



저작자표시-비영리-동일조건변경허락 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.
- 이차적 저작물을 작성할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



동일조건변경허락. 귀하가 이 저작물을 개작, 변형 또는 가공했을 경우에는, 이 저작물과 동일한 이용허락조건하에서만 배포할 수 있습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

문학박사학위논문

중·고등학교 영어교사들의 말하기
수행평가의 시행에 관한 인식 연구



2015년 2월

부경대학교 대학원

영어영문학과

고현우

문학박사학위논문

중·고등학교 영어교사들의 말하기
수행평가의 시행에 관한 인식 연구



2015년 2월

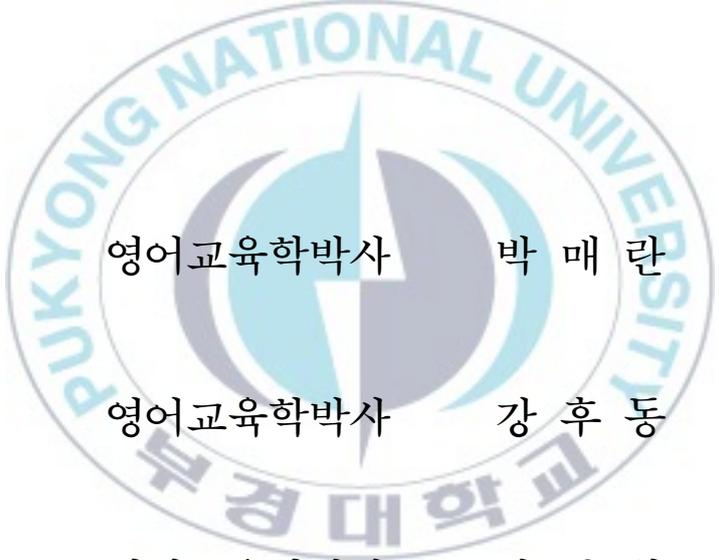
부경대학교 대학원

영어영문학과

고 현 우

고현우의 문학박사 학위논문을 인준함

2015년 2월 27일



주 심	영어교육학박사	박 매 란	(인)
위 원	영어교육학박사	강 후 동	(인)
위 원	영어교육학박사	박 종 원	(인)
위 원	교육학박사	조 윤 경	(인)
위 원	응용언어학박사	오 준 일	(인)

목 차

영문초록	viii
I. 서론	1
1.1 연구의 필요성	1
1.2 연구의 목적	5
1.3 연구 과제	6
1.4 연구의 구성	7
1.5 연구의 제한점	8
II. 이론적 배경	9
2.1 영어과 수행평가	9
2.2 말하기 수행평가의 효용성	12
2.2.1 성취도 측면	13
2.2.2 정의적 측면	14
2.2.3 수업 개선의 측면	16
2.3 말하기 수행평가의 전문성	18
2.4 말하기 수행평가의 시행한계성	23
2.5 교사의 평가 시행에 영향을 미치는 요인	28
III. 예비 연구	33
3.1 예비 연구의 과제	33
3.2 예비 연구의 방법	34
3.3 예비 연구의 분석 결과 및 논의	38

3.4 예비 연구의 보완점	41
IV. 연구 방법	43
4.1 연구 대상	43
4.2 연구 도구	46
4.2.1 설문지	46
4.2.2 신뢰도 및 타당도	48
4.3 자료의 수집 절차 및 분석 방법	52
V. 연구 결과 및 논의	55
5.1 학교급간 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 인식	55
5.1.1 학교급간 말하기 수행평가의 전문성에 대한 인식 차이	56
5.1.2 학교급간 말하기 수행평가의 효용성에 대한 인식 차이	70
5.1.3 학교급간 말하기 수행평가의 시행한계성에 대한 인식 차이	76
5.1.4 학교급간 말하기 수행평가의 시행한계성에 대한 개선 방안	91
5.1.5 요약	98
5.2 배경 변인에 따른 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 인식	101
5.2.1 배경 변인에 따른 말하기 수행평가의 전문성에 대한 인식 차이	101
5.2.2 배경 변인에 따른 말하기 수행평가의 효용성에 대한 인식 차이	112
5.2.3 배경 변인에 따른 말하기 수행평가의 시행한계성에 대한 인식 차이	115
5.2.4 요약	116
5.3 말하기 수행평가의 전문성, 효용성 및 시행한계성 간의 상관관계	119
5.3.1 전체 영어교사들을 대상으로 한 세 주요 요인들 간의 상관관계 분석	120
5.3.2 중학교 영어교사들을 대상으로 한 세 주요 요인들 간의 상관관계 분석	123
5.3.3 고등학교 영어교사들을 대상으로 한 세 주요 요인들 간의 상관관계 분석	124

5.3.4 요약	126
5.4 말하기 수행평가의 시행에 대한 구조방정식모형	127
5.4.1 이론적 모형	128
5.4.2 구조방정식모형의 정규성 검정	130
5.4.3 구조방정식모형의 적합도 검증	132
5.4.4 구조방정식모형의 해석	133
VI. 결론 및 제언	140
6.1 연구의 결론	140
6.2 제언	142
참고문헌	146
부록	158

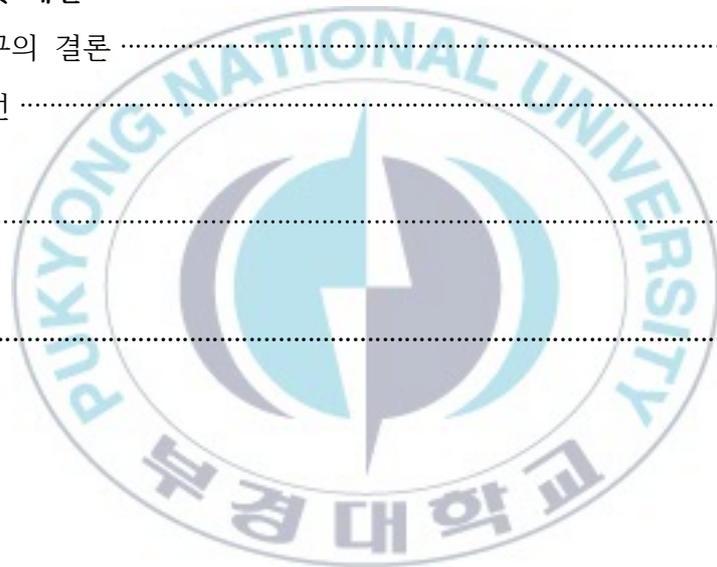


표 목차

<표 1> 설문지의 세부 문항 구성(예비 연구)	35
<표 2> 주요 요인들에 대한 요인 분석 결과(예비 연구)	36
<표 3> 연구 도구의 KMO와 Bartlett의 구형성 검정 결과(예비 연구)	37
<표 4> 연구 도구의 신뢰도(예비 연구)	37
<표 5> 세 주요 요인들 간의 상관관계(예비 연구)	38
<표 6> 잠재 변인들 간의 경로분석 결과(예비 연구)	39
<표 7> 지역별 설문 참여 학교 및 참여자 수	43
<표 8> 자유서술식 문항 응답자 수	44
<표 9> 설문 참여자들의 배경 변인별 분포	45
<표 10> 주요 요인별 세부 문항 구성	48
<표 11> 연구 도구의 신뢰도	49
<표 12> KMO와 Bartlett의 구형성 검정 결과	50
<표 13> 주요 요인들에 대한 요인 분석 결과	51
<표 14> 말하기 수행평가의 전문성 전체에 대한 기술통계	56
<표 15> 말하기 수행평가의 전문성 전체에 대한 중·고등학교 집단 간의 인식 차이 검증	56
<표 16> 말하기 수행평가의 전문성 세 하위 요인들에 대한 기술통계	57
<표 17> 말하기 수행평가의 전문성 세 하위 요인들에 대한 중·고등학교 집단 간의 인식 차이 검증	58
<표 18> 평가 방법과 도구의 선정 및 개발 각 문항에 대한 기술통계	59
<표 19> 평가 방법과 도구의 선정 및 개발 각 문항에 대한 중·고등학교 집단 간의 인식 차이 검증	60
<표 20> 평가의 실시 및 채점 각 문항에 대한 기술통계	63
<표 21> 평가의 실시 및 채점 각 문항에 대한 중·고등학교 집단 간의 인식 차이 검증	64

<표 22> 평가 결과의 해석 및 활용 각 문항에 대한 기술통계	67
<표 23> 평가 결과의 해석 및 활용 각 문항에 대한 중·고등학교 집단 간의 인식 차이 검증	67
<표 24> 말하기 수행평가의 효용성 전체에 대한 기술통계	71
<표 25> 말하기 수행평가의 효용성 전체에 대한 중·고등학교 집단 간의 인식 차이 검증	71
<표 26> 말하기 수행평가의 효용성 각 문항에 대한 기술통계	72
<표 27> 말하기 수행평가의 효용성 각 문항에 대한 중·고등학교 집단 간의 인식 차이 검증	73
<표 28> 말하기 수행평가의 시행한계성 전체에 대한 기술통계	76
<표 29> 말하기 수행평가의 시행한계성 전체에 대한 중·고등학교 집단 간의 인식 차이 검증	76
<표 30> 말하기 수행평가의 시행한계성 각 문항에 대한 기술통계	77
<표 31> 말하기 수행평가의 시행한계성 각 문항에 대한 중·고등학교 집단 간의 인식 차이 검증	78
<표 32> 자유서술식 응답에 나타난 말하기 수행평가의 시행상의 한계점들(중학교) ..	82
<표 33> 자유서술식 응답에 나타난 말하기 수행평가의 시행상의 한계점들(고등학교) ..	86
<표 34> 중·고등학교 말하기 수행평가의 시행상의 한계점들 순위 비교	90
<표 35> 말하기 수행평가의 시행한계성에 대한 개선 방안들(중학교)	91
<표 36> 말하기 수행평가의 시행한계성에 대한 개선 방안들(고등학교)	94
<표 37> 중·고등학교 말하기 수행평가의 시행한계성에 대한 개선 방안들 순위 비교 ..	97
<표 38> 말하기 수행평가의 전문성 전체에 대한 배경 변인별 기술통계	102
<표 39> 말하기 수행평가의 전문성 전체에 대한 배경 변인별 인식 차이 검증 ..	103
<표 40> 평가 방법과 도구의 선정 및 개발에 대한 배경 변인별 기술통계	105
<표 41> 평가 방법과 도구의 선정 및 개발에 대한 배경 변인별 인식 차이 검증 ..	106
<표 42> 평가의 실시 및 채점에 대한 배경 변인별 기술통계	107
<표 43> 평가의 실시 및 채점에 대한 배경 변인별 인식 차이 검증	108
<표 44> 평가 결과의 해석 및 활용에 대한 배경 변인별 기술통계	110

<표 45> 평가 결과의 해석 및 활용에 대한 배경 변인별 인식 차이 검증	110
<표 46> 말하기 수행평가의 효용성에 대한 배경 변인별 기술통계	112
<표 47> 말하기 수행평가의 효용성에 대한 배경 변인별 인식 차이 검증	113
<표 48> 말하기 수행평가의 시행한계성에 대한 배경 변인별 기술통계	115
<표 49> 말하기 수행평가의 시행한계성에 대한 배경 변인별 인식 차이 검증	116
<표 50> 세 주요 요인들에 대한 기술통계(전체 영어교사들)	120
<표 51> 세 주요 요인들 간의 상관관계(전체 영어교사들)	120
<표 52> 세 주요 요인들에 대한 기술통계(중학교 영어교사들)	123
<표 53> 세 주요 요인들 간의 상관관계(중학교 영어교사들)	123
<표 54> 세 주요 요인들에 대한 기술통계(고등학교 영어교사들)	124
<표 55> 세 주요 요인들 간의 상관관계(고등학교 영어교사들)	125
<표 56> 각 변인별 평균, 표준편차, 왜도 및 첨도(N=243)	131
<표 57> 구조방정식모형의 적합도	132
<표 58> 잠재 변인들 간의 경로분석 결과	136
<표 59> 잠재 변인들 간의 직·간접 효과 및 전체 효과	137

그림 목차

<그림 1> 말하기 수행평가 시행의 결정 요인들	31
<그림 2> 구조방정식모형(예비 연구)	40
<그림 3> 평가 방법과 도구의 선정 및 개발 각 문항별 신뢰구간(중학교)	61
<그림 4> 평가 방법과 도구의 선정 및 개발 각 문항별 신뢰구간(고등학교)	62
<그림 5> 평가의 실시 및 채점 각 문항별 신뢰구간(중학교)	65
<그림 6> 평가의 실시 및 채점 각 문항별 신뢰구간(고등학교)	66
<그림 7> 평가 결과의 해석 및 활용 각 문항별 신뢰구간(중학교)	68
<그림 8> 평가 결과의 해석 및 활용 각 문항별 신뢰구간(고등학교)	69
<그림 9> 말하기 수행평가의 효용성 각 문항별 신뢰구간(중학교)	74
<그림 10> 말하기 수행평가의 효용성 각 문항별 신뢰구간(고등학교)	75
<그림 11> 말하기 수행평가의 시행한계성 각 문항별 신뢰구간(중학교)	79
<그림 12> 말하기 수행평가의 시행한계성 각 문항별 신뢰구간(고등학교)	81
<그림 13> 잠재 변인들 간의 이론적 모형	129
<그림 14> 구조방정식모형(비표준화)	134
<그림 15> 구조방정식모형(표준화)	135

A Study on Perceptions of Speaking Performance Assessment Practice of Secondary School English Teachers in Korea

Hyun Woo Koh

*Department of English Language and Literature, The Graduate School,
Pukyong National University*

Abstract

This study examined Korean secondary school English teachers' perceptions of expertise in speaking performance assessment, their perceptions of how useful speaking performance assessment is, and their perceptions of what makes it difficult to implement speaking performance assessment in the classroom. It also identified the effects of such perceptions on their speaking performance assessment practice. The present study surveyed 243 secondary English teachers (128 in middle and 115 in high school) from 5 different districts. The data analysis about an open-ended question for qualitative research was included as well. To explore the four research questions, the results were compared and analyzed by T-test, ANOVA, Correlation Analysis, and SEM (Structural Equation Model).

The findings of the study were as follows. First, the perceptual differences in expertise, usefulness and infeasibility of speaking performance assessment between middle and high school English teachers were compared. High school English teachers had lower expertise in speaking performance assessment, while both of them generally had positive perception in its pedagogical effectiveness. High school English teachers also felt much more trouble in conducting speaking performance assessment than their counterparts did, owing to several

reasons such as the difficulty of guaranteeing objectivity in scoring, no proper methods or tools for assessment, the shortage of helpful materials or teacher training programs, or the possibility of giving students another learning burden.

Secondly, English teachers' perceptual differences in expertise, usefulness and infeasibility of speaking performance assessment were investigated, according to teaching career, teacher training related to speaking performance assessment, and perceived competence in English speaking. English teachers with 11-15 years of teaching career were more professional in implementation and scoring of speaking performance assessment, rather than those with 16-20 years of teaching career. To enhance English teachers' professional competence and raise their positive awareness of the effectiveness of speaking performance assessment, at least more than once or twice completing the teacher training related to speaking performance assessment proved helpful. Besides, English teachers who thought their perceived competence in English speaking was superior showed the excellent competence and affirmative awareness of speaking performance assessment.

Thirdly, English teachers' perceptions in expertise, usefulness and infeasibility of speaking performance assessment were correlated to one another. It indicated that English teachers with overall more positive awareness of the educational effectiveness of speaking performance assessment had much more expertise in speaking performance assessment, or vice versa. Meanwhile, English teachers with higher awareness of the infeasibility of speaking performance assessment held lower expertise in speaking performance assessment, or vice versa. The higher infeasibility English teachers perceived in speaking performance assessment, in addition, the lower they considered the usefulness of speaking performance assessment in English education, or vice versa.

Last but not least, this study revealed the causal effect of English teachers' perceptions in expertise, usefulness and infeasibility of speaking performance

assessment on their speaking performance assessment practice. Obviously, English teachers highly equipped with expertise in speaking performance assessment were strongly willing to carry out speaking performance assessment in the classroom. Although English teachers positively perceived its educational effectiveness, however, English teachers tended to implement speaking performance assessment, only if they were much ready with professional competence in speaking performance assessment. Overcoming the infeasibility in speaking performance assessment, English teachers tended to be minimally trying to execute speaking performance assessment, just in case English teachers not only highly recognized the educational merits of speaking performance assessment, but they also had splendid competence in speaking performance assessment.

The implications of the research findings were discussed. And pedagogical suggestions to activate speaking performance assessment practice in Korean educational settings were provided.

I. 서론

본 연구를 통해서 말하기 수행평가를 시행하는 우리나라 중·고등학교 영어 교사들의 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 인식 차이를 파악하고, 또한 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 영어교사들의 인식이 말하기 수행평가의 시행에 미치는 영향을 고찰하고자 한다. 이 장에서는 연구의 필요성, 연구의 목적, 연구 과제, 연구의 구성, 그리고 연구의 제한점이 제시된다.

