 보여주는 쓰기 평가 변인과 어떠한 관계를 맺고 있는지를 확인하는 것이 이 연구의 1차적인 목적이다. 그리고 이를 통해 국어교사 양성과정에서 어떠한 점이 고려되어야 하는지에 대해 논의함으로써 예비 국어교사 양성과 관련한 함의를 찾고자 하는 것이다.

II. 이론적 배경

가. 쓰기 평가 효능감

학생 필자들이 작성한 글을 채점하는 것은 결코 쉽고 편한 작업이 아니다. 전문적인 필자들이 작성한 글을 읽을 때는 몰입한다거나 일정한 재미에 빠져 쉽게 헤어나오지 못하는 경우도 종종 있지만, 학생 필자의 글을 읽으면서 이러한 경험을 하는 경우는 몹시 드물다. 반대로 크고 작은 오류들을 찾아내면서 학생 필자의 쓰기 능력 신장을 고민하며 읽다보면 흐름이 끊기기 일쑤이고, 이는 채점을 하는 국어교사에게 상당한 부담으로 다가온다. 이러한 부담은 비교적 긴 시간에 걸친 인지적인 부담으로 작용하기도 하고, 국어교사의 정서적인 면에도 부정적인 영향을 끼칠 수 있다. 그렇기 때문에 쓰기 평가자로서의 국어교사가 갖고 있는 쓰기 평가와 관련된 정의적 요인도 주

목해 볼 필요가 있는 것이다.

쓰기 평가 국면에 영향을 미치는 국어교사의 정의적 요인은 매우 다양하겠지만, 본 연구에서는 특히 쓰기 평가 효능감을 중점적으로 살펴보고자 하였다. 쓰기 평가 효능감은 Bandura(윤운성 외 역, 2003)의 자기 효능감(self-efficacy) 개념을 쓰기 평가 분야에 적용한 것이다. 자기 효능감은 특정한 상황에서 자신이 적절한 행동을 함으로써 문제를 해결할 수 있다고 믿는 신념 또는 기대감을 의미한다. 보다 구체적으로는, 자신이 적절한 수행을 할 수 있다고 믿어야 하며, 동시에 자신에게 주어진 상황이나 과업이 자신의 수행에 따라 적절히 달라질 것이라고 믿어야 한다.

일반적으로 자기 효능감은 특정 분야에서 높은 성취 수준을 보이는 것과 매우 밀접한 관련이 있는 것으로 알려져 있다. 어떤 분야에서 높은 성취 수준을 보이기 위해서는 주의와 집중, 끈기 등을 필요로 하는데, 이러한 요인들이 발현되기 위해서는 자기 효능감을 토대로 하는 것이 중요하다는 것이다. 물론 반대로 자기 효능감이 높다는 것이 반드시 높은 수준의 성취를 보장하는 것은 아니다. 하지만 자기 효능감이 성취 행동의 중요한 매개 변인으로 작용한다는 점은 일반적인 교육심리 이론에서 정설로 굳어지고 있다(임규혁 외, 2007).

쓰기 평가 효능감은 이 개념을 ‘쓰기 평가’라는 특정 상황에 적용한 것으로, 쓰기 평가 상황에서 자신이 적절한 쓰기 평가 수행을 할 수 있고, 이를 통해 학생의 글을 적절히 평가할 수 있다고 믿는 스스로에 대한 신념 또는 기대감이라고 할 수 있다(박영민, 2010). 쓰기 평가 효능감이 높은 교사는 인지적으로 상당히 높은 부담이 필요한 쓰기 평가에 보다 집중할 가능성이 높고, 학생 필자의 글이 갖고 있는 수준이나 특성을 찾아내기 위한 주의를 보다 지속적으로 기울일 수 있을 것이라 기대할 수 있다. 더불어 보다 나은 쓰기 평가를 수행하기 위한 노력을 게을리 하지 않으리라 예측할 수 있다.

쓰기 평가는 비교적 오랜 시간동안 상당한 정도의 집중을 해야 하는 과업이다. 또한 능숙한 수행을 위해서는 매우 다양한 변인들이 적절히 구현되어야 한다. 이러한 점을 고려해 본다면 쓰기 평가 효능감이 높다면 평가 과정

에서 쓰기 평가 효능감이 높은 교사는 평가 과정에서 쓰기 평가 자체를 기피하지 않고 쓰기를 활용한 수업에 적극적일 것이며, 평가 과정에서의 자기점검에도 능동적일 것이라 추측할 수 있다.

그런데 자기 효능감과 같은 정의적 특성은 짧은 시간에 형성되는 것이 아니라는 점에 주의할 필요가 있다. 자신이 무엇을 해야 하고, 실제로 할 수 있으리라 생각을 해 보지 않은 상황에서 영역 특수적인 효능감이 발생할 것이라고는 기대하기 어렵다는 것이다. 다시 말하면 ‘나도 저렇게 할 수 있어!’라는 마음을 품게 되기까지 상당한 정도의 경험과 더불어 긍정적인 체험이 필요하다는 것이다. 자신이 잘 모르거나 경험을 거의 해 보지 않은 생소한 분야에 대해 효능감을 갖게 되는 것이 흔한 일이 아니기 때문이다.

자기 효능감이 장기간에 걸쳐서 형성되는 것이라는 점에 비추어 보면, 현재 국어교사 양성 과정에서 쓰기 평가 효능감이 충분히 갖추어질 수 있는 환경인지에 대해서는 비판적으로 바라보아야 한다. 예비 국어교사들은 입직하는 순간부터 경력이 많은 국어교사와 큰 차이가 없는 쓰기 평가 업무를 담당하게 된다. 하지만 충분한 쓰기 평가 효능감을 갖추지 못한 상황에서 쓰기 평가가 이루어지는 것은 교사 개인을 위해서도, 학생을 위해서도 결코 바람직하다고 보기 어렵다.

나. 쓰기 평가 전문성

쓰기를 평가하는 것은 결국 학생의 쓰기 능력에 대한 정보를 체계적으로 수집하는 활동이다. 따라서 쓰기 평가를 체계적으로 시행하지 못하면, 쓰기 능력 신장을 위한 수업 프로그램을 올바르게 편성하기 어렵고 올바르게 운영하기도 어렵다. 학교 쓰기 교육에서 쓰기 평가를 중요하게 보는 이유는 바로 이 때문이다. 학교에서 쓰기 평가를 통해 학생의 쓰기 능력을 체계적으로 수집하려면, 국어교사는 쓰기 평가에 대한 전문성을 갖추고 있어야 한다.