1.1 연구의 필요성

영어교육에서 말하기는 쓰기와 함께 화자의 의사소통 능력을 측정하는 중요한 표현 기능 중 하나이다. 1980년대에 의사소통 중심 교수법(Communicative Language Teaching)이 처음 등장한 이후 지금까지 많은 영어교육 학자들은 기존의 영어교수법에서 벗어나 실제 생활에서 영어로 대화를 이끌어갈 수 있는 능력 함양의 중요성을 늘 강조해왔다. 그리고 1990년대부터 우리나라에서도 발달적 교육관에 입각하여 학생의 지식이나 능력을 여러 측면에서 지속적으로 평가(assessment)하는 교수·학습 활동 중심의 수행평가를 본격적으로 시행하게 되었다(교육부, 1998). 이러한 영어교육의 최근 경향인 의사소통 능력의 함양이라는 목표에 맞게 2009 개정 교육과정에서는 영어의 네 기능을 연계하고 통합된 언어사용 능력을 강조하면서, 특히 말하기와 같은 표현 기능은 수행평가로 실시할 것을 제안하고 있다(교육과학기술부, 2011). 또한, 영어교육 학자들(Huerta-Macias, 1995; O'Malley & Pierce, 1996)도 학습자의

언어능력을 정확하게 측정하는 중요한 평가 수단으로서 수행평가의 역할을 강조하였다. 이제는 단지 영어에 대한 지식을 보유하는 데 그칠 것이 아니라 사회문화적 상황에서 그 지식을 적절하게 사용하는 의사소통 능력의 향상과, 그리고 학습자의 의사소통 능력을 제대로 측정하고 보다 발전적인 방향을 제시하기 위해서 타당하고 신뢰성 있는 말하기 수행평가의 시행이 필요하다.

영어 말하기 및 쓰기 수업과 학습에 대한 학생들의 만족도를 조사한 최근의 연구에서 학생들은 영어 의사소통 능력의 향상을 위해서는 쓰기 보다는 오히려 말하기가 꼭 필요하다는 인식을 가지고 있었다(한국교육과정평가원, 2012). 그리고 실제로 우리나라 중·고등학교 학생들이 영어 수업시간에 직접적인 영어회화를 할 수 있는 기회가 적은 이유는 독해와 문법에만 매달리는 영어교육 방식 때문인데, 이제는 학교에서 의사소통 중심의 영어교육을 위한 근본적인 변화가 시급하다는 현장 보고(김광현, 2014.09.07)도 있다. 또한, 김진철(1998)도 영어 말하기 능력은 한 사람의 영어 실력을 판단하는 준거가 되고 외국어 학습은 바로 말하기 학습인 것으로 인식될 정도로 말하기가 외국어 학습에서 매우 중요하게 여겨진다고 주장하였다. 따라서 이러한 영어 말하기에 대한 시대적 요구를 반영하여 앞으로의 실용적인 영어교육이 나아가야 할 방향은 영어의 형태와 구조보다는 의미(meaning)와 기능(function)에 중점을 둔 말하기 교육이 되어야 할 것이다. 이를 실천하기 위해서는 영어교사들이 학생들의 말하기 수행을 직접 평가하고 그 결과를 활용하여 말하기 능력의 발달에 유용한 도움을 줄 수 있는 말하기 수행평가의 시행이 필요하다.

말하기 수행평가를 시행하는 일은 평가의 환류효과(washback effect)와 결부지어 보면 학교 현장의 교수·학습 과정에서 아주 중요하다. 평가가 교수와 학습 과정에 미치는 영향에 대한 실험적 연구에서 Cheng(1997), Jung(2008), Wall과 Alderson(1993), 그리고 Watanabe(1996)는 새로운 평

가의 도입으로 인해서 교사들의 수업 과정과 내용이 달라짐을 보여주었다. 그리고 말하기 수행평가는 교사의 수업 방법 개선을 위한 유용한 피드백(feedback)을 제공함으로써 궁극적으로 학교 교육의 정상화까지도 도모할 수 있다(O'Brien, 1997). 게다가 말하기 수행평가는 학생들의 학습 과정을 정확하게 파악하고 학생들의 의사소통 능력의 향상에 기여하며(이완기, 2012), 또한 말하기는 학습자의 인지적 구조나 문제해결 과정과 밀접한 연관성이 있기 때문에(이기수, 1999), 영어 말하기 수행평가의 시행을 통해서 학생 개인의 특성과 학습활동의 전 과정을 종합적으로 평가할 수 있는 장점이 있다. 그러므로 말하기 수행평가의 시행이 학교 현장에 미치는 긍정적인 영향을 고려해 볼 때, 현재 읽기 위주의 영어수업이 대세인 우리나라에서 말하기 수행평가의 적극적인 시행은 영어교사의 수업 방식의 개선뿐만 아니라 학생들의 영어 말하기 능력의 향상을 위한 획기적인 단초가 될 수 있다.

한편, 이러한 영어 말하기 수행평가의 교육적 효용성(effectiveness)과 중요성에도 불구하고 실제 학교 현장에서는 말하기 수행평가를 아예 시행하지 않거나, 혹은 시행을 하더라도 수많은 현실적인 한계에 부딪혀 제대로 시행되지 못하고 있는 실정이다(고현우, 2014; 김미은, 2002; 이소영, 2008; 이은주, 2003; 최윤경, 2001). 그리고 최근 각 시·도 교육청이 교육 예산의 부족으로 인해서 영어 원어민 교사 배치에 대한 지원을 축소하기 시작한 이후로(이태윤, 2012.03.20) 대부분의 중·고등학교에서는 영어 원어민 교사들이 주로 담당하던 영어회화 수업을 한국인 영어교사들이 대신해야 하는 부담이 생겼다. 더구나 영어 교과는 그 특성 상 타 교과에 비해서 수행평가의 영역에 말하기를 포함한 영어의 네 기능 모두를 반영해야 하는 문제도 영어교사들의 또 다른 심적 부담으로 작용하고 있다. 이처럼 일선 학교에서는 말하기 수행평가의 시행에 걸림돌이 되는 현실적인 한계성(infeasibility)이 높아서 사실 많은 영어교사들이 신뢰성과 타당성이 낮은 작의적인 방식으로 말하기

수행평가를 시행하고 있는 실정이다(김성연, 이호경, 2013; 방인자, 천세영, 2011; 장인순, 2004).

결국 말하기 수행평가가 지닌 교육적 효용성을 극대화하고 말하기 수행평가의 시행상의 한계점들을 극복하기 위해서는 말하기 수행평가에 대한 영어교사의 전문적인 능력과 자질이 선행되어야 한다. 즉, 학생들에게 말하기 능력의 향상이라는 교육적 효과를 안겨주기 위해서 영어교사의 말하기 수행평가에 대한 평가 전문성(assessment expertise)은 우선적인 필요조건이라고 할 수 있다. 전병만 외(2005)는 평가는 이해 당사자에게는 엄청나게 중요한 변수이므로 현장 교사들의 평가 전문성에 대한 사회적인 기대와 요구가 날이 높아지고 있음을 주장하였다. 그러므로 실생활에서의 의사소통 능력의 향상이라는 교육과정의 목표에 맞게 학생들의 말하기 능력을 평가하려면 무엇보다도 영어교사들은 말하기 수행평가의 계획부터 준비와 시행 및 결과 활용에 걸친 모든 단계에서 탁월한 평가 전문성을 지녀야 함은 당연한 일이다. 왜냐하면, 자신의 평가 전문성이 높다고 인식하는 교사는 평가의 시행을 성공적으로 수행하게 될 가능성이 높다(Bol, 2004; Jonson, 1999)는 점에서, 말하기 수행평가에 대한 전문성이 높은 영어교사는 말하기 수행평가를 더 잘 시행하게 될 것이기 때문이다.

이와 같이 우리나라의 학교 현장에서 말하기 수행평가를 시행하는 일은 말하기 수행평가가 지닌 긍정적인 환류효과, 말하기 수행평가의 시행을 어렵게 만드는 한계점들, 그리고 말하기 수행평가를 성공적으로 시행하기 위한 영어교사들의 전문성과 밀접한 관련이 있는 것으로 보인다. 교사의 평가 시행에 영향을 미치는 요인들 간의 상관관계를 연구한 Yang(2008)에 의하면, 교사 자신의 평가 능력에 대한 인식과 평가의 교육적 장점에 대한 인식은 교사의 평가 시행과 긍정적인 상관관계가 있는 반면, 교사의 평가 시행의 어려움에 대한 인식은 교사의 평가 시행과 부정적인 상관관계가 있음을 밝히고 있다.

즉, 영어교사들에게 있어서 말하기 수행평가의 전문성에 대한 인식은 말하기 수행평가의 효용성뿐만 아니라 말하기 수행평가의 시행한계성에 대한 인식과 서로 연관성이 있고, 또한 말하기 수행평가의 시행에도 어떠한 영향을 미칠 것으로 짐작된다. Bandura(1977)의 자기 효능감 이론(self-efficacy theory)에 따르면, 평가에 있어서 교사들의 자기 효능감은 자신의 평가 능력에 대한 자신감으로서 평가를 시행할 수 있는 능력에 영향을 미친다고 하였다. 또한, Borg(2003)의 연구에서도 평가에 대한 전문성 교육은 평가 시행의 장점에 대한 교사의 신념에 영향을 미친다는 결과를 발표하였다.

따라서 본 연구를 통해서 말하기 수행평가에 대한 전문성, 효용성 및 시행한계성에 대한 영어교사들의 인식은 학교급간과 배경 변인에 따라 어떻게 차이가 나고, 또한 서로 어떠한 연관성을 지니고 있는지를 고찰하고자 한다. 그리고 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 영어교사들의 인식이 말하기 수행평가를 시행하는 데 직·간접적인 영향을 얼마나 미치는지를 파악하는 일 또한 매우 큰 의미가 있으리라 본다.

1.2 연구의 목적

본 연구의 목적은 말하기 수행평가를 시행하는 우리나라 중·고등학교 영어교사들의 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 인식 차이를 살펴보고, 이들에 대한 영어교사들의 인식이 말하기 수행평가의 시행에 미치는 영향을 고찰하는 데 있다. 이를 위하여 연구자는 구체적인 연구 목표를 설정하였다.

첫째, 말하기 수행평가를 시행하는 중학교 영어교사들과 고등학교 영어교사들의 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 인식이 서로

어떻게 차이가 나는지를 살펴본다. 둘째, 말하기 수행평가를 시행하는 중·고등학교 영어교사들의 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 인식은 배경 변인(교직 경력, 관련 연수 이수 및 말하기 능력)에 따라 서로 어떠한 차이가 있는지를 살펴본다. 셋째, 말하기 수행평가를 시행하는 중·고등학교 영어교사들의 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 인식은 서로 어떠한 상관관계에 있는지를 살펴본다. 넷째, 말하기 수행평가를 시행하는 중·고등학교 영어교사들이 지닌 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 인식은 학교 현장에서 영어교사들이 실제로 말하기 수행평가를 시행하는 정도에 어떠한 영향을 미치고 있는지를 살펴본다.

1.3 연구 과제

말하기 수행평가를 시행하는 중·고등학교 영어교사들이 인식하는 말하기 수행평가의 전문성, 효용성, 시행한계성 및 말하기 수행평가의 시행도에 대한 인식 차이를 살펴보기 위해서 연구자는 앞 절에서 기술된 연구 목적에 따라 다음과 같이 네 가지의 연구 과제를 설정하였다.

첫째, 말하기 수행평가를 시행하는 중학교 영어교사들과 고등학교 영어교사들 간에 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 인식은 서로 어떠한 차이가 있는가?

둘째, 말하기 수행평가를 시행하는 중·고등학교 영어교사들의 배경 변인(교직 경력, 관련 연수 이수 및 말하기 능력)에 따라 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 인식은 서로 어떠한 차이가 있는가?

셋째, 말하기 수행평가를 시행하는 중·고등학교 영어교사들의 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 인식은 서로 어떠한 상관관계가 있는가?

넷째, 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 중·고등학교 영어교사들의 인식은 영어교사들이 실제로 말하기 수행평가를 시행하는 정도에 어떠한 영향을 미치는가?

1.4 연구의 구성

제 1장인 서론은 본 연구의 필요성과 목적, 연구 과제, 연구의 구성 및 연구의 제한점을 밝힌다. 제 2장에서는 영어과 수행평가, 말하기 수행평가의 효용성, 말하기 수행평가의 전문성, 말하기 수행평가의 시행한계성 및 교사의 평가 시행에 영향을 미치는 요인에 관한 선행 연구들을 통해서 본 연구를 위한 이론적 배경을 제시한다. 제 3장은 본 연구에 앞서 수행된 예비 연구의 과제와 방법 및 분석 결과를 간략하게 언급한 후 본 연구를 위한 보완점을 제시한다. 제 4장은 연구 방법에 관한 것으로서 본 연구를 위한 연구 대상, 연구 도구, 그리고 자료의 수집 절차 및 분석 방법을 구체적으로 제시한다. 제 5장에서는 다양한 통계 분석을 통해서 도출된 본 연구의 결과를 선행 연구와 연계하여 논의한다. 제 6장은 각 연구 과제에 대한 결과들을 토대로 본 연구의 결론 및 이에 따른 제언을 한다.

1.5 연구의 제한점

본 연구는 다음과 같은 제한점들이 있다. 첫째, 본 연구는 학교 현장에서 말하기 수행평가를 시행하는 영어교사 집단만을 대상으로 수집한 자료를 분석하였다. 차후 본 연구의 과제와 관련하여 말하기 수행평가를 시행하지 않는 영어교사 집단과의 비교 연구가 수행된다면, 말하기 수행평가에 대한 두 집단 간의 인식이 구체적으로 어떻게 유사하고 어떠한 차이가 있는지에 대해서 교육적으로 보다 의미 있는 정보를 얻는 데 도움이 될 것이다.

둘째, 본 연구는 질적 연구 방법의 일환으로서 자유서술식 응답의 결과를 분석하여 연구 결과와 연계시켜 보완하였음에도 불구하고 설문지 조사의 결과에 기초한 양적 연구 방법의 비중이 높은 편이다. 차후 본 연구의 과제와 관련하여 면접이나 관찰과 같은 보다 다양한 질적 연구 방법이 보완된 후속 연구도 필요하리라 본다.

셋째, 본 연구에서 말하기 수행평가의 시행도에 대한 신뢰도는 .436으로서 다소 낮게 나왔다. 이는 말하기 수행평가의 시행도라는 요인을 측정하는 문항이 단지 두 개에 불과하기 때문에 본 연구 도구의 신뢰도에 영향을 미친 것으로 판단된다. 차후 본 연구의 과제와 관련된 후속 연구를 시행할 때는 말하기 수행평가의 시행도를 측정하는 문항을 최소한 세 문항 이상으로 구성하여 연구 도구의 신뢰도를 보다 높일 필요가 있을 것이다.

II. 이론적 배경

우리나라에서 ‘수행평가’라는 용어는 1996년에 출간된 『수행평가의 이론과 실제』라는 책에서 본격적으로 소개되었다(조영기, 2006). 그 후 수행평가가 많은 시행착오를 거쳐 오늘날의 학교 현장에 자리를 잡기까지 수행평가에 대한 이론적인 측면뿐만 아니라 실제적인 측면에서 많은 연구들이 쏟아져 나왔다. 이 장에서는 영어과 수행평가, 말하기 수행평가의 효용성, 말하기 수행평가의 전문성, 말하기 수행평가의 시행한계성 및 말하기 수행평가에 영향을 미치는 변인들에 대한 이론과 선행 연구들을 주로 고찰하고자 한다.

2.1 영어과 수행평가

영어과 수행평가가 등장하게 된 배경은 1980년대 이후 학습자의 언어적 능력보다는 의사소통 능력이 중요시되면서 언어 평가 역시 의사소통적 언어 평가에 대한 관심으로 바뀌면서였다. 이는 언어를 읽기, 쓰기, 듣기 및 말하기라는 네 기능으로 나누어 그 수행 능력의 수준에 따라 각 언어 기능별 평가를 할 수 있다(Heaton, 1988)는 주장을 근거로 우리나라에서 영어과 수행평가를 도입하는 계기가 되었다. Nunan(1988)은 의사소통기능을 중시하고 학습자의 언어수행 능력 그 자체를 강조하는 이론과 실질적인 방법론의 변화에 따라 오늘날의 교육과정도 학습자 중심으로 옮겨가고 있음을 강조하였다. 이러한 시대적 경향을 반영한 의사소통적 언어 평가는 실제적인 언어사용에 대한 수행 능력을 측정하고 진정성(authenticity) 있는 평가를 지향하는 현재의 영어과 수행평가로서 자리매김하게 된 것이다.

이러한 맥락에서 영어과 수행평가는 “학습자의 영어 지식과 기능을 적용하여 실제 생활이나 인위적인 평가 상황에서 요구되는 과제를 수행하는 능력을 관찰 또는 면접 등의 직접적인 방법을 통하여 평가하는 것으로서 수행의 결과뿐만 아니라 그 과정도 평가하는 것”으로 정의된다(최연희, 2000: p. 22). 다시 말해, 영어과에서 수행평가는 영어를 통해서 이루어지는 과업 수행의 모든 과정에 대한 다면적인 평가라고 할 수 있다.

학습자의 의사소통 능력을 평가하는 언어 평가의 특징에 대한 연구(최연희, 2000; Bachman, 1990; Weir, 1990)에서 전통적 평가와 비교해서 의사소통적 언어 평가(communivative language testing)는 교사와 학습자 또는 학습자 간의 상호작용 및 실제 상황에 기반을 둔 수행에 대해서 직접 평가(direct testing)의 시행이 가능하다고 하였다. 따라서 기존의 지필평가가 단순한 언어지식에 대한 획일적이고 수동적인 평가였다면, 영어과 수행평가는 학습자의 동기나 필요에 초점을 두고 학습자가 중심이 되어 문제를 해결하는 과정에 대한 창의적이고 능동적인 평가라고 할 수 있다.

영어과 수행평가가 지닌 교육적 특징에 대해서 정리하자면 다음과 같다. 첫째, 영어과 수행평가를 통해서 가능한 한 실제 상황과 거의 유사한 자연스러운 상황에서 학생들의 의사소통 능력의 함양이라는 교육 목표의 달성여부와 학습내용의 성취정도를 파악할 수 있다(최연희 외, 1998). 둘째, 영어과 수행평가는 직접적인 수행에 의거한 언어 평가(performance-based language testing)로서 평가의 타당성을 제고할 수 있다. 셋째, 영어과 수행평가는 여러 가지 지식과 기능을 통합한 과제로 제시할 수 있어서 학생들의 영어능력을 균형 있게 측정하여 평가의 신뢰성을 높일 수 있다. 넷째, 영어과 수행평가는 교수·학습의 결과뿐만 아니라 과정까지 종합적으로 평가하므로 높은 예측 타당도(predictive validity)를 지닐 수 있다. 다섯째, 영어과 수행평가는 단순히 평가로서의 역할에만 그치는 것이 아니라 영어교육에 중요한 환류효

과를 주는 역할을 한다(McNamara, 2000). 즉, 학생들에게 큰 영향을 미치는 영어 평가에 수행평가와 같은 의사소통 능력을 측정하는 영역이 포함된다면 아무래도 수행평가를 통한 학생들의 의사소통 능력이 향상될 가능성이 높을 것이다.

이와 같은 영어과 수행평가의 특징으로 비춰봤을 때, 의사소통 중심의 평가로서 영어과 수행평가가 갖추어야 할 가장 우선적인 요건은 수행평가의 시행 시 학습자의 실제적인 영어 사용능력을 의미 있게 측정하고 있는가에 관한 타당성 여부이다. 의사소통적 언어 평가의 타당성에 대해서 Canale과 Swain(1980)은 제 2언어에 대한 의사소통 능력을 평가할 때는 반드시 학습자가 알고 있는 지식과 그 지식을 사용하는 방법에 대해서뿐만 아니라, 학습자가 의미 있는 의사소통적 상황에서 그 지식을 실제로 사용할 수 있는지에 대해서도 평가가 이루어져야 함을 강조하였다. 그렇다면 영어과 수행평가도 마찬가지로 학생들이 영어에 대해서 알고 있는 지식인 영어 능력(competence)을 실제와 거의 유사한 상황에서 얼마나 성공적으로 잘 수행(performance)하는지를 측정해서 평가해야 타당성이 있다고 볼 수 있다. 이는 학습자들의 언어능력에 대한 실제적이고 직접적인 수행을 지향하는 평가는 학습자들의 언어 수행을 성공적인 결과로 연결시켜 주기 때문이다(Clark, 1972; Jones, 1977; Morrow, 1977).

아울러 학교 현장에서 영어과 수행평가의 시행을 위한 또 다른 중요한 요건들에 대한 필요성도 제기되었다. 우선 의사소통과 관련한 수행평가의 시행자가 갖추어야 할 요건에 대해서 한국교육과정평가원(1999)은 다른 언어 기능의 평가자보다 실제적인 담화 능력이 더 우수해야 하고, 평가 시 단호하고 분명한 태도를 지녀야 하며, 또한 피평가자가 평가를 시행할 때 적극적인 참여 자세를 보여야 함을 강조하였다. 그리고 말하기 수행평가를 원활하게 진행하기 위해서는 평가 대상자가 평가를 받는 장소의 조명, 음향 및 안락감

등 시설에 대한 환경적인 요건도 충족시키는 것이 중요하다(교육과학기술부, 2008). 게다가 Weir(1990)는 실생활과 관련된 상황, 정보 격차의 관련성, 상호작용, 자기 수정의 기회 제공, 적절한 분량의 입력 처리 및 적절한 시간의 제한 등 총 여섯 가지를 제시하였다. 그러므로 학교 현장에서 학생들의 영어 능력을 향상시킬 수 있는 진정한 수행평가를 시행하기 위해서는 평가자가 갖추어야 할 자격 요건과 환경적 요건은 물론이고, 또한 말하기 수행평가의 시행과 관련한 여러 기타 요건들도 고려해야 할 필요가 있다.

이와 같은 선행 연구들을 통해서 영어과 수행평가는 학습자의 능동적이고 실제적인 언어사용에 대한 모든 과정을 다면적으로 측정하는 진정성 있는 언어 평가라는 것을 알 수 있다. 그리고 영어과 수행평가는 의사소통 능력의 향상을 위한 교수·학습 과정에서 높은 타당성과 진정성 및 긍정적인 환류효과 등 여러 가지 교육적 특징을 지니고 있음을 알 수 있다. 또한, 의사소통 중심의 평가로서 영어과 수행평가를 잘 시행하기 위해서는 평가의 타당성 확보와 이를 위한 평가자, 평가 환경 및 평가 시행과 관련한 여러 요건들의 충족이 선행되어야 함을 확인할 수 있었다.

2.2 말하기 수행평가의 효용성

이 절에서는 영어교육의 현장에서 말하기 수행평가의 시행이 지닌 교육적 효과에 관한 선행 연구들을 성취도와 정의적 및 수업 개선의 측면에서 살펴보고자 한다.

2.2.1 성취도 측면

언어의 네 기능 중 말하기의 특성을 언급한 김인철(1998)은 말하기라는 능력은 의미 전달의 유창성(fluency)과 정확성(accuracy)의 함양을 위해서는 실제 상황에서만 연습이 가능하고, 특히 수행평가를 통해서만 학습자의 말하기 성취도에 대한 직접적인 측정이 가능하다고 하였다. 즉, 외국어로서의 영어를 학습하는 우리나라의 학교 현장에서 유창성과 정확성을 고려한 실제성이 있는 말하기 수행평가를 자주 시행한다면, 영어교사들은 학생들의 말하기 능력의 성취 수준과 발달 과정을 잘 파악할 수 있을 것이다.

학생들의 수업 참여와 학업 성취도와의 연관성에 관한 연구(Lim, 1992; Wudong, 1994; Zhou, 1991)의 결과에 따르면, 일반적으로 학생들의 수업 참여는 말하기, 듣기, 읽기, 쓰기, 몸짓 언어나 신체 활동 등 많은 형태의 행동 양상으로 나타났다. 이들 중 주목할 만한 수업 참여 방법은 구두 참여(oral participation) 방법으로 학생들의 학업 수행에 가장 효과적인 것으로 보고되었다. 다시 말해, 학생들이 수업시간에 말하기 수행평가를 통해서 말하기 활동에 활발하게 참여할 기회를 많이 가진다면, 말하기에 대한 그들의 성취도 또한 높아질 것이다. Swain(1985)도 제 2언어의 습득에 있어서 말하기 능력의 향상을 위해서는 학습자들이 말을 할 수 있는 기회를 가급적 많이 가지는 것이 중요하다고 하였다. 게다가 다수의 학자들은 언어 습득과 수업에서의 상호작용 사이의 관계에 관한 실험적 연구(Busch, 1982; Day, 1984; Ellis, 1993; Ely, 1986; Spada, 1986)를 통해서 수업 시간에 구두적인 상호작용(oral interaction)에 들어가는 시간의 양은 언어 습득의 정도와 긍정적인 관계가 있음을 확인시켜 주었다.

또한, 말하기 수행평가와 영어 말하기 능력의 신장과의 관계에 대해서 비교반과 실험반으로 나누어 학생들의 인식을 연구한 연구에서 배원기(2004)는

말하기 수행평가는 학생들의 영어 말하기 능력의 신장에 도움이 된다는 결과를 얻었다. 아울러 말하기 수행평가의 효용성에 대한 중·고등학교 영어교사들 간의 인식을 비교한 연구(고현우, 2014)의 결과, 입시를 중시하는 고등학교보다 실용영어를 중시하는 중학교는 말하기 수행평가를 더 자주 시행할 수 있어서 말하기 수행평가는 학생들의 영어 말하기 성취 수준을 파악하는데 도움이 됨을 보고하였다.

이와 같은 선행 연구들을 통해서 학생들의 영어 말하기 성취 수준을 정확하게 측정하기 위해서는 말하기 수행평가를 반드시 시행해야 함을 알 수 있다. 그리고 가급적 영어 수업시간에 학생들이 영어로 말을 할 기회를 자주 제공하여 말하기 수업에 대한 참여도를 높이고, 또한 말하기 수행평가를 자주 시행함으로써 학생들의 영어 말하기 능력의 발달 과정을 꾸준히 파악할 수 있음을 알 수 있다. 결국 학교 현장에서 학생들의 영어 말하기 능력을 신장시킬 수 있는 가장 효과적인 방법은 바로 영어교사들이 말하기 수행평가를 시행하는 일이라는 것을 예측할 수 있다.

2.2.2 정의적 측면

영어 말하기 수업에 대한 인식과 관련된 연구로서 이문복 외(2012)에서 조사한 바에 따르면, 고등학교 학생들은 대부분이 학교 영어수업에서 쓰기보다는 말하기의 중요성에 공감하였다. 그 이유로서 영어 말하기는 해외에 나가거나 외국인과의 직접 의사소통을 해야 할 때 반드시 필요한 도구라고 인식하고 있기 때문이었다. 그리고 학교의 영어 말하기 수업이 영어 실력의 향상에 도움이 되는지에 대한 질문에 대부분의 학생들이 약간 향상되었다고 하였고, 또한 영어 말하기 수업이 긍정적인 동기 부여가 되어 영어 수업이 예전보다 더 재미있어 졌다는 대답도 있었다.

영어 말하기 수행평가의 시행이 학생들에게 미치는 정의적 효과에 관한 연구에서 윤말임(2001)은 학생들을 대상으로 역할극과 인터뷰로 말하기 수행평가를 시행한 결과, 영어 말하기에 대한 흥미와 자신감이 생겼다고 보고하였다. 배원기(2004)도 역시 말하기 수행평가를 통해서 학습자들의 학습 흥미와 학습 욕구 및 자신감과 같은 학습 태도가 긍정적으로 변하게 되었다는 결과를 도출하였다. 게다가 말하기 수행평가의 시행 후 학생들의 말하기 능력이 실제로 신장되었고 영어에 대한 자신감 또한 크게 상승했다는 은경자(2010)의 연구 결과도 이들과 일맥상통하고 있다. 즉, 말하기 수행평가는 학생들의 수업에 대한 적극적인 참여를 통해서 영어 말하기에 대한 자신감을 심어주며, 또한 학생들이 혼자서 말하기 연습을 많이 하도록 만드는 일종의 촉매제 (facilitator) 역할을 하는 것으로 보인다.

또한, 최주경(2009)도 영어 교사가 정확한 기준을 가지고 말하기 수행평가를 시행하게 되면 학생들의 태도가 좋아지고, 결과적으로 학생들의 의사소통 능력도 점점 향상된다는 결과를 얻었다. 이와 비슷한 연구에서 윤소영(1999)은 지필평가에 비해서 수행평가는 학생들의 학습 태도에 대한 부정적인 영향을 상대적으로 적게 준다고 하였다. 그리고 교사와 학생 모두 수행평가에 긍정적인 반응을 보였고, 특히 학생들은 수행평가에 흥미를 가지고 주어진 과제를 해결해가는 것으로 조사되었다. 더구나 영어 수업시간에 말하기를 통한 학습 동기의 신장에 관한 연구를 한 MacIntyre 외(2001)는 학교의 정규 수업시간에 교과서에서 나오는 소재로 말하기 과제를 수행하게 되면 학생들의 의사소통에 대한 의지가 증가한다고 보고하였다.

이러한 선행 연구들을 통해서 영어 말하기 수행평가는 학생들에게 다양한 활동을 제공하여 말하기에 대한 동기와 흥미, 자신감 및 학습 태도의 상승으로 연결하는 중요한 매개체 역할을 수행함을 알 수 있다. 왜냐하면 수업시간에 이루어지는 평가는 교사가 학생들에게 학습 동기를 불어넣어 수업의 질

을 향상시키는 연속적인 활동이기 때문이다(Gronlund, 2006). 따라서 말하기 수행평가는 학생들이 긍정적인 학습 태도를 가질 수 있고, 궁극적으로 학생들의 말하기 능력을 배양하는 데 도움을 주는 훌륭하고 바람직한 평가임을 확인할 수 있다.

2.2.3 수업 개선의 측면

평가란 교사의 수업과 같은 교육 프로그램의 개선을 위한 의사결정에 있어서 필요한 정보를 모으거나 이용하는 과정이다(Cronbach, 1982). 그리고 평가는 교수·학습 과정의 적절성을 가늠하는 잣대이며 교육 내용과 수단에 강력한 영향을 미치는 교수설계의 구성요소이다. 이처럼 평가가 교육의 대상과 주체 및 교육 내용과 형식 등 모든 교육과정의 결정에 영향을 미치는 현상을 언어 평가에서는 환류효과라고 한다. Oller(1979)는 학습자의 학습 과정을 촉진하기 위해서 평가를 적극 활용할 것을 주장하면서, 진정한 타당도를 지닌 평가를 개발하려면 교수과정과 평가가 밀접한 관련성이 있어야 함을 강조하였다. 이러한 평가의 교육적 가치가 학습자에게 미치는 직접적인 효과에 대한 용어로 발전한 것이 바로 환류효과이다.

원래 평가의 환류효과는 중립적인 개념으로서 교육 내용이나 방법에만 영향을 미치는 범위로만 한정하였으나, 최근에 들어서는 교육 참여자뿐만 아니라 교육 체계 전반에 미치는 영향까지 포괄하고 있다(Bachman & Palmer, 1996). 가령, 우리나라와 같이 대학입시 중심의 교육 환경이나 사회적인 책임도가 큰 시험이 존재하는 경우 평가의 환류효과로 인해서 시험의 준비과정에서 결과까지 학습자들에게 큰 영향을 미치게 된다. 또한, 환류효과는 시험이 교사에게 미치는 직접적인 영향으로도 규정될 수 있는데(Buck, 1988),

이는 교사가 시험에 출제될 범위와 내용을 바탕으로 자신의 수업을 구성하도록 영향을 받게 됨을 의미한다. Swain(1984)은 학교에서 교사들은 시험의 내용과 형식에 따라 학생들을 가르치는 경향이 있다고 지적하였다. 이는 의사소통 능력의 배양을 중시하는 교육과정에서 모든 학교들이 말하기 수행평가의 비중을 높이게 되면 그에 따라 영어교사의 수업 방법 또한 바뀌게 될 가능성을 시사하고 있다. 이와 같은 맥락에서 Pearson(1988)도 시험은 학습에 대한 학습자와 교사 및 학부모의 태도와 행동까지에도 영향을 끼친다고 주장하였다.

무엇보다도 환류효과가 한 사회의 교육체제 전반에 영향을 미치는 가운데 가장 큰 영향을 받는 대상은 바로 수험자이다. Bachman과 Palmer(1996)는 평가의 환류효과가 수험자에게 미치는 영향을 세 가지 측면에서 설명하였다. 즉, 시험을 치른 경험과 시험을 준비한 경험, 시험의 수행 능력에 대해서 받은 피드백 및 시험 점수에 근거하여 그들에게 내려질 의사결정이다. 우선 시험을 치른 경험과 시험을 준비한 경험의 측면에서 보았을 때, 학습자가 사회적으로 중요한 시험을 준비하는 과정에서 축적된 경험은 앞으로의 언어 사용과 사고 과정에까지 지대한 영향을 미친다는 것이다. 그리고 시험의 수행 능력에 대해서 수험자가 받은 피드백의 측면에서는 언어 평가의 경우 주로 점수의 형태로 주어지는데, 시험 결과에 대한 피드백에 따라 수험자의 언어 지식이 향상되거나 오히려 그 반대로 더 떨어질 수도 있다. 또한, 시험 점수에 근거하여 수험자들에게 내려질 의사결정의 측면을 보면, 언어 평가의 결과인 점수가 입학이나 취업과 같은 한 개인의 운명을 결정하는 데 큰 영향력을 행사하는 경우가 많다. 이러한 이유로 Bachman과 Palmer(1996)는 국가적으로나 사회적으로 중대한 시험을 개발할 때는 평가의 진정성 확보를 위해서 시험의 설계과정에 교사뿐만 아니라 수험자도 반드시 참여시킬 것을 주장하였다.