하지만 쓰기 수행은 대부분 필자의 머릿속에서 이루어지기 때문에 관찰하는 것이 매우 어려울 뿐만 아니라, 매우 복잡한 인지적, 정의적 반응을 통해 일어난다. 그렇기 때문에 쓰기 수행에 대한 평가를 전문적으로 해 낸다는 것

은 결코 쉬운 일이 아니다(이성영, 2005). 더군다나 학생의 학습이나 성취와 관련된 정보를 수집하고 해석하여 이를 활용하는 것이 쓰기 평가의 궁극적인 목적이라면(이인제 외, 2004), 전문적을 쓰기 평가를 수행해 낸다는 것은 경험이 풍부한 국어교사에게도 결코 쉬운 일이 아니라는 점을 알 수 있다.

쓰기 평가에서는 매우 복합적인 전문성이 요구된다(박영목, 2007:13). 쓰기 평가의 목표와 내용이 적절해야 하고, 평가 방법과 도구 역시 타당하고 신뢰할 수 있어야 한다. 평가의 실시와 채점이 공정해야 하며 평가 결과의 해석과 적용은 정확하고 충실하게 이루어져야 한다. 평가 환경 역시 국어교사가 책임을 져야 하는 부분이다.

이처럼 복합적인 쓰기 평가 전문성에 대해 이성영(2005:360)에서는 그 하위 요인으로 지식과 수행, 태도를 제시한 바 있다. 이 세 가지 요인 중 지식에는 ‘글’, ‘쓰기’, ‘학생’, ‘맥락’에 대한 지식과 같은 것이 있다. 국어교사는 좋은 글의 요건, 쓰기의 본질과 특성, 쓰기의 상황이나 방법과 관련된 지식, 학생 필자의 발달 수준 등과 관련된 지식을 갖추고 있어야 한다. 또한 쓰기 평가에 관여하는 다양한 맥락에 대한 지식도 갖추어야 한다.

태도 요인은 일정한 대상에 대한 태도를 의미하는데, 평가 자료, 학생, 평가 관계자, 자기 자신에 대한 태도 등으로 구분해 볼 수 있고 하였다. 이러한 구분법을 따른다면 평가 자료에 대해서는 성실하게 임할 필요가 있고 학생에게는 공정하게 대할 필요가 있다고 보았다. 또한 쓰기 평가 관계자들과 협동하려는 열린 마음을 가져야 하고, 스스로에 대해서도 자기 발전과 점검을 위해 노력하는 태도를 가져야 한다고 보았다.

그런데 지식과 태도 요인은 결국 수행 요인으로 드러나야 의미가 있다고 볼 수 있다. 수행은 실제로 쓰기 평가를 실천하는 것이기 때문이다. 수행은 지식과는 달리 분절적이거나 언어적인 표현으로 그 실체를 온전히 드러내기 어려우나, 교육적인 목적을 위해서는 수행을 하위 과정으로 쪼개어 접근하는 것이 일반적이다. 이와 관련하여 이인제 외(2004:67-74)에서는 국어교사의 평가 전문성 구성 요소를 ‘수행’에 초점을 두어 다섯 가지로 구분하여 제시한 바 있는데, 그 중에 본 연구에서 주목하고자 하는 것은 평가의 실시.채점.

성적 부여 능력이다. 이 능력 역시 수행을 기반으로 이루어질 필요가 있는데, 핵심은 이러한 평가 수행에 어떠한 요인들이 영향을 주고받았는지 그 실체를 확인하기 어렵다는 것이다.

이와 관련하여 쓰기 평가 수행에 어떠한 요인이 영향을 미칠지 확인하기 위한 탐색적 시도로, 본 연구에서는 쓰기 평가를 마친 뒤에 학생들이 쓰기 평가 과정에 서 겪은 어려움을 묻는 문항에 자유 서술식으로 답을 하도록 하였다. 이를 통해서 무엇이 쓰기 평가 과정에 영향을 미치는지를 살피고자 한 것이다.

다. 쓰기 평가 수행의 적절함을 평가하는 방법

일반적으로 평가가 적절한지 아닌지의 여부는 신뢰도와 타당도의 개념으로 설명된다. 그런데 쓰기 평가를 직접 평가 방식으로 실시할 경우 특히나 신뢰도가 상당한 문제로 제기된다. 쓰기에 대한 직접 평가의 요구가 매우 높은 수준임에도 불구하고, 고부담 시험이나 대규모 평가에서 직접 평가 방식이 사용되지 않고 있는 중요한 이유 중 하나는 신뢰도를 보장할 수 없기 때문이다. 많은 연구들이 쓰기 평가를 할 때 평가자의 개인적인 특성이 평가에 상당한 영향을 미친다는 점을 보고하고 있다. 그리고 이러한 차이점을 해결하고 쓰기 평가의 신뢰도를 높이기 위해 평가자 협의회 등을 통한 평가자 훈련 및 교육, 평가 지침 및 기준 개발 등을 제안하는 다양한 연구들이 수행된 바 있다.

이러한 점을 고려하여, 여기에서는 평가자의 개인적 특성인 엄격성과 일관성에 주목하였다. 쓰기 평가에서의 엄격성이란 쓰기 평가 점수를 부여할 때 엄격하게 평가하는 정도를 말한다. 일반적으로 엄격성이 높다고 한다면 점수를 낮게 주는 경향이 있으며, 엄격성이 낮다고 한다면 점수를 높게 주는 경향이 있다고 말할 수 있다. 또한 일관성이란 학생 글을 평가할 때 채점의 엄격성을 처음부터 끝까지 동일하게 유지하는 정도를 의미한다.

엄격성에 대한 논의가 필요한 이유는, 쓰기 평가의 신뢰성 문제와 직결되어 있기 때문이다. 쓰기 평가를 하면서 평가자가 엄격성을 일정하게 유지한

다는 것은 매우 어려운 일이다. 그렇기 때문에 엄격성이 얼마나 일정하게 유지되는지를 확인하는 것은 쓰기 평가자의 전문성을 확인할 수 있는 중요한 척도로 삼을 수 있다.

쓰기 평가에서 엄격성은 다시 말하면 일관성 있게 유지하는 정도를 의미한다. 그렇기 때문에 일관성은 결국 학생 글을 평가하는 국어 교사가 갖추어야 할 ‘엄격성의 일관성’이라고 할 수 있다(박영민, 2012:267). 이 점에서 보면, 일관성은 엄격성은 개념적으로는 분리할 수 있지만 실제 평가 수행 국면에서는 분리해서 살피기 어렵다는 특성이 있다.

이처럼 엄격성을 일관되게 유지하기 위해서 다양한 논의들이 있어왔으나 여기에서는 다국면 Rasch 모형을 활용하여 엄격성의 일관성을 진단하고 이를 토대로 논의하는 방안을 중요하게 활용하였다. 다국면 Rasch 모형을 활용하면 지나치게 관대한지 엄격한지(엄격성), 엄격성이 일관되게 유지되는지(일관성)를 수리적으로 파악할 수 있다는 점이 다양한 선행연구들을 통해 확인되었다(장은주 2015, 이인혜 2014, 박종임 2013, 권태현 2012, 박영민 2012, 박종임·박영민 2012, 박종임·박영민 2011, 박영민·최숙기, 2010, 박찬홍, 2018 등).

Ⅲ. 연구 방법

가. 연구 대상

이 연구의 목표는 예비 국어교사들의 쓰기 평가 효능감에서 나타나는 특성과 쓰기 평가 수행에서 나타나는 엄격성과 일관성을 파악하여 예비 국어교사 양성과정에 필요한 교육적 함의를 논의하는 것이다. 이를 위하여 2019-1학기 J대학교 사범대학 국어교육 (복수)전공으로 작문교육론을 이수하고 있는 3학년 학생 20명을 연구 대상으로 선정하였다. 총 수강생은 26명이었으나, 누락 데이터가 발생 한 경우 해당 학생의 데이터를 모두 제외하여 총 20명의 데이터만을 분석 대상으로 삼았다. 데이터 분석에 사용된 표집 현

황은 아래 <표 1>과 같다.

	남	여	계
주전공	6	13	19
복수전공	1	-	1
계	7	13	20

<표 1> 연구 대상

이들은 모두 사범대학에 속해 동일학년, 학기를 이수하고 있었고 학기 초에 실시한 진단을 통해 쓰기 평가 경험이 거의 없었다는 점을 확인하였다. 또한 작문교육론 수업을 통해 쓰기 평가 관련 항목들에 대한 수업을 진행하였기 때문에 연구 데이터를 수집한 학기말에는 쓰기 평가 역량 차원에서 비교적 동일한 집단 특성을 갖고 있다고 판단하였다. 또한 학생 글을 평가하거나 쓰기 평가 효능감에 응답할 때에는 다른 학생들과 최대한 협의를 하지 않고 개별적으로 응답할 수 있도록 안내하였다. 이들은 평가에 임하기 전에 쓰기 평가 효능감 검사에 응답하였고, 응답을 마친 뒤에 쓰기 평가 수행을 시작하였다.

나. 검사 도구

(1) 쓰기 평가 효능감 검사지

쓰기 평가 효능감과 같은 정의적인 요인은 관찰이나 주관적 판단을 통한 측정이 쉽지 않기 때문에 ‘검사 도구’를 통해서 양적으로 파악할 수 있다(박영민 외, 2016:405). 쓰기 평가 효능감 검사지는 박영민(2010)에서 예비 교사 321명을 대상으로 쓰기 평가 효능감을 검사하기 위해 개발된 것을 사용하였다. 이는 국어교사 133명을 대상으로 쓰기 평가 효능감을 조사·분석한 연구(박영민 2011), 국어교사 93명을 대상으로 한 연구(장은주 2015), 예비 국어교사 85명을 대상으로 한 연구(가은아, 2018)에서도 사용되어 충분히 타당화 과정을 거친 것으로 보았다. 따라서 이 연구에서 예비검사를 따로 실시하지

않고 그대로 활용하였다. 쓰기 평가 효능감 검사 도구는 모두 18개의 문항으로 구성되었으며, 쓰기 평가 효능감 요인은 2개로, 제1요인은 ‘일반적 평가 효능감’이며 제2요인은 ‘실행적 평가 효능감’이다.²⁾ 제1요인은 1~5번, 제2요인은 6~18번에 해당한다.

(2) 쓰기 평가 수행 자료

이 연구에서는 설명문 과제 문항³⁾에 따라 작성된 중학생의 설명문 30편을 채점하도록 하였다. 학생의 글씨 요인이 평가에 미치는 영향을 통제하기 위해 학생 글을 워드프로세서로 입력한 뒤 출력한 종이를 예비 국어교사들

-
- 2) 쓰기 평가 효능감 검사지의 문항은 다음과 같다. 1. 나는 학생 글을 잘 평가(채점)할 수 있다. 2. 나는 학생들의 쓰기 능력을 정확하게 평가할 수 있다. 3. 나는 높은 신뢰도 수준을 유지하면서 학생 글을 평가할 수 있다. 4. 나는 학생 글을 ‘잘 쓴 글’과 ‘잘 못 쓴 글’로 잘 변별하며 평가할 수 있다. 5. 나는 학생 글을 평가할 때 평가 기준을 일관성 있게 적용할 수 있다. 6. 나는 채점의 엄격성(점수를 잘 주는지 아닌지)을 일관성 있게 유지하며 학생 글을 평가할 수 있다. 7. 나는 학생 글에서 ‘내용’과 관련된 요소(명료성, 통일성, 창의성, 풍부성 등)를 잘 평가할 수 있다. 8. 나는 학생 글에서 ‘조직’이나 ‘구성’이 적절한지를 잘 평가할 수 있다. 9. 나는 학생 글에서 ‘단어 선택’이 적절한지를 잘 평가할 수 있다. 11. 나는 학생 글에서 어법(구두점, 띄어쓰기, 정서법 등)이 정확한지를 잘 평가할 수 있다. 12. 나는 쓰기 평가와 관련하여 어떤 문제나 갈등이 발생했을 때 이를 적절하게 해결할 수 있다. 13. 나는 학생 글을 평가할 때 쓰기 윤리를 위반했는지(예, 표절)를 가려낼 수 있다. 14. 나는 다른 동료 평가자와 협의하며 학생 글을 평가할 수 있다. 15. 나는 쓰기 평가의 상황을 고려하여 평가 과정과 평가 결과를 조절할 수 있다. 16. 나는 평가 영향 요인(예, 글씨 모양, 학생 성별 등)을 적절히 통제하며 학생 글을 평가할 수 있다. 17. 나는 마음이 편한 상태에서 학생 글을 읽고 평가할 수 있다. 18. 나는 쓰기 평가와 관련하여 어떤 문제가 발생했을 때 문제의 원인이 어디에 있는지 파악할 수 있다.
- 3) 생활하다 보면 내가 아는 정보를 다른 사람에게 설명해야 하는 기회가 많습니다. 나는 잘 알고 있는 내용이지만 다른 사람은 그것에 대해 잘 모르다면 어떻게 설명해야 할까요? 지금부터 내가 잘 아는 사물, 운동 경기, 음식 중에서 하나를 골라 그것에 대해 잘 모르는 사람에게 설명문을 써 주세요. 내 글을 읽는 사람이 쉽게 이해하고 오래 기억할 수 있도록 씁니다.(사물, 운동경기, 음식에 속하거나 관련된 글감도 가능합니다. 예, 축구, 김밥 등)

에게 제공하였다. 글에 나타난 맞춤법이나 띄어쓰기 등의 오류는 수정하지 않고 그대로 입력하였다. 여기에서 사용된 분석적 평가 기준은 Spandel & Culham(1996)에서 제시한 6특성(trait) 평가 요소 중 일부를 수정하여 다섯 가지로 제시한 것이다.⁴⁾ 채점 기준은 5점 척도를 사용하였고, 각 채점 요소에 따라 5, 3, 1점에 해당하는 글의 특성을 간략하게 제시하였다.