이와 같은 선행 연구들을 통해서 평가의 환류효과는 모든 교수·학습의 주체와 대상에게 중요한 영향을 미친다는 것을 알 수 있다. Shohamy(1993)는 언어 평가는 학생들에게 동기나 흥미를 부여하고 교사의 교수 활동을 촉진시키는 등 교육이라는 활동이 일어나는 모든 과정에 그 영향력을 행사하고 있음을 역설하였다. 이러한 언어 평가의 환류효과를 고려했을 때, 우리나라의 대학수학능력시험에서 말하기 평가가 꼭 반영될 필요가 있다. 그렇게 된다면 영어교사들의 수업 방법은 현 교육과정과 연계된 진정한 의사소통 중심으로 개선될 여지가 있을 것이다.

2.3 말하기 수행평가의 전문성

교육 현장에서 가장 관심을 많이 받는 영역 중 하나는 교사의 학생 평가이다. 또한, 학교에서 교사들이 가장 많은 시간을 투자하는 일들 중 하나도 또한 평가관련 업무이다. 교육의 과정과 결과에 중요한 영향을 미치는 평가에 대한 교사의 전문성은 수업의 효과를 극대화하고 교사의 수업 방법을 개선하여 학생의 성장과 발달 과정을 도와주는 필수적 요소임을 부인할 수 없다. 과거 학생에 대한 평가는 목표와 결과, 비교와 서열 및 지필평가 중심의 제한적인 평가였고, 또한 학생에 대한 평가 결과를 유용하게 활용하지 못하는 등 평가로서의 기능을 제대로 수행하지 못하였다. 그러다가 최근에서야 평가의 교육적 효과에 대한 관심의 증대로 인해 기존의 평가 방법에 대한 대안적 평가로서 수행평가가 등장하게 되면서부터 학생의 능력과 특성을 다양하게 평가할 수 있게 되었다(배호순, 2000). 이에 따라 학생의 학습 과정과 결과에 대한 다면적인 평가가 가능하고 평가 결과를 적절히 활용하는 수행평가에 대한 교사의 전문성 또한 요구되는 것은 당연할 것이다.

그럼에도 불구하고 영어교사들은 평가와 관련된 훈련이나 경험을 받은 적이 거의 없을 뿐더러 평가와 관련된 일을 수행하는 데 있어서 준비가 부족한 것으로 조사되었다(김신영, 2002b; Masole, 2011; Schafer, 1993; Zhang & Burry-Stock, 2003). 이 결과는 영어과 평가의 특성 상 언어능력이라는 것이 다소 추상적이어서 평가의 내용을 선정하기가 어렵고, 영어에 대한 지식보다는 영어를 사용할 수 있는 능력을 측정해야 하며, 그리고 비영어민인 한국인 영어교사들이 본인부터 완벽한 영어를 구사하지 못하기 때문에 평가도구를 제작하고 채점하는 과정에서 항상 부담감을 느끼기 때문일 것이다(김수동, 이의갑, 2005). 또한, 교사들을 위한 연수 프로그램은 실제로 학생 평가와 관련하여 발생한 문제를 해결하는 지식이나 기술에 대한 내용과 동떨어져 있고(Stiggins, 1991), 특히 수행평가의 시행을 위해서 교사들이 갖춰야 할 전문적인 지식과 기술 및 활용 방안에 대한 구체적인 지침이나 연수에 관한 내용은 거의 없는 실정이다.

Zhang과 Burry-Stock(2003)은 어떠한 학습 환경에서 교사가 학생 평가를 효과적으로 시행할 수 있느냐의 문제는 바로 교사가 이수한 평가 연수의 질(the quality of assessment training)에 달려있다고 하였다. Howie(2006)와 Stiggins(2002)도 역시 학생들을 평가하기 위한 연수 등의 경험적 지식이 부족한 교사는 평가 시행에 대한 전문적인 능력이 떨어진다고 하였다. 다시 말해, 평가 연수의 경험이 있는 교사가 그렇지 못한 교사에 비해서 학생 평가에 대한 전문성이 높으므로(이정애, 2003), 학생들을 대상으로 시행하는 평가에 대한 준비와 계획, 시행 및 결과의 활용 등에 있어서의 교육 전문가로서 영어교사의 평가 전문성은 꼭 필요하다.

교사의 학생 평가 전문성(teacher competence in student assessment)이란 학생의 학습과 성취에 관한 평가 정보를 수집하고 해석하여 활용할 수 있는 능력이다(김수동 외, 2005). 그리고 교사가 갖추어야 할 학생 평가 전

문성은 평가 활동을 위한 기본적이고 핵심적인 능력이기 때문에 평가 소양 (assessment literacy for teacher)이라고도 불린다(Plake, Impara & Fager, 1993; Schafer, 1993). 즉, 교사의 학생 평가 전문성은 모든 교육활동과 관련하여 학생을 평가하는 과정에서 교사가 반드시 알고 있는 지식뿐만 아니라 이를 실천할 수 있는 능력과 자질을 말한다. 그러므로 영어교사들의 학생 평가 전문성은 평가 도구를 선정하고 개발하며, 평가 결과를 도출하여 타당하게 해석하고, 평가 결과를 수업에 반영하여 학생이나 학부모에게 적절한 피드백을 제공하는 등 학생 평가의 전 과정에 필요한 전문적인 능력을 포괄하고 있다(전병만 외, 2005).

영어교사의 학생 평가 전문성에 대한 체계적인 연구는 많지 않지만, 국내외의 주요 전문가나 교육관련 기관에서 연구한 학생 평가 전문성을 구성하는 요소에 대해서 살펴보고자 한다. 우선 학교 교육에 있어서 교수·학습의 질적 향상은 평가자의 질적 요소에 따라 좌우된다는 관점에서 김신영(2002a)은 6개 영역의 학생 평가의 전문성 구성 요소를 제시하였다. 첫째는 건전한 평가의 질에 대한 이해 능력, 둘째는 평가하고자 하는 방법과 대상의 특성에 대한 이해 능력, 셋째는 다양한 평가 도구에 대한 이해 능력, 넷째는 교실 내 평가의 상호관계적인 측면에 대한 이해 능력, 다섯째는 평가 결과의 활용 방법에 대한 이해 능력, 그리고 마지막으로 교육적 의사결정의 수단으로서 평가를 교실 내에서 활용하는 능력이다.

그리고 김수동 외(2005)에서는 교사가 학생을 대상으로 평가 활동을 수행할 때 발휘해야 하는 전문적인 능력의 구성 요소를 5가지 영역으로 제안하였다. 첫째는 평가 방법의 선정 능력, 둘째는 평가 도구의 개발 능력, 셋째는 평가의 실시·채점·성적 부여 능력, 넷째는 평가 결과의 분석·해석·활용의 사소통 능력, 그리고 마지막으로 평가의 윤리성 인식 능력이다.

한편, 1990년에 미국 교사 연합회(American Federation of Teachers: AFT)와 교육 측정 위원회(National Council on Measurement in Education: NCME) 및 국립 교육 연합회(National Education Association: NEA)에서 공동으로 학생 평가의 시행을 위해서 반드시 필요한 영어교사의 전문적 능력 요소를 7가지 영역으로 나누어 제안하였다. 첫째는 교수 결정에 적합한 평가 방법을 선택할 수 있는 능력, 둘째는 교수 결정에 적합한 평가 방법을 교사가 직접 개발할 수 있는 능력, 셋째는 외부에서 개발한 평가 도구와 교사가 개발한 평가 도구를 실시하고 채점하며 그 결과를 해석할 수 있는 능력, 넷째는 학생 평가를 활용하여 타당한 성적 부여 절차를 개발할 수 있는 능력, 다섯째는 개별 학습자, 수업 계획, 교육과정 개발 및 학교 개선 등에 관한 결정을 할 때 평가 결과를 적절하게 활용할 수 있는 능력, 여섯째는 평가 결과를 학생, 학부모, 관련 인사 및 동료 교사 등에게 의사소통할 수 있는 능력, 그리고 마지막으로 비윤리적이거나 적절치 못한 평가 방법과 평가 결과의 사용이 미치는 폐해를 인식할 수 있는 능력이다.

이상으로 언급된 교사의 학생 평가 전문성 구성 요소들은 공통적인 내용을 담고 있지만 평가 환경과 평가 목적에 따라 서로 조금씩 차이를 보이고 있다. 우선 김신영(2002a)이 제시한 6가지 평가 전문성 구성 요소들은 평가 전문성의 전반에 대한 인지적 이해에 초점을 두고 있다. 반면, 김수동 외(2005)에서 제시한 5가지 평가 전문성 구성 요소들은 AFT, NCME, 그리고 NEA(1990)에서 제시한 전문성 구성 요소들을 바탕으로 하면서 이인제 외(2004)의 연구 자료에서 조사된 교사의 평가 능력이 요구되는 업무를 기준으로 추출된 내용으로서, 주로 평가와 관련된 실행적인 측면에서 교사가 지녀야 할 구체적인 행동 영역을 기술하고 있다.

본 연구에서는 위의 연구들에서 제시한 평가 전문성 구성 요소들을 평가 시행에 대한 시기적인 관점에서 크게 평가의 준비 단계, 평가의 시행 단계, 그

리고 평가 결과의 활용 단계 등 세 단계로 구분하였다. 각 단계별로 평가의 시행에 있어서 교사들이 발휘해야 하는 전문성의 구성 요소들은 다음과 같다. 첫째, 평가의 준비 단계에서는 평가의 시행을 위한 평가 방법이나 도구를 선정하고 개발하는 일에 대한 교사의 전문성이 필요하다. 즉, 김신영(2002a)의 경우 건전한 평가의 질, 평가하고자 하는 방법과 대상의 특성 및 다양한 평가 도구에 대한 이해 능력이, 김수동 외(2005)의 경우는 평가 방법의 선정과 평가 도구의 개발 능력이, 또한 AFT, NCME, 그리고 NEA(1990)에서는 교수 결정에 적합한 평가 방법을 선택하거나 교사가 직접 개발할 수 있는 능력이 이에 해당되는 것으로 본다. 둘째, 평가의 시행 단계에서는 평가의 시행을 위해서 학생들에게 평가에 관한 세부사항을 공지하고 평가 방법과 채점 기준을 정확하게 숙지한 후 채점하는 일에 대한 교사의 전문성이 필요하다. 즉, 김신영(2002a)의 경우 교실 내 평가의 상호관계적인 측면에 대한 이해 능력이, 김수동 외(2005)의 경우는 평가의 실시, 채점 및 성적 부여의 능력이, 또한 AFT, NCME, 그리고 NEA(1990)에서는 교사나 외부에서 개발한 평가 도구를 실시하고 채점하여 그 결과에 맞게 해석 및 성적을 부여할 수 있는 능력이 이에 해당되는 것으로 본다. 셋째, 평가 결과의 활용 단계에서는 평가를 시행한 후 평가 결과에 대한 타당한 해석을 통해서 차후 교수·학습의 의사결정 자료로 활용하는 일에 대한 전문성이 필요하다. 즉, 김신영(2002a)의 경우 평가 결과의 활용 방법을 이해하고 교육적 의사결정의 수단으로서 평가를 교실 내에서 활용하는 능력이, 김수동 외(2005)의 경우는 평가 결과의 분석, 해석, 활용 및 의사소통 능력이, 또한 AFT, NCME, 그리고 NEA(1990)에서는 평가 결과를 활용하여 교육과 관련한 여러 인적·물적 환경을 개선하고 평가 관계자들과 의사소통할 수 있는 능력에 대한 전문성이 이에 해당되는 것으로 본다.

참고로, 김수동 외(2005)가 제시한 평가의 윤리성 인식 능력과 AFT, NCME, 그리고 NEA(1990)에서 제시한 비윤리적이거나 적절치 않는 평가 방

법과 평가 결과의 사용이 미치는 폐해를 인식하는 능력이라는 전문성 구성 요소는 평가 시행과 관련한 실행적인 측면이 아니라 가치판단의 측면에서 바라본 교사의 전문적인 능력으로 간주되어 본 연구에서는 제외되었다.

지금까지 영어교사의 학생 평가 전문성의 구성 요소에 대한 국내외의 연구들을 살펴보았다. 이들 중 어느 한 가지만을 채택하여 절대적인 평가 전문성의 기준으로 삼거나, 혹은 미국의 AFT, NCME, 그리고 NEA(1990)가 제시한 전문성 요소들을 우리나라의 상황에 그대로 적용하는 것은 무리가 있다고 생각된다. 따라서 교사의 평가 전문성이 교육의 질적 향상과 밀접하게 관련되어 있다는 점을 감안하여 영어교사들은 말하기 수행평가를 시행할 때 우리의 교육환경에 최적화된 학생 평가의 전문성 요소들을 평가의 기준으로 삼고 실천해야 할 것이다.

2.4 말하기 수행평가의 시행 한계성

말하기 수행평가는 그 교육적 효용성이 높음에도 불구하고 수많은 현실적인 문제나 한계점들로 인해 성공적인 시행에 있어서는 사실상 많은 어려움이 따르는 것으로 보고되고 있다. 따라서 이번 절에서는 영어교육의 현장에서 말하기 수행평가의 운영 실태와 말하기 수행평가의 시행을 어렵게 만드는 여러 가지 한계점들에 관한 선행 연구들을 살펴보고자 한다.

우선 말하기 수행평가의 운영 실태에 관한 선행 연구들을 살펴보면, 김성연과 이호경(2013)의 연구에 참여한 대부분의 중학교 영어교사들은 듣기나 따라 읽기를 위주로 말하기 수업을 진행하고 있었고, 이에 따라 말하기 수행평가의 과업은 다양했지만 주로 진정성이 없는 암기식 발표 위주의 평가가 지배적이어서 그 진정성이 떨어진다고 밝혔다. 한편, 많은 고등학교에서 시

행되는 수행평가는 주로 듣기 평가나 어휘 평가가 지배적이었고 말하기는 거의 평가에 반영하지 않았다(장인순, 2004). 더구나 고등학교에서의 말하기 수행평가는 수업이나 평가 체제가 대체로 대학입시를 위주로 운영되기 때문에(방인자, 천세영, 2011) 진정한 의사소통 중심의 말하기 교육이 제대로 이행되기가 힘든 상황을 여실히 보여주고 있었다.

최근 중·고등학교에서 말하기 수행평가가 잘 시행되지 못하는 주요한 원인들을 조사한 결과, 학급 규모의 크기나 학생 수의 과다, 채점의 객관성 확보의 어려움, 적절한 말하기 수행평가 도구의 부재 및 학생에게 또 다른 학습 부담의 제공 등의 이유로 영어교사들은 말하기 수행평가의 시행이 어려움을 호소하고 있었다(고현우, 2014; 이소영, 2008). 그리고 이은주(2003)는 영어교사의 말하기 능력의 부족, 한 학급의 학생 수의 과다, 말하기 수업자료의 부족 및 원어민 영어교사의 부재 등의 이유로 말하기 수업과 평가를 진행하기가 어려움을 털어놓았다. 특히, 김미은(2002)은 고등학교에서 말하기 수행평가의 시행이 어려운 원인으로서 교사의 업무 과다, 학급당 인원 수의 과다 및 지필 고사 위주의 대학 입학시험 등을 사유로 들었다. 또한, 최윤경(2001)은 많은 수의 학생들을 대상으로 한 명의 교사가 말하기 수행평가를 위한 다양한 활동 과제를 제시하면서 각각의 활동에 대한 채점 기준을 정확히 제시하기가 힘들다는 점이 학생들의 말하기 능력을 개별적으로 평가하는데 있어서 큰 걸림돌이 된다고 하였다.

이처럼 말하기 수행평가를 시행하는 데 있어서 학급 규모가 크거나 평가 대상자인 학생 수가 많다는 한계점은 아무래도 채점의 신뢰도와 높은 연관성이 있을 것이다. 말하기 수행평가 시행의 신뢰성의 측면에서 Weir(1990)은 말하기 평가를 가장 어렵게 만드는 것은 어떻게, 어떤 기준으로, 그리고 누가 채점하는가에 따른 신뢰도의 문제가 가장 크다고 주장하였다. 그리고 Brown(1994)도 언어 평가에서 말하기가 아주 중요하지만 그 기능상 복잡한

특성 때문에 의사소통 능력의 평가에 있어서 신뢰성(reliability)을 확보하기가 가장 어렵고 힘들다고 하였다. 즉, 말하기 수행평가를 시행하고자 할 때 영어교사의 채점 신뢰도가 확보되지 않으면 결국 학생의 말하기 능력보다는 채점 시 채점자 변수가 평가 결과에 영향을 줄 위험이 있다.

말하기 수행평가의 채점 신뢰도에 대한 보다 구체적인 사례 연구를 한 Park(2000)은 말하기 평가 채점 시 학습자의 언어능력보다는 교사가 지니고 있던 수업 중의 학습자에 대한 기억이 채점에 지대한 영향을 미치고 있음을 지적하였다. 그리고 말하기 수행평가의 신뢰성을 떨어뜨리는 주요 원인으로서 교사의 독단적인 채점, 학급당 과다 인원으로 인한 평가 시간의 부족, 평가 점수의 산출에만 목적을 둔 평가 방식, 교수·학습의 내용을 반영하지 못하는 평가 내용, 말하기 평가에 대한 교사의 인식 부족 및 채점자 훈련의 부족 등을 언급하였다. 그리고 Lantolf와 Frawley(1985)는 말하기 수행평가가 짧은 시간에 일대일로 여러 명의 수험자들을 차례로 만나는 인터뷰 평가인 경우 상대평가의 방식으로 이루어져 이에 준하는 규범이 부적절하게 설정되어 있어서 보이지 않는 불이익을 받는 학생들이 많이 생길 수 있다고 하였다.

말하기 수행평가의 시행을 어렵게 만드는 또 다른 요인은 적절한 평가 방법이나 도구가 없다는 점인데, 이 또한 말하기 수행평가의 채점 시 객관성 유지의 어려움과 과다한 학급이나 학생 수와 밀접한 관련이 있는 것으로 보인다. 이소영(2008)은 수행평가의 시행과 관련한 가장 큰 문제점으로서 취지에 맞지 않는 평가 도구의 사용을 지적하면서 상당수의 교사들이 진정한 의미의 수행평가를 시행하지 못하고 있다고 하였다. 즉, 학생 수의 과다로 인하여 채점 시 객관성을 확보하기가 어려워짐에 따라 많은 영어교사들이 보다 손쉽게 채점할 수 있는 방법에 의존함으로써 수행평가라는 본래의 취지를 살리지 못하는 것이다. 이는 앞서 언급된 현재 우리나라 중·고등학교의

말하기 수행평가의 운영 실태에 관한 연구들(김성연, 이호경, 2013; 방인자, 천세영, 2011; 장인순, 2004)을 통해서 여실히 증명되고 있다.

아울러 말하기 수행평가와 관련된 도움이 되는 자료나 연수가 부족하여 말하기 수행평가를 시행하기가 어렵다는 지적도 있다. 문지형(2007)은 영어교사들이 수행평가를 시행하는 데 있어서 나타난 문제점으로서 수행평가를 위한 참고할 만한 자료가 부족하여 수행평가의 준비와 자료 개발에 시간이 많이 소요됨을 지적하였다. 그리고 윤미현(2012)은 현재 우리나라의 영어교과서가 학습자의 수준이나 흥미를 고려하지 않은 채 개발된 경우가 많아서 실질적으로 학생들의 말하기 능력의 평가에 큰 도움이 되지 않음을 언급하였다. 또한, 현직 영어교사들은 기존의 문법과 독해 수업에 익숙해져 있어서 말하기 수업을 진행하는 일에 부담을 느끼고 있으므로 영어교사들을 위한 말하기 수업과 평가에 관한 교사 연수의 강화와 지원을 강조하였다.

마지막으로, 말하기 수행평가는 학생들에게 또 다른 학습 부담이 된다는 이유로 학교 현장에서 말하기 수행평가를 시행하기가 어렵다는 연구들이 보고되었다. 문지형(2007)은 영어교과의 특성상 듣기, 말하기, 읽기 및 쓰기라는 영어의 네 기능에 대한 수행평가를 모두 준비해야 하는 학생들의 심적 부담을 언급하였다. 정기적으로 행해지는 교과 시험뿐만 아니라 내신 성적에 반영되는 수행평가까지 시행하고 성적을 산출하는 일은 영어교사나 학생들 모두에게 큰 부담이 아닐 수가 없는 것이다. 그리고 말하기 수행평가를 단순한 평가 방식들 중 하나로 인식을 하고 있는 관계로 대부분의 영어교사들은 그 반영비율을 10%이상 20%미만으로 정하여 학생들의 부담을 줄이려는 경향이 있는 것으로 나타났다(염지연, 2011). 또한, 박소현(2010)의 연구에서는 말하기 수행평가의 과제 양과 시행 빈도가 과다하다는 점이 문제점으로 지적되었다. 학생들의 말하기 수행평가에 대한 준비 시간이 부족하여 심적인 부담이 클 뿐더러 말하기 수행평가로 인하여 수업시간이 감소한다는 점도

불만으로 제기되었다. 더구나 수행평가에 대한 학생들의 인식 연구(이재호, 2012)에서도 학생들은 말하기 수행평가가 말하기 능력의 발달 과정을 평가하는 것이 아니라 암기식 평가와 별다른 차이가 없어서 말하기 수행평가의 시행 자체에 약간의 부담을 느끼고 있는 것으로 조사되었다.

지금까지 살펴 본 우리나라의 중·고등학교가 안고 있는 말하기 수행평가의 시행 상의 한계점들에 대한 개선 방안을 연구한 선행 연구들은 다음과 같다. 먼저 고현우(2014)와 이소영(2008)은 말하기 수행평가의 취지에 맞는 도구를 다양하게 개발하는 일이 최우선적으로 중요한 방안이고, 그 다음으로 말하기 수행평가의 채점 기준을 구체화하고 객관화하는 일이며, 또한 말하기 수행평가의 여건을 개선하는 일도 필요함을 제안하였다. 김정임(2006)과 염지연(2011) 또한 학생들의 말하기 능력을 효과적으로 측정할 수 있는 다양한 말하기 수행평가 자료의 개발이 가장 시급함을 강조하였다. 그리고 학급 규모의 크기나 학생 수의 과다라는 문제에 있어서 김미은(2002)과 박상수(2004)는 말하기 수행평가의 시행을 위한 가장 시급한 개선 방안으로서 학급당 학생 수를 반드시 감축해야 한다고 주장하였다. 윤미현(2012)과 황정선(2007)은 영어 말하기 수업의 효율성을 높이고 실용영어 능력의 향상을 위해서 말하기 수업과 평가에 관한 교사 연수를 강화할 것과 바람직한 말하기 평가를 위한 평가 환경의 조성을 중요한 해결책으로 꼽았다. 또한, 이문복 외(2012)와 문지형(2007)은 정부나 교육청 차원에서 영어교사들을 위해서 말하기 수행평가에 대한 유용한 자료를 제공하고 수행평가와 관련된 연수를 마련해 줄 것을 주문하였다. 게다가 박소현(2010)은 말하기 수행평가의 준비로 인한 학생들의 학습 부담을 축소하기 위해서 교과 목표 성취에 필요한 최소한의 학습량을 선정하고, 과제물 위주의 평가를 지양하며, 또한 수업 시간을 통해서 학습된 내용만을 본 수업 시간 중에 평가함을 원칙으로 할 것을 제안하였다.

2.5 교사의 평가 시행에 영향을 미치는 요인

학교의 많은 교육활동들 중에서 평가는 학생들의 학습 결과를 확인하고 유용한 학습 정보를 제공하며, 또한 학생들의 학습 과정을 촉진시키는 중요한 역할을 한다. 그래서 학생들의 평가를 책임지는 교사는 평가가 지니는 이러한 교육적 효용성을 고려했을 때, 말하기 수행평가의 시행 상의 어려움이 다소 있더라도 이를 현명하게 극복해내는 전문적인 능력과 자질을 갖추는 것이 중요하다. 왜냐하면, 학생들의 학업 성취에 영향을 주는 요인들 중에서 가장 중추적인 역할을 하는 요인은 바로 평가 주체인 교사이기 때문이다. 즉, 교사는 교육과정을 조절하는 문지기(gatekeeper) 역할을 수행한다고 볼 수 있다(Eisner, 2002). 그러므로 학생 평가를 주관하는 교사가 평가에 대해서 어떠한 신념을 보이는가에 따라서 평가 시행의 양상이 달라질 수 있는 것이다.

평가에 대한 교사의 신념(teachers' beliefs)에 대해서 Brown, Bull, 그리고 Pendlebury(1997)는 학생의 학습(learning)이 바뀌려면 먼저 교사 자신의 평가 방법부터 바뀌어야 한다고 주장하였다. 언어 평가가 학생의 발전적인 변화를 유도하는 것이 목적이라면, 교사는 학생의 언어적 능력에 대한 유용한 평가 정보를 산출할 수 있는 효과적인 평가 방법을 구축해서 사용해야 한다. 그리고 교사는 자신이 시행하는 평가 방법이 학생의 학습 행동에 바람직한 영향을 미칠 것이라는 신념을 가질 때 올바른 평가가 될 수 있는 것이다. 가령, 한 영어교사가 말하기 수행평가는 학생들의 말하기 능력의 향상에 도움을 줄 수 있다고 생각하고 말하기 수행평가의 시행을 위한 타당한 도구나 방법을 채택하는 일은 바로 자신의 신념에서 비롯된 행동인 것이다. 중요하게도, 평가에 대한 교사의 신념과 교사의 평가 시행(assessment practice)과의 상관관계를 연구한 결과, 평가의 시행이 교수·학습에 효과적이라고 믿는 교사들은 평가를 더 자주 시행하고 평가 관련 과제나 업무 또한 더 많이

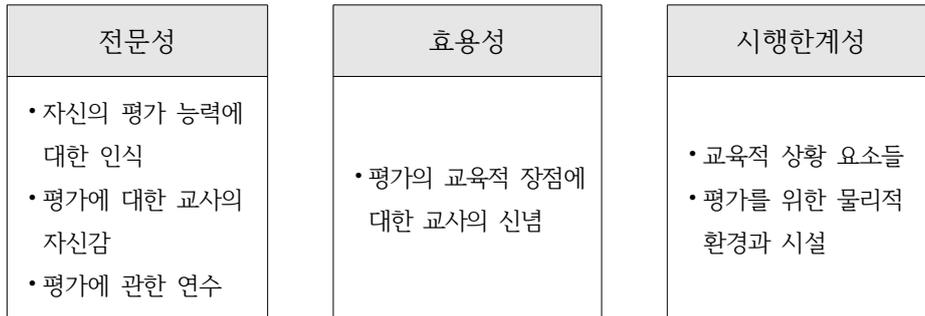
수행하는 경향이 있는 것으로 밝혀졌다(Chan, 2007; Green & Stager, 1987; Vitali, 1993).

이러한 평가에 대한 교사의 신념과 관련하여 최근에 교수·학습의 상황에서 교사의 평가 시행을 결정하는 주요 요인에 관한 연구들이 조명을 받으면서 다양한 현장 연구들이 쏟아져 나왔다(Chan, 2007; Edelenbos & Kubanek-German, 2004; Gattullo, 2000; Hsu, 2005; McKay, 2006). 이들 연구에서 교사의 평가 시행을 결정하는 주요 요인들 중 공통적으로 가장 많이 보고된 요인들은 교사의 신념과 평가에 관한 연수(assessment education) 및 자신의 평가 능력에 대한 인식(self-perceived competency)이었다. 그리고 초등학교 교실에서 EFL 교사들의 다양한 평가 시행에 영향을 미치는 요인에 관한 연구(Yang, 2008)에서는 평가에 대한 교사의 자신감(feelings of self-confidence), 평가의 교육적 장점에 대한 교사의 신념(teacher beliefs about the pedagogical benefits of assessment) 및 평가에 관한 연수 등이 교사가 평가의 시행을 위해서 다양한 평가 기법을 폭넓게 사용하는 데 긍정적인 영향을 미치고 있었다. 반면, 교사의 과도한 업무(a heavy workload), 시간적 제약(time constraints) 및 학부모들의 압력(pressure from parents) 등은 교사의 평가 업무나 과제 수행을 저해하는 요인으로 나타났다.

또한, Shih와 Wang(2010)은 영어교사들의 평가 시행에 중대한 영향을 미치는 네 가지 요인으로서 교육적 상황(educational context), 평가 관련 전문 과정이나 연수(professional coursework or training), 시범 수업을 포함한 수업 시연(classroom practice including practice teaching) 및 교육 경력(schooling history)을 제시하였다. 이들 중 교사의 평가 시행에 가장 큰 영향을 미치는 요인은 교육적 상황으로서, 예를 들면 중대한 이해관계가 걸린 시험, 학교 현장의 내부 정책과 관리자의 결정, 평가를 위한 물리적

환경과 시설 및 학부모의 개입 등이 여기에 해당되었다. 또한, Robin(2006)은 교사의 평가 시행에 영향을 미치는 요소로서 여섯 가지의 교육 관련 정보의 출처(sources of information in education)를 언급하였다. 여기에는 교육적 연구(educational research), 평가에 관한 조사(evaluative inquiry), 대규모 평가(large-scale assessment), 교육 정책(educational policy), 전문성 개발(professional development) 및 교사의 신념 등이 있었다. 게다가 Zhang과 Burry-Stock(2003)은 교과 영역(content area), 학년 수준(grade level), 교직 경력(teaching experience) 및 평가 연수(measurement training) 등이 교사의 평가 시행에 영향을 미치는 주요한 요인들이라고 하였다.

교사의 평가 시행을 결정하는 요인들은 학자들마다 아주 다양하게 제시되어 있으나, 그 각각의 용어는 개념적 범주에 있어서 서로 공통적인 부분이 많다. 따라서 선행 연구들에서 조사된 교사의 평가 시행을 결정하는 각 요인들 중에서 말하기 수행평가의 전문성과 효용성 및 시행한계성과 관련이 있는 요인들을 세 가지 범주로 정리하면 다음의 <그림 1>과 같다. 즉, 자신의 평가 능력에 대한 인식과 평가에 대한 교사의 자신감 및 평가에 관한 연수는 말하기 수행평가의 전문성이라는 요인과 관련이 있을 것이다. 그리고 평가의 교육적 장점에 대한 교사의 신념은 말하기 수행평가의 효용성이라는 요인과 관련이 있을 것이다. 또한, 평가 시행을 어렵게 만드는 여러 가지 교육적 상황과 평가를 위한 물리적 환경과 시설은 말하기 수행평가의 시행한계성이라는 요인과 관련이 있을 것이다.



말하기 수행평가 시행

<그림 1> 말하기 수행평가 시행의 결정 요인들

이상과 같이 학생들의 학업 성취를 좌우하는 중대한 역할을 맡고 있는 당사자는 교사이고, 그러한 교사가 평가를 시행하는 데 있어서 가장 큰 영향을 주는 요인들도 역시 교사의 주체적인 사고나 행동과 관련이 많은 것으로 나타났다. 그러므로 학생들의 영어 말하기 능력을 평가하는 데 있어서 무엇보다도 영어교사들의 말하기 수행평가의 시행을 결정하는 주요 요인들, 즉 말하기 수행평가의 전문성과 효용성 및 시행한계성에 대한 영어교사들의 인식이 절대적으로 중요하다는 것을 알 수 있다. 그러나 앞서 말하기 수행평가의 운영 실태에 관한 선행 연구들(김성연, 이호경, 2013; 방인자, 천세영, 2011; 장인순, 2004)에서 밝혀진 대로, 중학교 영어교사들은 다양한 과업 중심의 말하기 수행평가를 시행하고 있지만 그 진정성이 떨어진다는 인식을 하고 있었고, 한편 고등학교에서는 대학입시의 부담으로 인해서 말하기 수행평가

를 제대로 시행하지 못하고 있다는 인식을 지닌 것으로 나타났다. 이와 같이 중학교 영어교사들과 고등학교의 영어교사들은 일반적으로 말하기 수행평가의 시행과 관련한 상황적인 측면에서 서로 다른 인식을 보이는 것으로 추측된다. 따라서 중학교 영어교사들과 고등학교의 영어교사들 집단 간의 말하기 수행평가에 대한 전반적인 인식이 서로 어떻게 다른지를 연구하는 일은 큰 의미가 있을 것으로 사료된다.

지금까지의 선행 연구들을 살펴봤을 때, 우리나라 중·고등학교 영어교사들의 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 인식이 서로 어떠한 관계에 있고, 또한 학교급간과 배경 변인에 따라 어떠한 인식 차이를 보이는지를 다룬 연구는 거의 찾아보기가 힘들다. 특히, 영어교사들의 말하기 수행평가의 전문성과 효용성 및 시행한계성에 대한 인식이 말하기 수행평가를 시행하는 데 있어서 유의미한 영향을 주는지에 관한 연구는 거의 없는 실정이다. 비록 중학교와 고등학교의 말하기 수행평가에 대한 연구들이 많이 이루어지긴 했으나, 주로 말하기 수행평가에 대한 단순한 인식이나 실태 조사, 시행 현황 및 문제점에 관한 사례 연구들이 대부분이었다.