채점 결과 보고의 뒷부분에는 “학생 글을 평가할 때 평가 범주별로 점수 결정이 어려웠던 정도는 어느 정도였나요?”라는 문항을 통해 다섯 가지 평가 범주 중 어떤 범주가 어려웠는지에 6점 척도로 답하게 하였다. 그리고 마지막으로 쓰기 평가 과정에 대한 주관적인 경험을 확인하고 어떠한 요인들이 쓰기 평가 수행에 영향을 미쳤을지를 파악하기 위한 탐색적 시도로 다음 질문에 자유기술형으로 답하게 하였다. “전반적인 평가 과정에서 어려웠던 것은 무엇입니까?” 본 연구는 이러한 문항에 대한 자유기술식 반응을 확인함으로써 쓰기 평가 효능감이나 쓰기 평가 전문성에 영향을 미치는 요인들이 무엇인지 개괄적으로나마 탐색할 수 있을 것이라 보았다.

다. 연구 절차

설문조사 채점 과정에서 글의 유형과 분석적 채점의 특성, 방법, 하위 기준에 대해 간략히 안내하였다. 또한 이들은 작문교육론을 수강하면서 쓰기 평가의 이론과 원리에 대한 내용을 학습하였다. 하지만 쓰기 평가 효능감에 따라 평가 경험이 많지 않은 채점자가 갖고 있는 특성을 파악하는 것이 본 연구의 주요 목적이기 때문에 채점자 훈련은 별도로 실시하지 않았다. 예비 국어교사들은 2019년 6월 13일부터 28일까지 약 2주에 걸쳐 각자 편한 시간 및 장소에서 평가를 실시했다. 이들은 채점 결과와 함께 설문지를 하나의 워드 파일로 저장하여 제출하였다.

라. 분석 방법

4) 내용, 조직, 표현, 단어, 형식

(1) 쓰기 평가 효능감

쓰기 평가 효능감의 경우 학생들의 응답 자료를 코딩한 후 일반적 평가 효능감과 실행적 평가 효능감의 평균과 표준편차를 산출하였다. SPSS 18.0을 사용하여 집계를 낸 뒤, 쓰기 평가 효능감의 수준에 따른 집단을 상/하로 구분하였다.

(2) 쓰기 평가 수행 자료

예비 국어교사 20명의 분석적 쓰기 평가 수행에서 나타나는 특성을 파악하기 위해 우선적으로 채점 결과 데이터를 FACETS 프로그램을 사용하여 분석하였다. 피험자, 채점자, 평가기준 등 3가지 국면을 설정하여 평가자들의 엄격성과 일관성을 분석하였다. 그리고 이러한 분석 결과를 토대로 엄격성과 일관성 수준에 따라 이들의 평가 과정에 어떠한 특성이 확인되는지를 살펴 보았다. 또한 다섯 가지 평가 범주의 어려움을 묻는 문항에 대한 답변 내용을 별도로 코딩하여 SPSS 18.0을 사용하여 집계하였다.

쓰기 평가 수행 결과물의 마지막에 제시된 세 가지 자유기술식 문항의 경우 연구자가 학생들의 반응을 읽으며 이론적 배경 2절을 통해 제시한 내용 중 쓰기 평가의 지식, 수행, 태도 요인으로 구분하여 분류하였다. 또한 이론적 배경 1절을 통해 제시한 쓰기 평가 효능감과 관련된 언급들도 별도로 구분하여 정리하였다. 이러한 자유기술식 문항의 답변 내용은 계량적인 데이터를 해석하고 향후 논의점을 제안하는 데에 보조 및 참고 자료로서의 역할로 삼고자 하였다.

IV. 연구 결과

가. 쓰기 평가 효능감 기술 통계

문항		평균	표준편차	문항		평균	표준편차
일반적 평가 효능감	1	3.95	0.9734	실행적 평가 효능감	10	4.55	0.8646
	2	3.70	0.9000		11	4.50	0.9747
	3	3.75	0.9937		12	4.05	0.8047
	4	4.70	0.8426		13	3.95	1.0235
	5	4.10	0.7681		14	4.90	0.8307
실행적 평가 효능감	6	4.15	0.9097		15	4.20	0.9274
	7	4.05	1.0712		16	4.50	0.8660
	8	4.30	1.1446		17	4.60	0.8000
	9	4.35	0.9097		18	4.10	0.9434

<표 2> 쓰기 평가 효능감 문항별 기술 통계

쓰기 평가 효능감에 대한 기술 통계 결과는 위의 <표 2>와 같다. 일반적 평가 효능감 평균은 4.04, 표준편차는 .0834로 확인되었다. 또한 실행적 평가 효능감 평균 4.32, 표준편차는 .0994로 나타났으며 18개 문항에 대한 Cronbach α 는 .913을 얻었다. 두 요인의 상관도는 .773으로 0.01수준에서 통계적으로 유의하여 검사도구로서의 양호도는 확인되었다.

국어교사를 대상으로 한 연구인 박영민(2011)의 경우 일반적 평가 효능감과 실행적 평가 효능감은 각각 4.318과 4.410로 나타났으며 장은주(2015)의 경우 각각 4.49와 4.71로 나타났다. 예비 국어교사를 대상으로 한 가은아(2018)의 경우 각각 4.07과 4.32로 나타났다.

가은아(2018)에서는 2번 문항의 평균(3.49)이 가장 낮게 나타났는데, 본 연구의 분석 결과도 마찬가지로의 값을 보여준 부분은 흥미롭다. 해당 문항은 '나는 학생들의 쓰기 능력을 정확하게 평가할 수 있다.'이다. 선행 연구와의 비교 결과, 현직 국어교사보다는 예비 국어교사의 쓰기 평가 효능감이 상대적으로 낮다는 점을 확인할 수 있었다.

일반적 평가 효능감과 실행적 평가 효능감의 대응표본 t검정 결과 유의수

준 .002에서 차이가 있는 것으로 확인되었기 때문에 일반적 평가 효능감보다 실행적 평가 효능감이 더 높은 값을 보인다고 해석할 수 있다. 이는 일반적 효능감의 문항 진술이 전반적으로 추상적이고 쓰기 평가 전반을 다루고 있는 반면에 실행적 효능감의 문항 진술은 전반적으로 구체적이고 세부적인 내용을 다루고 있기 때문인 것으로 보인다. 다른 선행연구들에서도 공통적으로 일반적 평가 효능감에 비해 실행적 평가 효능감이 높다는 점도 확인되었다. 이와 관련하여 일반적 평가 효능감이 낮은 원인에 대한 다양한 연구 방법을 활용한 세밀한 분석이 필요할 것으로 판단된다.