이에 본 연구에서는 우선 영어교사들의 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 인식이 중·고등학교 집단 간에 서로 어떠한 차이가 있고, 여러 배경 변인에 따라 서로 어떻게 다르며, 그리고 서로 어떠한 상관관계에 있는지를 파악하고자 한다. 또한, 앞서 살펴 본 이론적 배경에 근거하여 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 영어교사들의 인식이 그들의 말하기 수행평가를 시행하는 정도에 어떠한 영향을 미치는지를 검증하고자 한다.

III. 예비 연구

본 연구에 앞서 연구자는 연구 도구의 신뢰도와 타당도를 검증하고 연구 설계의 과정에서 예상되는 문제점 등을 미리 파악하기 위해서 2014년 2월경에 예비 연구를 실시하였다. 이는 예비 연구에서 드러난 연구 방법과 연구 결과에 대한 문제점들을 진단하고 보완하여 본 연구의 질적인 완성도를 제고하는 데 그 목적이 있었다. 이번 장에서는 예비 연구의 과제와 연구 방법, 연구 결과와 논의 및 본 연구를 위한 보완점에 대해서 기술하고자 한다.

3.1 예비 연구의 과제

예비 연구에서는 말하기 수행평가의 전문성, 효용성 및 시행한계성이라는 요인에 대한 중·고등학교 영어교사들의 인식은 서로 어떠한 상관관계를 가지고 있고, 또한 이 세 요인들에 대한 영어교사들의 인식은 각각 말하기 수행평가의 시행도에 어떠한 영향을 미치는지를 알아보기 위해서 다음과 같이 두 가지의 연구 과제를 설정하였다.

첫째, 말하기 수행평가를 시행하는 영어교사들의 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 인식은 서로 어떠한 관계가 있는가?

둘째, 말하기 수행평가를 시행하는 영어교사들의 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 인식은 말하기 수행평가의 시행도에 영향을 미치는가?

3.2 예비 연구의 방법

예비 연구에서는 중·고등학교 영어교사들의 말하기 수행평가의 시행에 있어서 그들의 전문성, 효용성 및 시행한계성에 대한 인식이 서로 어떻게 연관되어 있는지, 그리고 말하기 수행평가의 시행에 어떠한 영향을 미치는지를 알아보고자 하였다. 이러한 연구 과제를 수행하기 위하여 현직 중·고등학교 영어교사들을 대상으로 한 설문조사와 수집된 자료에 대한 통계 분석을 실시하였다. 영어교사들이 학교 현장에서 말하기 수행평가를 보다 효과적으로 시행할 수 있는 방향을 모색하고자 진행된 예비 연구의 구체적인 연구 방법은 다음과 같다.

예비 연구의 참여 대상자는 2013년 7~9월에 걸쳐 서울과 경기지역의 중·고등학교 영어교사 42명(중학교 17명과 고등학교 25명)이었다. 그리고 예비 연구 도구인 설문지는 성별, 근무학교, 교직 경력, 주당 수업시수, 말하기 수행평가 관련 연수 이수정도 및 영어 말하기 능력에 대한 교사의 자기평가 등 영어교사의 배경 변인에 관한 6개의 문항들과 함께 말하기 수행평가의 전문성, 효용성, 시행한계성에 대한 인식 및 말하기 수행평가의 시행도를 묻는 28개의 문항들로 구성되었다. 예비 연구를 위해 사용되었던 설문지의 세부 문항들에 대한 내용은 <표 1>에 제시하였다.

예비 연구의 과제에 따른 자료 분석의 방법은 통계 분석 프로그램인 SPSS(Statistical Package for Social Science) 15.0 프로그램을 사용하여 요인 분석과 신뢰도 분석 및 상관관계 분석을 하였다. 또한, AMOS 7.0 프로그램을 사용하여 구조방정식모형을 설정한 후 각 잠재 변인 간의 상관관계 및 인과 관계를 검증하였다.

<표 1> 설문지의 세부 문항 구성(예비 연구)

주요 요인	문항 수	문항별 내용
말하기 수행평가 시행도	2	·학기당 말하기 수행평가의 시행 횟수 ·학기당 말하기 수행평가의 시행 방법 수
말하기 수행평가 전문성	14	·말하기 수행평가의 목적과 내용에 적합한 평가 방법 선택 ·학생의 특성에 적합한 말하기 수행평가 방법의 선택 ·학생의 성취 수준에 적합한 말하기 수행평가 방법의 선택 ·적합한 말하기 수행평가 도구의 부재시 평가 도구의 직접 개발 ·말하기 수행평가의 과제 및 채점 기준의 개발 원리에 맞는 평가 도구의 개발 ·말하기 수행평가 계획의 사전 공지 ·평가 방법 및 평가 도구의 정확한 숙지 후 말하기 수행평가의 시행 ·말하기 수행평가 시행 전 과제 및 채점 기준에 대한 충분한 사전 설명 ·말하기 수행평가의 채점 기준을 정확히 숙지한 후 채점 ·말하기 수행평가의 각 영역별 검사 점수의 이해와 점수 부여 ·말하기 수행평가의 목적 및 내용에 기초한 평가 결과의 해석 ·말하기 수행평가의 결과 해석 시 다른 여러 변수의 고려 ·말하기 수행평가의 결과에 기초한 학생의 특징 파악 ·말하기 수행평가의 종합적 결과의 교수·학습 자료로의 활용
말하기 수행평가 효용성	5	·학생의 말하기 능력 향상에의 도움 ·학생의 말하기 능력 발달 과정 파악에의 도움 ·학생의 말하기 성취 수준 파악에의 도움 ·학생의 말하기에 대한 학습 동기 향상에의 도움 ·교사의 말하기 수업 방법 개선에의 도움
말하기 수행평가 시행한계성	7	·학급 규모의 크기와 학생 수의 과다 ·채점의 객관성 확보의 어려움 ·적절한 평가 도구의 부재 ·학생의 학습 부담의 가중 ·학교 업무로 인한 시간의 부족 ·교사 연수나 자료의 부족 ·사교육비의 증가에 대한 우려
총 문항 수	28	

예비 연구에서는 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 인식과 말하기 수행평가의 시행도를 묻는 총 28개 문항 전체에 대해서 요인

분석을 하였고, 그 결과는 <표 2>와 같다. 표에서 보듯이, 전체 문항들 중에서 말하기 수행평가의 효용성과 시행한계성의 일부 문항들은 각각 하나의 요인으로 묶이지 않고 서로 다른 요인을 설명하고 있는 것으로 나타났다. 이는 아마도 예비 연구의 설문 도구가 다듬어지지 않은 초안인데다가, 특히 연구 대상자들의 표본 크기가 너무 작았기 때문인 것으로 추측된다.

<표 2> 주요 요인들에 대한 요인 분석 결과(예비 연구)

세부 문항	요인1	요인2	요인3	요인4	요인5	요인6	요인7	요인8
채점성적5	.869	-.057	.144	-.167	.079	.047	.008	-.014
선정개발4	.849	-.110	.147	-.137	.090	.019	-.198	.070
선정개발5	.846	-.298	.098	-.060	.013	.088	.011	.091
채점성적4	.829	-.045	.130	-.032	.158	.268	.130	.090
선정개발1	.818	-.032	.146	.153	.132	.237	.045	-.054
선정개발2	.791	-.047	-.051	.058	-.093	.368	.018	.039
채점성적2	.765	-.244	.180	-.017	.041	-.089	.214	-.259
선정개발3	.764	.062	.183	.186	-.361	.051	.009	-.147
채점성적3	.759	-.076	.111	-.072	.263	.144	.316	-.182
채점성적1	.710	.022	.039	-.134	.254	.307	.367	-.163
해석활용1	.700	.017	.105	-.092	.249	.031	.326	-.050
해석활용4	.661	.291	.349	.003	-.212	-.078	-.144	.213
해석활용2	.631	.089	.411	-.142	-.013	-.071	-.381	-.057
해석활용3	.516	-.044	.192	.302	-.374	-.273	.485	-.156
효용성3	.512	-.062	.461	-.032	-.071	.499	.093	-.108
한계성1	-.135	.898	-.086	-.105	.083	-.031	-.005	-.080
한계성6	-.010	.824	-.087	.045	.108	-.003	-.139	.052
효용성1	.243	-.150	.863	-.016	-.059	.198	.167	-.052
효용성2	.474	-.279	.586	-.086	-.015	.153	.222	.067
시행방법	.006	-.127	-.197	.917	.122	.062	-.068	.003
시행횟수	-.193	.059	.114	.912	.034	-.053	-.057	.088
한계성5	.116	.086	-.147	.128	.822	.107	.107	-.094
한계성7	.310	.268	.199	.110	.708	-.258	-.224	.264
효용성5	.348	.071	.310	-.014	.062	.774	.128	.063
한계성3	-.255	.371	.027	-.097	.050	-.532	-.041	.531
효용성4	.093	-.180	.206	-.172	.019	.157	.822	.113
한계성4	.158	-.258	.072	.149	.046	-.025	.141	.797
한계성2	-.345	.415	-.245	-.052	-.065	.097	-.104	.658

<표 3>은 예비 연구에서 사용된 연구 도구의 타당성을 검증하기 위해서 실시한 요인 분석이 적절한지를 나타낸 결과이다. 표본의 적절성을 측정하는 KMO 값이 .555로서 상관행렬이 요인 분석을 위한 적합한 기준값인 1에 다소 미치지 못하는 것으로 나타났다.

<표 3> 연구 도구의 KMO와 Bartlett의 구형성 검정 결과(예비 연구)

Kaiser-Meyer-Olkin	Bartlett의 구형성 검정		
	근사 카이제곱	df	p
.555	955.380	378	.000

예비 연구에서 사용된 연구 도구의 신뢰도를 분석한 결과는 <표 4>에 제시하였다. 말하기 수행평가의 전문성에 대한 인식을 측정하는 문항들의 신뢰도 계수는 .949, 말하기 수행평가의 효용성에 대한 인식을 측정하는 문항들의 신뢰도 계수는 .820, 말하기 수행평가의 시행한계성에 대한 인식을 측정하는 문항들의 신뢰도 계수는 .639, 그리고 말하기 수행평가의 시행도에 대한 인식을 측정하는 문항들의 신뢰도 계수는 .868로 나타났다. 따라서 예비 연구에서 사용된 설문지 전체 문항들에 대한 신뢰도 계수는 .841로서 신뢰할 만한 수준임을 알 수 있었다.

<표 4> 연구 도구의 신뢰도(예비 연구)

주요 요인	문항 수	Cronbach α
전문성	14	.949
효용성	5	.820
시행한계성	7	.639
시행도	2	.868
전 체	28	.841

3.3 예비 연구의 분석 결과 및 논의

예비 연구의 첫 번째 과제는 중·고등학교 영어교사들이 인식하는 말하기 수행평가의 전문성, 효용성 및 시행한계성이라는 세 요인들이 서로 관련성이 있는지를 파악하는 것이었다. 그래서 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대해서 영어교사들이 지닌 인식 간의 상관관계를 분석하였고, 그 결과는 <표 5>와 같다.

<표 5> 세 주요 요인들 간의 상관관계(예비 연구)

요인	전문성	효용성
전문성		
효용성	.639 (.000)	
시행한계성	-.170 (.282)	-.287 (.065)

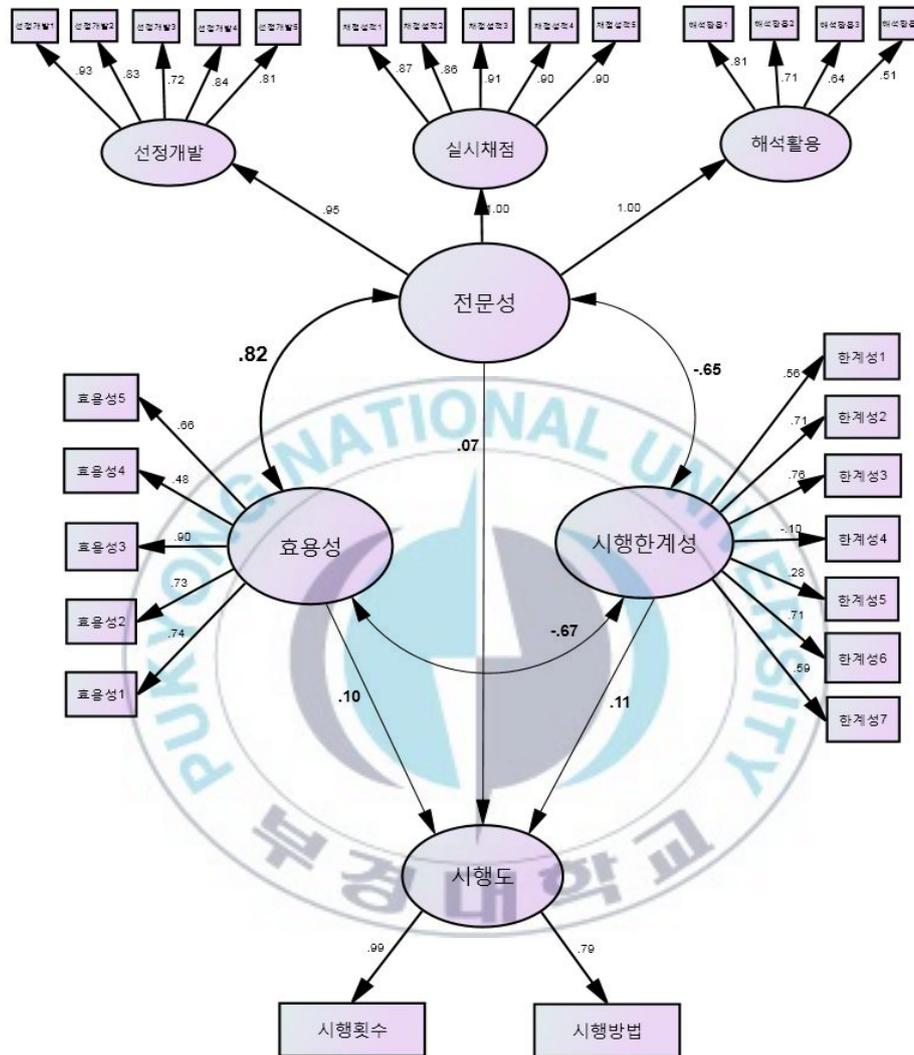
<표 5>에서 보듯이, 말하기 수행평가의 효용성은 말하기 수행평가의 전문성과 서로 높은 정의 상관관계($r=.639, p<.05$)를 가진 것으로 나타났다. 즉, 말하기 수행평가의 효용성에 대한 긍정적인 인식이 높은 영어교사들일수록 말하기 수행평가의 전문성에 대한 인식이 높으며, 또한 그 역의 관계도 성립할 수 있음을 보여준다. 한편, 말하기 수행평가의 시행한계성은 말하기 수행평가의 전문성($r=-.170, p>.05$)뿐만 아니라 말하기 수행평가의 효용성($r=-.287, p>.05$)과도 서로 음의 상관관계가 있는 것으로 나타났지만 통계적으로 유의미하지는 않았다.

예비 연구의 두 번째 과제를 위해서 <그림 2>와 같이 말하기 수행평가의 전문성, 효용성 및 시행한계성이라는 잠재 변인들이 말하기 수행평가의 시행도라는 잠재 변인에 미치는 경로계수의 추정치를 보여주는 구조방정식모형을 설정하였다. 그림에서 보듯이, 말하기 수행평가의 전문성과 효용성은 서로 높은 상관관계를 가지면서 통계적으로 유의미한 것으로 나타났으나, 말하기 수행평가의 전문성, 효용성 및 시행한계성은 말하기 수행평가의 시행도에 직접적인 영향을 미치는 것처럼 보이지만 통계적으로 유의미하지 않은 것으로 나타났다. 구조방정식모형에 나타난 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 영어교사들의 인식이 말하기 수행평가의 시행도에 미치는 직접적인 영향에 대한 경로계수는 <표 6>에 제시하였다.

<표 6> 잠재 변인들 간의 경로분석 결과(예비 연구)

경로 및 방향	비표준화 계수	표준화 계수	유의확률
전문성 → 시행도	.08	.07	.153
효용성 → 시행도	.10	.10	.209
시행한계성 → 시행도	.10	.11	.218

이상과 같이 예비 연구의 결과를 정리해 보면, 말하기 수행평가의 효용성에 대한 긍정적인 인식을 가진 영어교사들은 말하기 수행평가의 전문성에 대한 인식도 아주 높게 지니고 있었고, 또한 말하기 수행평가의 전문성에 대한 인식이 높은 영어교사들은 말하기 수행평가의 효용성에 대해서도 긍정적으로 인식하고 있음을 알 수 있었다. 게다가 예비 연구에서는 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 중·고등학교 영어교사들의 인식이 각각 말하기 수행평가를 시행하는 데 있어서 직접적인 영향을 주지 않는 것으로 나타났다.