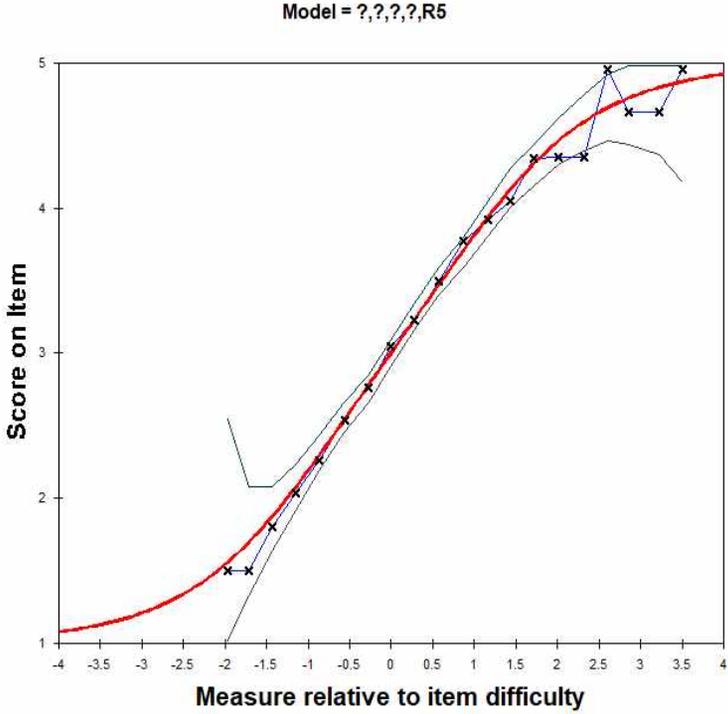
나. 다국면 Rasch 모형 분석 결과

(1) 모형 적합도

FACET 프로그램을 통해 확인할 수 있는 모형 적합도 그래프는 아래 <그림 2>와 같다. 가운데 누운 S자 모형의 곡선이 전반적으로 위/아래 선 가운데에 위치하는 것으로 보아 시각적으로 데이터가 다국면 Rasch 모형으로 분석하기에 적합하다는 것을 확인할 수 있다.⁵⁾

5) 가로축은 문항의 난도와 학생의 능력을 고려한 값을 나타내고 세로축은 그에 따른 기대 점수를 나타낸다. 그래프의 'x'는 예비 국어교사가 부여한 점수이며, 중앙의 굵은 선은 문항 반응 이론에 따른 문항 특성 곡선을 나타낸다. 문항 특성 곡선을 중심으로 양쪽의 가는 선은 95%의 신뢰구간을 의미하기 때문에 x 표시가 위아래의 두 신뢰구간 안에 위치하면 분석이 적합하다고 판단할 수 있다(지은림 외, 2000).

<그림 1> 모형 적합도 그래프



(2) 요인별 분포

측정 (logit)	+ 학생글	-평가자	-평가기준	-성별	척도
2					(5)
	4				4
1	2 12	09			

	8 9 11 18 26 27 29	16	형식	여	
	1 6 15 17	1 18 19	표 현 조직		3
0	14 30	4 11 12	내 용 단어	남	
	3 19 22 23 24 28	2 3 13 14 20			
	7 16 21 25	15 17			
	13 20				
-1	5 10	6 8 10			
		7			2
		5			
-2					(1)

〈표 3〉 학생글 × 평가자 × 평가기준 × 성별 분포도

본 연구에서는 학생 글, 평가자, 평가기준, 성별 4가지 요인을 설정하여 분석하였다. 동일 척도로 해당 요인들의 분포를 확인한 결과는 위의 <표 3>과 같다.⁶⁾ 학생 글만 출력 결과를 내림차순으로 설정하였다. 표를 통해 4번 학생 글이 가장 높은 점수를 받았고 5, 10번 글이 낮은 점수를 받았다는 것을 확인할 수 있다. 또한 9번 평가자가 가장 엄격한 평가자였고, 5번 평가자가 가장 관대하였다. 또한 평가 기준의 경우 형식 요인의 평균 점수가 가장 낮았고, 내용과 단어 범주는 비슷한 수준으로 높은 평균 점수를 보였다. 성별 요인에서는 여학생 평가자가 남학생 평가자에 비해 점수를 낮게(엄격하게) 주었다는 것을 확인할 수 있다. 다만 이러한 결과는 표본 수가 충분하지 않다는 점을 고려하여 제한적으로 해석할 필요가 있다.

(3) 엄격성 수준과 적합도

6) FACET 프로그램은 윈도의 메모장 프로그램을 사용하여 해당 정보를 제공하거나 시각적 편의를 위해 별도의 도표로 해당 데이터를 옮겨 정리하였다.

평가 자	엄격성 (logit)	표준 오차 (SE)	내적합 지수 (Infit MS)	외적합 지수 (Outfit MS)	평가 자	엄격성 (logit)	표준 오차 (SE)	내적합 지수 (Infit MS)	외적합 지수 (Outfit MS)
5	-1.74	.11	1.08	.99	20	-.19	.09	.83	.83
7	-1.21	.10	1.89	2.01	3	-.17	.09	.71	.71
6	-1.03	.09	.57	.59	11	-.05	.09	.93	.93
8	-1.02	.09	1.18	1.17	12	.06	.09	1.73	1.73
10	-1.01	.10	.80	.83	4	.08	.09	.63	.63
15	-.62	.09	.94	.93	19	.14	.09	1.33	1.33
17	-.41	.09	.76	.76	18	.25	.09	1.39	1.39
13	-.36	.09	1.48	1.47	1	.27	.09	.73	.73
14	-.27	.09	.92	.93	16	.41	.09	.49	.49
2	-.26	.09	.83	.83	9	.75	.09	.80	.80
M	logit = -.327)		SE = .09		내적합 평균=1.00		외적합 평균=1.00		
s	logit = .61		SE = .01		내적합 평균= .38		외적합 평균= .39		

〈표 4〉 예비 국어교사(평가자 국면)의 엄격성 수준과 적합도

분석 결과 평가자 국면의 엄격성 수준과 적합도 관련 결과를 위의 <표 4>와 같이 정리하였다. 카이자승 763.6, 유의수준 .00, 분리신뢰도 R=.98으로, 표집한 예비 국어교사의 데이터는 엄격성 측면에서 통계적으로 유의한 차이가 있는 것으로 확인되었다. 가장 관대한 평가자인 5번의 logit 값은 -1.74, 가장 엄격한 평가자인 9번의 logit 값은 .75이다. 외적합 지수 1.3이상인 경우 부적합⁸⁾으로 볼 수 있는데 여기에 해당하는 평가자는 7, 13, 12, 19, 18번이었다. 외적합 지수 0.7이하인 과적합⁹⁾은 6, 4, 16번이었다. 위의 두 경우에 해당하지 않으면서 엄격성이 적정 수준에 유지된 것으로 나타난 평가자는 5, 8, 10, 15, 17, 14, 2, 20, 3, 11, 1, 9번 등 총 12명이었다.

7) 채점자 항목은 logit 평균값을 0으로 보정하지 않도록 설정하였다.(명령어 Noncenter)

8) 모형을 통해 점수가 낮을 것으로 예상되는 글에 높은 점수를 주거나, 점수가 높을 것으로 예상되는 글에 낮은 점수를 주는 경향이 강하다고 해석

9) 모형을 통해 예상되는 학생들의 수준을 상대적으로 정교하게 변별하지 못한 것으로 해석

엄격성 수준은 연구에 참여한 예비 국어교사들의 엄격성이 얼마나 일관되었는지를 보여주는 값이다. 엄격성 수준에 따라 과적합, 적합, 부적합의 3집단으로 구분한 뒤 쓰기 평가 효능감을 종속변수로 하여 일원배치분산분석을 실시하였으나 문항별, 일반/실행 효능감별로도 통계적으로 유의한 차이는 확인되지 않았다. 반대로 효능감 수준을 상/하 집단으로 구분하고 외적합지수 및 엄격성의 logit 값을 종속변수로 하여 독립표본 t검정을 실시하였으나, 역시나 통계적으로 유의한 차이는 확인되지 않았다.

예비 교사들이 갖고 있는 쓰기 평가 효능감이 실제 쓰기 평가 수행이나 쓰기 평가 전문성에 토대를 두고 형성된 것이 아닐 수 있다는 점을 시사한다. 다시 말하면, 자신이 새롭게 바뀐 행동을 수행할 수 있다는 믿음과 더불어 환경이 자신의 변화된 행동에 따라 변화된 반응을 보일 것으로 믿어야 하는 조건을 충족시키기에는 쓰기 평가와 관련된 경험을 쌓을 기회가 매우 제한되었기 때문으로 볼 수 있다.

일반적으로 영역 특수적 자기 효능감은 해당 영역의 역량의 매개 변인인 경우가 많은 것으로 알려져 있다(김아영 외, 2007). 쓰기 평가를 적절히 하기 위해서는 쓰기 평가 효능감도 뒷받침되어야 한다. 하지만 이것만으로 충분하지는 않으며 쓰기 평가 전문성을 갖추기 위해 자신이 일정한 변화를 위한 노력을 기울여야 하며, 이에 따라 쓰기 평가 전문성이 생겨날 것이라는 믿음으로 이어져야 하는 것이다.¹⁰⁾ 이는 쓰기 효능감은 높으나 실제 쓰기 수행 수준이 높지 않은 학생이 존재한다는 박영민 외(2016)의 지적과도 맥이 통한다. 자신이 글을 잘 쓰기 위해 어떠한 노력을 기울여야 하는지를 알지 못하는 상태에서라면 쓰기 효능감이 높을 수는 있으나 쓰기 수행 수준이 높지

10) 본 연구에 참여한 학생들은 작문교육론을 수강하였지만 실험 데이터를 수집하기 이전에 쓰기 평가의 실제에 참여해본 경험이 거의 없었다. 작문교육론 시간에 총 2주(6시간)를 할애하여 쓰기 평가와 관련된 이론적 설명을 접한 것만으로는 쓰기 평가 전문성이 확보될 수 없음은 자명하다. 이러한 특성은 쓰기 평가 효능감이 확립되는 데에도 큰 영향을 미치지 못했을 것으로 추정할 수 있다. 이와 관련하여, IV장의 자유기술식 반응에 대한 논의를 통해서 구체적인 원인을 추정해 보고자 하였다.

않은 경우가 쉽게 확인되기 때문이다.11)

이러한 점에 비추어보면, 우선적으로는 예비 국어교사나 국어교사의 쓰기 평가 전문성 신장을 위해서는 평가자가 스스로 무엇을 해야 하는지를 명확히 안내해 줄 필요가 있다. 그리고 이러한 토대 위에서야 쓰기 평가 효능감 신장을 위한 고민이 이루어져야 한다는 시사점을 얻을 수 있는 것이다.

(4) 채점 요인에 따른 난도와 적합도

여기에서는 주관적으로 어렵다고 밝힌 정도와 실제 엄격성 및 내적합, 외적합 지수의 관련성을 살펴보았다. 우선 평가 범주에 따른 난도와 적합도는 아래와 같다.

채점 요인	엄격성 (logit)	점수 평균 (표준편차)	표준오차 (SE)	내적합지수 (Infit MS)	외적합지수 (Outfit MS)
내용	-.31	3.45 (0.5839)	.05	1.20	1.25
단어	-.13	3.30 (0.5647)	.05	.80	.80
표현	.02	3.18 (0.4241)	.05	1.00	.99
조직	.03	3.18 (0.4036)	.05	.94	.94
형식	.39	2.89 (0.5595)	.05	1.04	1.03
M	.00	16.01	.05	1.00	1.00
s	.23	2.166162	.00	.13	.15

〈표 5〉 채점 요인에 따른 난도와 적합도

그리고 예비 국어교사들이 스스로 인지한 평가 범주별 점수 선정의 어려움에 대한 응답 결과는 다음과 같다.

11) 반대로 쓰기 효능감이 낮으나 자신이 글을 잘 쓰기 위해 무엇을 해야 하는지 아는 학생들의 경우도 쓰기 수행 수준이 높지 않다는 점은 영역 특수적 효능감이 특정 영역의 역량에 매개변인으로 작용한다는 점을 뒷받침해주는 사례가 된다.

내용	조직	표현	단어	형식
3.4	2.95	3	2.95	4.3
1.019804	1.071214	1.224745	1.32193	1.345362

<표 6> 채점 요인별로 응답한 점수 결정의 어려움

채점 요인에 따른 엄격한 정도와 적합도를 확인한 결과는 위의 <표 6>과 같다. 각 요인들은 통계적으로 유의한 차이를 보여주었다(카이 제곱: 124.8 유의수준 .00, 분리신뢰도 R=.96). 평균 점수나 엄격성을 기준으로 가장 많은 점수를 준 것은 내용-단어-표현-조직-형식의 순서였다. 내적합지수나 외적합지수는 모두 0.7에서 1.3사이에 분포하여 적절한 수준을 보였다. 다만 로짓 값 1.0을 가장 적합한 수준의 기준으로 삼아본다면, 외적합지수를 기준으로 많이 벗어나 있는 수준으로 나열을 하면 내용-단어-조직-형식-표현의 순서이다. 이는 아래 표에서 언급한 예비 국어교사들의 주관적인 채점 요인 판정의 어려움 순서와는 일치하지 않는다. 이 점은 스스로 인지하는 어려움과, 다국면 Rasch 모형을 통해 확인되는 쓰기 채점의 적절함이 상응하지 않는다는 점을 시사한다. 이 점은 앞의 (3)항에서 논의한 시사점과 맥을 같이 한다.