<그림 2> 구조방정식모형(예비 연구)¹⁾

1) 각 경로에서 추정된 미지수는 표준화된 경로계수임($\chi^2=688.5$, $df=367$, $N=42$): TLI=.614, CFI=.625, RMSEA=.149

3.4 예비 연구의 보완점

예비 연구의 목적은 본 연구를 진행하기에 앞서 연구 도구와 연구 방법 및 연구 결과에 대한 큰 윤곽을 잡기 위한 것이었다. 이에 연구자는 다음과 같이 예비 연구를 통해서 밝혀진 네 가지 문제점들을 개선하고 난 후에 본 연구를 진행하였다.

첫째, 예비 연구의 가장 큰 문제점은 무엇보다도 표본의 크기가 작다는 것이었다. 연구 표본의 크기는 요인 분석과 상관관계 분석의 결과뿐만 아니라 최소 150명 이상의 표본의 자료가 필요한 구조방정식모형의 분석 결과에도 지대한 영향을 미친다. 보통 50개 이하의 표본 크기는 매우 열악(very poor), 100개는 열악(poor), 200개는 보통(fair), 그리고 300개 이상의 표본 크기는 우수한(good) 것으로 간주된다(Comrey & Lee, 1992). 특히, 예비 연구에서는 말하기 수행평가의 전문성, 효용성, 시행한계성 및 시행도에 대한 구조방정식모형의 적합도(TLI=.614, CFI=.625, RMSEA=.149)가 그다지 만족스럽지 못하여 잠재 변인들의 인과 관계를 검증할 수 있는 구조방정식모형을 적용하는데 어려움이 있었다. 따라서 본 연구에서는 타당한 요인 분석과 높은 적합도를 지닌 구조방정식모형의 설정을 위해서 최소 200명 이상의 영어 교사들을 전국적으로 확보하여 표본의 대표성뿐만 아니라 연구 결과의 신뢰성과 타당성을 높이고자 한다.

둘째, 예비 연구의 도구인 설문지의 문항들 중 묻고자 하는 내용이 중복되거나 다소 모호한 일부 문항들이 있었다. 이는 검사 도구의 특성에서 기인한 것으로 연구 도구의 타당성을 떨어뜨리는 주요한 원인들 중 하나이다. 그러므로 본 연구에서는 설문지의 모든 문항들에 대해서 동료 영어교사들이나 영어교육을 전공한 교수들의 의견을 반영하여 연구자가 신중하게 재점검할 필요가 있다. 즉, 내용이 중복되는 문항들은 하나로 합치거나 부분적인

수정을 가하든지, 혹은 아예 제외시켜서 본 연구 도구의 타당성을 높여야 할 것이다.

셋째, 예비 연구에서는 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대해서 전체 중·고등학교 영어교사들이 지닌 인식 간의 단순한 상관관계에 대한 결과만을 얻을 수 있었다. 그런데 앞서 많은 선행 연구들에서도 밝혔듯이, 말하기 수행평가에 대한 영어교사들의 인식과 말하기 수행평가의 운영 실태는 중학교와 고등학교라는 학교급에 따라 서로 많은 차이가 있었다(김성연, 이호경, 2013; 방인자, 천세영, 2011; 장인순, 2004). 또한, 교사 개인을 둘러싼 많은 다양한 개인적 배경과 자질 및 경험적 인식에 따라서도 교사의 말하기 수행평가 시행의 정도나 양상에는 차이가 있다(Chan, 2007; Green & Stager, 1987; Shih & Wang, 2010; Vitali, 1993; Yang, 2008; Zhang & Burry-Stock, 2003). 따라서 본 연구에서는 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 영어교사들의 인식과 관련하여 학교급 간과 배경 변인에 따른 비교 연구를 연구 과제로 설정하여 연구의 범위를 보다 확대하고 세분화하고자 한다.

넷째, 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 영어교사들의 인식은 정적이든 부적이든 서로 연관성을 가지면서 말하기 수행평가를 직접 시행하는 영어교사들의 주체적인 결정과 행동에 영향을 미칠 수 있다. 이는 영어교사들의 말하기 수행평가의 전문성, 효용성 및 시행한계성이라는 변인들에 대한 인식이 서로 간에 어떠한 영향을 주고받는지, 또한 말하기 수행평가를 시행하는 정도에는 얼마나 영향을 미치는지에 대한 연구의 동기를 제공한다. 그러므로 본 연구에서는 구조방정식모형을 통해서 말하기 수행평가의 전문성, 효용성과 시행한계성 및 말하기 수행평가의 시행도에 대한 영어교사들의 인식 간의 직·간접적인 인과 관계를 구체적으로 확인하고자 한다.

IV. 연구 방법

본 연구는 2014년 2월에 실시한 예비연구에서 밝혀진 결과를 기초로 연구 과제를 보다 구체화하고 연구 대상을 확대하였다. 이 장에서는 연구 대상, 연구 도구, 그리고 자료의 수집 절차 및 분석 방법 등에 대해서 구체적으로 기술하고자 한다.

4.1 연구 대상

본 연구의 자료는 설문조사를 위주로 얻었고, 서울, 부산, 경기도, 경상도 및 전라도 등 5개 지역에 근무하는 중·고등학교 영어교사들이 자발적으로 설문조사에 참여하였다. 설문조사에 참여한 영어교사들은 공·사립 일반계 중학교와 고등학교에 근무하는 정교사와 기간제교사로서 학생들을 대상으로 말하기 수행평가를 직접 시행하고 있는 영어교사들이었다. 각 지역별로 설문에 참여한 구체적인 학교 수와 참여자 수는 <표 7>과 같다.

<표 7> 지역별 설문 참여 학교 및 참여자 수

구분	지역별 참여 학교 수						회수된 설문지 수	유효한 설문지 수	제외된 설문지 수	유효 참여율(%)
	서울	부산	경기	경상	전라	소계				
중학교	2	25	3	6	0	36	131	128	3	97.7
고등학교	9	11	4	4	2	30	120	115	5	95.8
총계				66			251	243	8	96.8

영어교사들의 말하기 수행평가의 전문성, 효용성 및 시행한계성에 관한 인식을 조사하기 위한 본 연구의 설문조사에 중학교 36개교와 고등학교 30개교의 영어교사들이 참여하였다. <표 7>에서 보듯이, 총 251명의 설문 응답자들 중 중학교 영어교사는 131명이고 고등학교 영어교사는 120명으로서 학교별 참여자의 수는 크게 차이가 나지 않았다. 이들 중 모든 설문지 문항에 성실하게 응답한 대상자들은 중학교 영어교사 128명(97.7%)과 고등학교 영어교사 115명(95.8%)으로서, 영어교사 총 243명(96.8%)의 설문 자료를 토대로 본 연구의 통계 분석을 실시하였다. 회수된 설문지들 중 설문 문항에 대해서 불성실하거나 부분적으로 응답이 된 설문지 8부는 연구 사례에서 제외하였다.

폐쇄된 설문에 대한 응답 이외에도 말하기 수행평가의 시행과 관련하여 설문 참여자들의 자유로운 의견을 조사하였다. <표 8>에서 보듯이, 설문지의 자유서술식 문항에 본인의 의견을 피력한 대상자들은 총 85명(35.1%)의 영어교사들로서, 중학교 영어교사 46명(35.9%)과 고등학교 영어교사 39명(33.9%)이었다.

<표 8> 자유서술식 문항 응답자 수

구분	유효한 설문지 수	자유서술식 문항 응답자 수	참여율(%)
중학교	128	46	35.9
고등학교	115	39	33.9
총계	243	85	35.1

본 연구에서는 영어교사들의 몇 가지 중요한 개인적 배경 변인들을 고려하였다. 본 연구에 조사된 배경 변인들은 성별, 학교급, 교직 경력, 말하기 수행평가 관련 연수 이수 및 영어 말하기 능력이었으며, 각 배경 변인에 따른 분포는 <표 9>에 제시하였다.

<표 9> 설문 참여자들의 배경 변인별 분포

배경 변인	구분	N	비율(%)
성별	남	57	23.5
	여	186	76.5
학교급	중학교	128	52.7
	고등학교	115	47.3
교직 경력	5년 이하	69	28.4
	6~10년	39	16.0
	11~15년	43	17.7
	16~20년	18	7.4
	21년 이상	74	30.5
말하기 수행평가 관련 연수 이수	전혀 없음	97	39.9
	1회	81	33.3
	2회 이상	65	26.8
영어 말하기 능력	부족함	21	8.7
	보통임	141	58.0
	우수함	81	33.3

<표 9>에서 보듯이, 설문 참여자들 중 남자는 23.5%이고 여자는 76.5%로 상대적으로 여자의 비율이 훨씬 높았다. 근무 학교급별 분포에서는 중학교 영어교사의 비율이 52.7%이고 고등학교 영어교사의 비율은 47.3%로, 상대적으로 중학교에 근무하는 영어교사들의 수가 약간 많았다. 교직 경력의 경우 21년 이상의 교직 경력을 지닌 영어교사의 비율이 30.5%로 가장 높았고 5년 이하의 교직 경력을 가진 영어교사의 비율이 28.4%로 두 번째로 높았으며, 그 다음으로 11~15년, 6~10년 및 16~20년의 교직 경력 순으로 그 비율이 나타났다. 말하기 수행평가 관련 연수 이수의 경우 말하기 수행평가와 관련된 연수를 전혀 이수하지 않는 영어교사의 비율이 39.9%로 가장 높았고, 그 다음으로 1회의 비율인 33.3%와 2회 이상의 비율인 26.8%의 순으로 나타났다. 설문 응답자들의 영어 말하기 능력에 따른 분포율을 보면 본인의

말하기 능력이 보통이라고 생각하는 영어교사들의 비율이 58.0%로 가장 높았고, 본인의 말하기 능력이 우수하다고 생각하는 영어교사들의 비율은 33.3%이며, 본인의 말하기 능력이 부족하다고 생각하는 영어교사들의 비율은 8.7%로 가장 적었다.

이상으로 설문 참여자들의 배경 변인을 살펴본 결과, 그 특성이 대체로 고르게 분포되어 있어서 표집이 그다지 편향되지 않음을 알 수 있다. 따라서 차후 본 설문 자료를 이용한 통계 분석을 하는 데 있어서 별다른 무리가 없는 것으로 확인되었다.

4.2 연구 도구

본 연구에 사용된 도구는 중·고등학교 영어교사들의 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 인식을 묻는 설문지(부록 참조)이다. 지난 예비 연구의 결과를 참조하여 설문지의 일부를 수정 및 보완하였고, 측정 도구로서의 신뢰도 계수는 Cronbach's Alpha 계수를 채택하였다.

4.2.1 설문지

말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 중·고등학교 영어교사들의 인식을 조사하기 위하여 직접 면담을 통한 방법도 있지만, 본 연구에서는 시간의 제약과 자료의 정확한 분석을 위해서 설문지를 활용하였다. 본 연구에 사용된 설문지는 선행 연구들에서 다루어진 말하기 수행평가의 전문성, 효용성 및 시행한계성에 관한 내용을 바탕으로 연구자가 설문 문항

을 직접 제작하였다. 설문지 제작을 위한 선행 연구들로서 말하기 수행평가의 전문성은 김수동 외(2005), 말하기 수행평가의 효용성은 이문복 외(2012)와 이종윤(2002), 그리고 말하기 수행평가의 시행한계성은 이소영(2008)을 주로 참고하였다.

설문지의 첫 페이지는 설문조사의 취지를 설명하고 협조를 당부하는 내용과 함께 성별, 근무 학교, 교직 경력, 말하기 수행평가 관련 연수 이수정도 및 영어 말하기 능력에 대한 자기평가 등 설문 응답자들의 배경 정보를 묻는 문항들로 구성하였다. 그리고 두 번째와 세 번째 페이지는 응답자의 말하기 수행평가의 시행도를 묻는 2문항, 말하기 수행평가의 전문성에 대한 인식을 묻는 12문항, 말하기 수행평가의 효용성에 대한 인식을 묻는 5문항, 그리고 말하기 수행평가의 시행한계성에 대한 인식을 묻는 5문항 등 총 24문항으로 구성하였다. 이 설문 문항들은 모두 Likert의 5단계 척도로서, '전혀 그렇지 않다=1', '그렇지 않다=2', '보통이다=3', '그렇다=4', 그리고 '매우 그렇다=5'로 구성되어 설문 응답자는 각 문항을 읽고 이들 중 하나를 선택하도록 하였다. 또한, 네 번째 페이지는 영어 말하기 수행평가의 시행과 관련하여 설문 응답자의 자유로운 의견을 진술할 수 있도록 하였다. 이와 같은 설문지의 전문은 [부록]으로 제시하였다.

일차적으로 완성된 연구 도구의 타당성과 완성도의 제고를 위해서 연구자가 근무하는 학교 및 인근 학교의 영어교사 5명과 영어교육학을 전공한 대학교수의 검토 의견을 얻어 불명료한 질문이나 중복된 내용을 일부 수정한 후 예비 조사를 거쳐 최종적인 연구 도구를 완성하였다. 완성된 연구 도구의 각 주요 요인별 설문 문항들의 세부적인 내용과 문항 수는 <표 10>과 같다.

<표 10> 주요 요인별 세부 문항 구성

주요 요인	문항 수	문항별 내용
말하기 수행평가 전문성	평가 방법과 도구의 선정 및 개발	4 <ul style="list-style-type: none"> · 평가 목적과 내용에 적합한 방법과 도구 선택 · 학생의 성취 수준에 적합한 평가 방법과 도구 선택 · 적합한 평가 방법이나 도구의 부재 시 직접 개발 · 과제 및 채점 기준과 원리에 맞게 직접 개발
	평가의 실시 및 채점	4 <ul style="list-style-type: none"> · 평가 실시 전 학생들에게 평가 계획 공지 · 평가 실시 전 평가 방법과 도구의 숙지 · 평가 실시 전 학생들에게 과제 및 채점 기준 설명 · 채점 기준 숙지 후 채점 실시
	평가 결과의 해석 및 활용	4 <ul style="list-style-type: none"> · 평가 목적과 내용에 기초한 평가 결과 해석 · 평가 결과 해석 시 여러 변수 등의 고려 · 평가 결과에 근거한 학생의 성취적 특성 파악 · 평가 결과의 종합과 교수·학습 자료로의 활용
말하기 수행평가 효용성		5 <ul style="list-style-type: none"> · 학생의 말하기 능력 향상에 기여 · 학생의 말하기 능력 발달 과정 파악에 기여 · 학생의 말하기 성취 수준 파악에 기여 · 학생의 말하기에 대한 학습 동기 향상에 기여 · 교사의 말하기 수업 방법 개선에 기여
		5 <ul style="list-style-type: none"> · 학급 규모의 크기와 학생 수의 과다 · 채점의 객관성 확보의 어려움 · 적절한 평가 방법과 도구의 부재 · 교사 연수나 자료의 부족 · 학생의 학습 부담 가중
말하기 수행평가 시행도	2	<ul style="list-style-type: none"> · 말하기 수행평가의 학기당 시행 횟수 · 말하기 수행평가의 학기당 시행 방법 수
총 문항 수	24	

4.2.2 신뢰도 및 타당도

먼저 연구 도구의 신뢰도를 검증하기 위해서 말하기 수행평가의 전문성, 효용성 및 시행한계성과 말하기 수행평가의 시행도라는 네 주요 요인들에 대한

문항내적 일관성 신뢰도 계수를 추정한 결과는 <표 11>과 같다. 즉, 영어교사들의 말하기 수행평가의 전문성에 대한 인식을 측정하는 문항들의 신뢰도 계수는 .899, 말하기 수행평가의 효용성에 대한 인식을 측정하는 문항들의 신뢰도 계수는 .863, 말하기 수행평가의 시행한계성에 대한 인식을 측정하는 문항들의 신뢰도 계수는 .828, 그리고 말하기 수행평가의 시행도를 측정하는 문항들의 신뢰도 계수는 .436으로 나타났다. 한편, 말하기 수행평가의 시행도에 대한 신뢰도 계수가 다소 낮게 나온 이유는 아마도 시행도라는 요인을 측정하는 문항이 고작 두 개에 불과하기 때문인 것으로 추측된다. 결과적으로, 전체 24개 문항들에 대한 신뢰도 계수는 .841로 설문지의 전반적인 내적 신뢰도가 비교적 높은 편으로 연구 도구로서 신뢰할 만한 수준이다. 따라서 본 연구 도구는 대체적으로 네 가지 요인을 일관되게 평가하는 것으로 판단할 수 있다.