다. 자유기술식 반응에 대한 분석 및 논의

본 연구에 참여한 예비 국어교사들은 쓰기 평가 효능감과 쓰기 평가 수행에서 모두 일정한 특성과 고유한 차이를 드러내었다. 하지만 이 두 가지 변인 사이에 통계적으로 유의미한 관련성을 찾기는 어려웠다. 이러한 연구 결과에 대한 보완 논의를 수행하고자, 추가적으로 예비 국어교사들이 쓰기 평가 과정에 대해 자유기술식으로 반응한 결과를 탐색하였다.¹²⁾

첫째, 과반수를 넘는 12명의 학생이 자신의 쓰기 평가 수행에 대한 확신

12) 수강생들은 띄어쓰기 포함 평균 762자 분량으로 다양한 응답을 제시하였다. 이 중에서 본 연구에서 관심을 갖고 있는 쓰기 효능감 및 쓰기 평가 전문성과 관련된 하여 제시한 학생들의 반응을 별도로 발췌하였고, 이를 크게 세 가지 범주로 분류하였다. 각각의 범주는 아래와 같으며, 논문 전체 분량을 고려하여 각각 대표적인 반응 세 가지만을 소개하고자 한다.

을 갖지 못하고 있다는 점을 명시적으로 드러내었다. 일관성이라는 단어를 직접 사용한 경우도 있었고, 신뢰성이나 내면화하고 있는 평가 기준이 흔들렸다는 것을 경험한 점을 분명히 드러낸 경우도 있었는데, 공통적으로 자신의 쓰기 평가 수행에 대해 만족하지 못하고 있었다.

- 2. 나름 기준을 정해 채점해보았는데도 일관성을 지키기가 힘들었다. 그래서 일관성 있는 평가를 하지 못한 것 같다.
- 8. 학생이 문제를 제기했을 경우 이해시킬 수 있을까 하는 생각이 들었다.
- 20. '내가 교사가 되었을 때, 잘 할 수 있을까'라는 불안감이 생겨 스스로에게 실망감도 없지 않아 가지게 된 시간

하지만 적절한 엄격성이 유지된 것으로 나타난 평가자 중 5명이 이러한 반응을 보였고, 엄격성이 유지되지 않은 것으로 나타난 평가자 중 3명은 이와 관련된 언급이 전혀 없었다. 이러한 점은 자신의 수행에 대한 적절함을 스스로 인지하지 못하는 경우가 상당수라는 점을 보여주는 결과라 할 수 있다.

둘째, 10명의 수강생이 30편의 학생 필자 글을 채점하는 것에 대한 피로감을 호소하였다.

- 9. 완성도가 낮은 글을 읽는다는 것 자체가 피로한 것 같다. 그리고 많은 양의 글을 읽다보니 되게 기계적으로 평가를 하는 것 같아서 하나의 무의미한 작업을 하는 것 같아 기분이 좋지 않았다.
- 14. 실제 한 학급이 쓴 만큼의 양을 읽는 게 이렇게 힘들 줄 짐작은 했으나 실제로 해 보니 정말 어려운 것 같다.
- 16. 단순 점수만으로 30명의 작문을 평가하는 일도 이렇게 부담 됐는데, 모든 글을 자세히 피드백 해준다는 건 사실 불가능에 가까울 것 같다.

이론적 배경의 2장에서 논의한 바와 같이, 일정한 쓰기 평가 효능감을 유지하고, 또한 적절한 쓰기 평가 수행을 위해서는 평가 대상 글에 성실하게 임하면서 동시에 스스로의 평가 태도에 대해 점검하 필요도 있다. 하지만 절

반의 수강생들은 30편의 학생 글을 평가하는 것에 대해 상당한 정도의 피로감을 호소하였다. 이는 단순히 의지의 문제라기보다는 수강생들이 처음으로 학생글을 채점해본 것과 관련지어 해석할 수 있다. 특히 실제로 해 보았더니 생각한 것과 다르다는 반응이 다수 등장했다는 점에 비추어보면, 실제로 쓰기 평가 과정을 체험하는 경험이 예비 국어교사 양성과정에서 보다 늘어날 필요가 있다는 점을 시사한다.

셋째, 8명의 수강생들이 학생 필자의 발달 수준에 대한 이해 부족에서 비롯된 어려움을 호소하였다.

1. 아이들이 어느 정도의 어휘 수준을 가지고 있는지 몰라 판단하기 힘들었다. 나의 어휘수준과는 다르기에 어떤 기준으로 점수를 주어야하는지가 곤란하였다. 아이들 수준에서는 독창적이며 쉽고 정확하게 표현하였을 수도 있어 평가가 어려웠다.
5. 학생들의 발달상태를 알지 못했기 때문에 어떤 기준으로 점수를 매겨야 하는지가 모호했다. 특히 단어나 표현부분의 점수는 발달상태에 따라 변동성이 크다고 생각했기 때문에 학생들의 발달상태를 모른 채, 글을 평가하는 점이 곤란했다.
12. 학생의 글을 보며 이정도면 괜찮은 글인지 아닌지 판단하는 것이 어려웠다.

이러한 반응은 쓰기 평가 전문성의 지식 관련 요소로 설명할 수 있다. 쓰기 평가에는 쓰기 교육 분야에서 다루는 전반적인 정보들이 폭넓게 요구된다. 그런데 유독 지식 요인 중 수강생들의 반응은 학생 필자의 발달 수준과 관련된 것으로 모아졌다. 작문교육론 강의를 통해 다양한 작문 발달 이론에 대해 다루었지만, 이것만으로는 충분하지 않았다고 해석할 수 있다. 특히 실제 학생 필자들이 작성한 다양한 자료글을 토대로 학생들의 적절한 발달 수준을 확인하는 교육 내용이 마련될 필요가 있다는 점을 의미한다. 또한 쓰기 평가 지식의 주요 구성 요인 중 하나가 특정 발달 수준의 학생 필자가 작성할 수 있는 글의 수준은 어느 정도인지에 대한 지식이라는 점도 확인할 수 있다.

위와 같은 세 가지 유형의 반응을 보았을 때, 작문교육론 교과목을 통해 약 2주간에 걸친 쓰기 평가와 관련된 수업 경험은 쓰기 평가 전문성을 함양하거나 근거가 분명한 쓰기 평가 효능감을 형성토록 해 주는 데에는 충분하지 못했다는 점을 추정할 수 있다. 특히나 이론적 배경의 1절에서 논의한 바와 같이, 자기 효능감은 자신이 적절한 수행을 할 수 있다는 믿음뿐만 아니라, 자신이 수행한 결과가 적절한 방식으로 받아들여질 것이라는 점에 대한 믿음도 필수적으로 요구한다. 이러한 차원에서 살펴보자면, 본 연구에 참여한 예비 국어교사들이 일관성이나 엄격성 차원에서 적절한 쓰기 평가 수행을 보였거나, 혹은 그 반대일지라도 자신이 수행한 것이 어느 정도로 적절한 것인지를 가늠할 방법이 명확하지 않다는 점을 확인할 수 있었다.

V. 결론

본 연구에서는 집단 비교를 통해 쓰기 평가 효능감과 엄격성, 평가 범주별로 주관적으로 느끼는 어려움 및 실제 채점 수행의 적합도 등을 살펴보았다. 본 연구는 위의 연구들과 맥을 같이 하되, 예비 국어교사를 대상으로서는 최초의 시도이다. 연구 결과 쓰기 평가 효능감과 쓰기 평가에서 나타나는 엄격성(적합도)에서는 관련성이 확인되지 않았다. 또한 평가 범주별로 확인된 적합도와 예비 국어교사들이 느끼는 평가 범주별 모호함의 정도 역시 밀접한 관련이 확인되지 않았다.

참고로 쓰기 평가 효능감과 다국면 Rasch 모형을 모두 활용하여 국어교사를 대상으로 이루어진 연구인 장은주(2015)에서는 쓰기 평가 효능감 수준이 높은 집단이 관대한 평가를 하는 경향이 나타났다. 반면에 백유진(2018)에서는 쓰기 평가 효능감 수준과 평가의 엄격성 사이에서는 일정한 연관성을 찾기 어려웠다. 국어교사를 대상으로 한 두 편의 연구에서 이러한 차이가 나타난 원인은 선행 연구 수준에서는 명확히 밝혀지지 않았다. 하지만 예비 국어교사를 대상으로 이루어진 본 연구 결과를 토대로 본다면, 쓰기 평가 경

험이 충분하지 않은 상황에서 측정된 쓰기 평가 효능감의 실체가 분명하지 않을 수 있다는 점을 추정하게 한다. 자기 효능감은 충분한 수행 경험을 토대로 해야만 형성되는 특성을 갖고 있기 때문이다. 물론 앞으로도 양적 접근 외에 다양한 연구 방법을 동원하여 쓰기 평가 효능감이 쓰기 평가 수행에 어떠한 영향을 미치는지를 탐색할 필요가 있다.

더불어, 본 연구에 참여한 예비 국어교사들이 쓰기 평가와 관련된 교육이나 훈련을 특별히 많이 받았다고 보기는 어렵다. 하지만 현재 사범대 교육과정에서 쓰기 평가 전문성 신장을 위한 집중적인 교육이 보편적으로 이루어지고 있는지는 않다는 점을 고려하면 이들이 여느 예비 국어교사들과 비교해 보더라도 쓰기 평가 전문성 교육 차원에서 이질 집단이라고 볼 수는 없다. 본 연구의 결과만 가지고 일반화하기에는 조심스러운 부분이 있지만, 현재 이루어지고 있는 수준의 쓰기 평가 교육으로는 임용 후 곧바로 쓰기 평가 실전에 투입할 수 있을 정도의 쓰기 평가 전문성을 확보하는 데에는 무리가 따른다는 점, 향후 예비 국어교사들이 쓰기 평가 전문성 신장 과정에서 겪고 있는 구체적인 어려움을 확인한 것이 본 연구의 가장 큰 시사점이라 할 수 있다.

참 고 문 헌

- 가은아(2018), 「예비 국어교사의 쓰기 평가 지식과 쓰기 평가 효능감 분석」, 『리터러시연구』9권 4호, 125-160, 한국리터러시학회.
- 권태현(2012), 「쓰기 채점 방식에 따른 국어교사의 채점 신뢰도 비교 연구」, 『교원교육』28권 4호, 113-134, 한국교원대 교육연구원.
- 김아영 외(2007), 『학업적 자기효능감 :이론과 현장연구』, 학지사.
- 박영목(2007), 「쓰기 평가 연구의 주요 과제」, 『작문연구』제6집, 9-37, 한국작문학회.
- 박영민 외(2016), 『작문교육론』, 역락.
- 박영민(2010), 「예비국어교사의 쓰기 평가 효능감 분석」, 『청람어문교육』 제42집, 181-207, 청람어문교육학회.
- 박영민(2012), 「예비국어교사의 중학생 논설문 평가에서 발견되는 엄격성 및 일관성의 특성」, 『국어교육학연구』 제43집, 253-283, 국어교육학회.
- 박영민.최숙기(2010), 「Rasch 모형을 활용한 국어교사의 쓰기 평가 특성 분석」, 『국어교육학연구』 제37집, 367-391, 국어교육학회.
- 박종임(2013), 『국어교사의 쓰기 평가 특성 연구』, 한국교원대 대학원 박사학위논문.
- 박찬홍(2018), 「대학생의 유형별 글쓰기에서 나타나는 채점요인 간의 관계 분석」, 『작문연구』 제38집, 65-97, 한국작문학회.
- 이성영(2005), 「국어과 교사의 쓰기 영역 평가 전문성 기준과 모형」, 『국어교육』 제117집, 353-376, 한국국어교육학회.
- 이인제 외(2004), 『교사의 학생 평가 전문성 신장 모형과 기준』, 한국교육과정평가원.
- 임규혁 외(2007), 『교육심리학』, 학지사.
- 장은주(2015), 『채점의 일관성 유형에 따른 국어교사의 쓰기 평가 특성 분석』, 한국교원대학교 박사학위논문.
- 지은림 외(2000), 『RASCH모형의 이론과 실제』, 교육과학사.

Bandura, A. (1995). 『변화하는 사회 속에서의 자기효능감 (윤운성 외 역.)』.
학지사.

Spandel, V., & Culham, R. (1996). Assessing portfolios. *A handbook for student performance assessment in an era of restructuring*, 1-7.

【Abstracts】

A Study on Writing Evaluation Efficacy and Writing Evaluation Performance of Preliminary Korean Teachers

Park, Chan - heung

In this study, the efficacy of writing evaluation was noted to find empirical connotations for the professional development of writing evaluation of prospective Korean language teachers. Specifically, the purpose of the study was to collect variables such as the effectiveness of writing evaluation by prospective Korean language teachers, the difficulties they experienced in the writing evaluation process and opinions on their experiences in conducting writing evaluations, and consistency and rigor in the results of writing evaluations. The group comparison looked at the efficacy and rigour of the writing evaluation, the difficulty of feeling subjective by the assessment category, and the suitability of performing the actual grading, and found no relevance in the writing evaluation efficacy and the rigour shown in the writing evaluation. In addition, the degree of suitability identified for each assessment category and the degree of ambiguity felt by prospective Korean language teachers were also not closely related. In order to interpret these findings more closely, the difficulties experienced by prospective Korean language teachers in conducting write evaluations were brought to a free technical response, and the factors related to the subject of the study were looked at during these responses. Exploring the

free technical responses, it was found that the current writing evaluation experience provided by the College of Education did not suffice to form a writing evaluation efficacy or writing evaluation expertise.

Key words : writing evaluation efficacy, writing evaluation expertise, consistency, preliminary Korean language teacher, multi-facets Rasch model

이 논문은 2019년 10월 31일에 투고되었으며, 2019년 11월 18일에 심사 완료되어 2019년 11월 20일에 게재가 확정되었음.