<표 11> 연구 도구의 신뢰도

주요 요인	문항 수	Cronbach α
전문성	12	.899
효용성	5	.863
시행한계성	5	.828
시행도	2	.436
전체	24	.841

아울러 영어교사들의 말하기 수행평가의 시행과 관련한 심리적 특성이나 행동 양상을 측정하기 위해서 본 연구 도구가 얼마나 타당성을 지니고 있는지를 확인하는 방법으로서 연구 대상자들의 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 인식과 말하기 수행평가의 시행도를 묻는 총 24개 문항 전체에 대해서 요인 분석(factor analysis)을 실시하였다. 요인 분석은 많은 측정 변수들을 공통적인 요인들끼리 묶어 자료의 복잡성을 줄이고 측

정된 변수들이 동일한 구성 개념을 측정하고 있는지를 파악하기 위한 방법으로서, 검사나 측정 척도의 개발 과정에서 측정 도구의 타당성을 파악하기 위해서 많이 사용한다(성태제, 2007). <표 12>에서 보듯이, 표본의 적절성을 측정하는 KMO 값이 .873으로서 1에 가깝고, 또한 요인들 간의 상관이 0인지를 검정하는 Bartlett의 구형성 검정 통계치가 3082.430(df=276, p=.000)으로서 유의수준 .05에서 유의하므로 상관행렬이 요인 분석을 실시하기에 적합한 것으로 나타났다.

<표 12> KMO와 Bartlett의 구형성 검정 결과

Kaiser-Meyer-Olkin	Bartlett의 구형성 검정		
	근사 카이제곱	df	p
.873	3082.430	276	.000

이를 바탕으로 주성분 분석(principal component analysis)과 요인 적재치의 단순화를 위해서 Kaiser 정규화가 있는 직교회전(varimax rotation) 방식을 통해서 요인 추출을 한 결과, <표 13>과 같이 총 여섯 개의 뚜렷한 요인이 도출되었다. 즉, 말하기 수행평가의 ‘전문성’이라는 요인을 구성하는 세 하위 요인들인 평가 방법과 도구의 선정 및 개발에 관한 4문항(요인 적재치: .846~.533), 평가의 실시 및 채점에 관한 4문항(요인 적재치: .852~.779), 그리고 평가 결과의 해석 및 활용에 관한 4문항(요인 적재치: .804~.725)과 더불어, 말하기 수행평가의 ‘효용성’이라는 요인을 설명하는 5문항(요인 적재치: .813~.710), 말하기 수행평가의 ‘시행한계성’이라는 요인을 설명하는 5문항(요인 적재치: .840~.697), 또한 말하기 수행평가의 ‘시행도’라는 요인을 설명하는 2문항(요인 적재치: .853~.732)은 각각 동일한 구성 개념을 측정하고 있는 것으로 나타났다.

<표 13> 주요 요인들에 대한 요인 분석 결과

세부 문항	요인1	요인2	요인3	요인4	요인5	요인6
	전문성			효용성	시행 한계성	시행도
	개발 및 선정	실시 및 채점	해석 및 활용			
개발3	.846	.113	.202	.145	-.090	.078
개발4	.825	.132	.169	.256	-.055	.070
개발2	.619	.238	.278	.301	-.182	-.017
개발1	.533	.180	.349	.408	-.182	-.018
채점3	.011	.852	.058	.143	-.005	-.067
채점1	.104	.845	.151	.044	-.052	-.024
채점2	.226	.790	.277	.059	-.072	.016
채점4	.167	.779	.210	.100	-.070	.035
활용2	.117	.194	.804	.217	-.003	.106
활용3	.249	.201	.753	.323	.049	.056
활용1	.211	.267	.745	.198	-.081	.017
활용4	.232	.122	.725	.340	-.004	.015
효용1	.155	.108	.183	.813	-.052	-.013
효용2	.181	.138	.137	.781	-.117	.004
효용4	.068	-.008	.193	.757	-.035	.001
효용3	.320	.185	.175	.724	-.020	-.023
효용5	.092	-.003	.201	.710	-.098	.067
한계3	-.104	-.113	.047	-.100	.840	.048
한계2	-.068	-.057	.024	-.140	.786	-.033
한계4	-.109	-.172	.040	-.014	.758	.055
한계5	-.035	-.059	-.107	-.040	.731	.036
한계1	.005	.238	-.063	-.009	.697	-.054
시행횟수	-.062	.040	-.052	.020	.045	.853
시행방법	.166	-.082	.181	.005	-.002	.732

<표 13>에서 보듯이, 개별 문항들이 해당 요인 각각에 비중을 얼마나 차지하고 있는지를 나타내는 값인 요인 적재치(factor loading)가 모두 0.5이상으로 나와 각 해당 요인으로 묶이는 데 큰 무리가 없음을 보여준다. 따라서 본 연구의 설문지는 측정 도구로서의 타당성이 있는 것으로 확인되었다.

4.3 자료의 수집 절차 및 분석 방법

본 연구의 자료 수집은 예비 연구의 결과를 기초로 보다 많은 연구 참여자를 확보한 양질의 연구를 위해서 2014년 4월 초순에서 5월 중·하순에 걸쳐 약 7주 동안 이루어졌다. 설문지를 배포하기에 앞서 서울, 부산, 경기도, 경상도 및 전라도 등 5개 지역에 근무하는 중·고등학교 영어교사들에게 직접 전화를 하거나 개인 이메일을 통해서 말하기 수행평가에 관한 설문조사에 참여해 줄 것을 부탁하였다. 설문지는 설문지 원본을 복사하여 직접 전달하거나, 혹은 개인 이메일을 통해서 설문지 파일을 첨부하여 보내주고 회수할 수 있도록 하였다. 전화나 개인 이메일로 연락을 할 때 본 설문조사는 다른 용도나 목적으로는 절대로 사용되지 않고 순수하게 학위 논문을 위한 연구 자료로만 사용됨을 분명히 밝히었다. 그러므로 설문 문항에 응답 시 영어 말하기 수행평가에 대한 솔직한 생각이나 느낌을 성의껏 표시해 줄 것을 당부하였다. 응답이 완료된 설문지는 주로 우편을 통하거나, 혹은 스캔 파일이나 사진 파일의 형태로 개인 이메일이나 휴대전화를 통해서 회수되었다.

이렇게 수집된 설문지들 중에서 불성실하게 응답된 8개의 사례들을 제외한 총 243개의 사례들을 통계처리 프로그램인 SPSS(Statistical Package for Social Science) 18.0 프로그램을 이용하여 분석하였고, 또한 AMOS (Analysis of Moment Structure) 20.0 프로그램을 이용하여 주요 요인들 간의 직·간접적인 인과 관계를 검증하였다. 또한, 연구자는 주로 설문지를 이용한 양적 연구 방법을 중심으로 설문 결과를 분석하였는데, 양적 연구 결과의 신뢰성을 높이기 위한 질적 연구 자료로서 연구 참여자들이 자유서술식으로 응답한 내용을 정리하고 범주화하여 본 연구의 결과를 보완하였다.

연구자가 제시한 각 연구 과제에 따른 통계 분석 방법은 다음과 같다. 첫째, 중학교 영어교사들과 고등학교 영어교사들이 인식하는 말하기 수행평가

의 전문성, 전문성의 세 하위 요인들, 효용성 및 시행한계성이 중·고등학교 집단 간에 어떠한 차이가 있는지를 파악하기 위해서 두 독립표본 t 검정을 실시하였다. 이 때 요인 분석을 통해서 추출된 주요 요인별 하위 문항들의 총점을 합쳐 문항수로 나눈 평균값을 각각의 요인 지수로 사용하였다. 분석된 모든 결과들은 기술통계량과 함께 t값 및 통계적 유의확률을 도표로 제시하였다. 그리고 여섯 개의 각 요인에 해당하는 문항들에 대해서 중학교와 고등학교 영어교사들의 인식 차이가 구체적으로 어떻게 다른지를 살펴보기 위해서도 마찬가지로 두 독립표본 t 검정을 실시하였다. 이 때 각 문항 간의 유의미한 차이를 알아보기로 집단별 모평균에 대한 95% 신뢰구간을 추정하여 비교하였다. 또한, 설문지에서 조사된 말하기 수행평가에 대한 영어교사들의 자유서술식 의견들 중 말하기 수행평가의 시행한계성과 개선 방안에 관한 의견들을 모아 범주화하였다.

둘째, 중·고등학교 영어교사들의 말하기 수행평가의 전문성, 전문성의 세 하위 요인들, 효용성 및 시행한계성 등에 대한 인식이 교직 경력, 관련 연수 이수 및 말하기 능력이라는 배경 변인에 따라 유의미한 차이가 있는지를 파악하고자 일원배치분산분석(ANOVA)을 실시하였다. 그리고 각 요인별로 배경 변인에 따른 유의미한 차이가 있을 경우 그 차이가 과연 어느 집단 간에 나타나는지를 알아보기로 Scheffe 사후검정을 실시하였다. 이 때도 요인별 평균값을 각각의 요인 지수로 사용하였는데, 영어교사들의 배경 변인을 독립변수로, 그리고 각 요인 지수를 종속변수로 두고 분석하였다. 아울러 기술통계량과 함께 F값, 통계적 유의확률 및 Scheffe 사후검정의 결과를 배경 변인에 따라 도표로 제시하였다.

셋째, 중·고등학교 영어교사들이 인식하는 말하기 수행평가의 전문성, 효용성 및 시행한계성이라는 요인들 간에 상관관계의 존재 여부와 상관성의 정도를 파악하고자 상관분석을 실시하였다. 이 때 말하기 수행평가의 각 요인

별 평균값인 요인 지수들끼리의 상관관계를 중학교와 고등학교에 따라 어떻게 다른지도 분석하였다.

마지막으로, 중·고등학교 영어교사들이 인식하는 말하기 수행평가의 전문성, 효용성 및 시행한계성이라는 세 요인들이 말하기 수행평가의 시행도라는 요인에 어떠한 영향을 미치는지를 확인하고자 구조방정식모형(Structural Equation Modeling: SEM)을 사용하였다. 말하기 수행평가와 관련된 문헌들과 선행 연구들 및 연구자의 경험적 근거를 토대로 구조방정식모형을 설정하고 그 모형의 정규성과 적합도를 검증한 후, 각 잠재 변인 간의 인과 관계 및 직·간접 효과를 분석하였다. 그리고 본 연구에서 모든 통계검증의 유의수준은 .05로 미리 정해두었다.



V. 연구 결과 및 논의

본 장에서는 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 중·고등학교 영어교사들의 학교급간과 배경 변인에 따른 인식 차이를 비교하여 그 결과를 논의하고자 한다. 그리고 말하기 수행평가의 전문성, 효용성 및 시행한계성이라는 요인들 간의 상관관계를 살펴보고 이들이 영어교사들의 말하기 수행평가의 시행에 미치는 영향과 그 효과가 어떠한지를 검증하고자 한다.

5.1 학교급간 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 인식

학교 현장에서 말하기 수행평가를 시행하는 중·고등학교 영어교사들이 말하기 수행평가의 전문성, 전문성의 세 하위 요인들, 효용성 및 시행한계성 등 여섯 가지 요인들에 대해서 학교급간에 따라 통계적으로 유의미한 인식 차이를 보이는지 알아보기 위해서 두 독립표본 t 검정을 실시하였다. 그리고 말하기 수행평가의 전문성, 효용성 및 시행한계성을 측정하는 각 하위 문항에 대해서도 중학교 영어교사 집단과 고등학교 영어교사 집단 간의 인식 차이가 구체적으로 어떻게 다른지를 알아보기 위해서 마찬가지로 두 독립표본 t 검정을 실시하였다. 뿐만 아니라, 말하기 수행평가의 전문성, 효용성 및 시행한계성의 각 요인 내에 속하는 문항들의 평균이 중학교와 고등학교별로 모평균에 대한 95% 신뢰구간에서 서로 어떻게 다른지를 설명하고자 그림으로 비교하였다. 덧붙여, 말하기 수행평가의 시행한계성에 대한 영어교사들의 자유서술식 의견을 범주화하여 정리한 결과를 보고하였다.

5.1.1 학교급간 말하기 수행평가의 전문성에 대한 인식 차이

우선 중학교 영어교사들과 고등학교 영어교사들 집단 간에 말하기 수행평가의 전문성 전체에 대한 그들의 인식 차이를 비교하였다. 중학교와 고등학교 영어교사들의 말하기 수행평가의 전문성 전체에 대한 기술통계는 <표 14>와 같다. 표에서 보듯이, 말하기 수행평가의 전문성 전체에 대한 중학교 영어교사들의 평균점은 4.02점으로 고등학교 영어교사들의 평균점인 3.76점보다 높은 것으로 나타났다.

<표 14> 말하기 수행평가의 전문성 전체에 대한 기술통계

구분	N	최소값	최대값	평균	표준편차	표준오차
중학교	128	2.17	5.00	4.02	.464	.041
고등학교	115	1.42	5.00	3.76	.571	.053

이러한 중학교 영어교사들 집단과 고등학교 영어교사들 집단 간에 말하기 수행평가의 전문성 전체에 대한 인식 차이가 통계적으로 유의미한지를 파악하고자 두 독립표본 t 검정을 실시하였으며, 그 결과는 <표 15>에 제시하였다.

<표 15> 말하기 수행평가의 전문성 전체에 대한 중·고등학교 집단 간의 인식 차이 검증

구분	N	평균	표준편차	표준오차	t	p
중학교	128	4.02	.464	.041	3.909	.000
고등학교	115	3.76	.571	.053		

<표 15>에서 보듯이, 말하기 수행평가의 전문성 전체에 있어서 두 집단 간의 평균점의 차이는 통계적으로 유의미한 것으로 나타났다($t=3.909$, $p<.05$). 이 결과는 입시 위주의 교육과정에 매여 있는 고등학교 영어교사들

에 비해서 아무래도 표현 기능 중심의 실용적인 영어를 가르치는 중학교 영어교사들이 말하기 수행평가를 더 자주 시행하게 되고, 그러다 보니 고등학교 영어교사들보다 중학교 영어교사들이 말하기 수행평가의 전반에 걸친 전문적인 능력과 자질을 갖추게 될 가능성이 더 높기 때문일 것이다. 방인자와 천세영(2011)은 고등학교에서 입시제도에 초점이 맞춰진 영어 수업과 평가 체제로 인해 의사소통 중심의 영어교육의 부재를 지적하였는데, 이는 곧 고등학교 영어교사들의 말하기 수행평가 전문성의 부족과 연관이 있을 것으로 짐작된다.

그리고 말하기 수행평가의 전문성을 구성하는 세 하위 요인들, 즉 말하기 수행평가의 시행에 있어서 평가 방법과 도구의 선정 및 개발, 평가의 실시 및 채점, 그리고 평가 결과의 해석 및 활용에 대한 중학교 영어교사들과 고등학교 영어교사들 간의 인식 차이를 비교하였다. 중학교 영어교사들과 고등학교 영어교사들의 말하기 수행평가의 전문성 세 하위 요인들에 대한 기술 통계는 <표 16>과 같다.

<표 16> 말하기 수행평가의 전문성 세 하위 요인들에 대한 기술통계

하위 요인	구분	N	최소값	최대값	평균	표준편차	표준오차
평가 방법과 도구의 선정 및 개발	중학교	128	1.50	5.00	3.67	.629	.056
	고등학교	115	1.00	5.00	3.35	.772	.072
평가의 실시 및 채점	중학교	128	3.25	5.00	4.62	.456	.040
	고등학교	115	2.25	5.00	4.40	.569	.053
평가 결과의 해석 및 활용	중학교	128	1.00	5.00	3.79	.683	.060
	고등학교	115	1.00	5.00	3.54	.715	.067

<표 16>에서 보듯이, 말하기 수행평가의 전문성 세 하위 요인들 중 평가 방법과 도구의 선정 및 개발이라는 전문성 요인에 대한 중학교 영어교사들의 평균점은 3.67점으로 고등학교 영어교사들의 평균점인 3.35점보다 높은

것으로 나타났다. 말하기 수행평가의 실시 및 채점에 관한 전문성 요인에서도 중학교 영어교사들의 평균점은 4.62점으로 고등학교 영어교사들의 평균점인 4.40점보다 높았다. 게다가 말하기 수행평가 결과의 해석 및 활용이라는 전문성 요인에서도 역시 중학교 영어교사들의 평균점은 3.79점으로 고등학교 영어교사들의 평균점인 3.54점보다 훨씬 높았다. 따라서 말하기 수행평가의 전문성 세 하위 요인들 전부에 있어서 고등학교 영어교사들보다 중학교 영어교사들의 평균점이 높은 것으로 나타났다.

이러한 중학교 영어교사들 집단과 고등학교 영어교사들 집단 간의 말하기 수행평가의 전문성 세 하위 요인들에 대한 인식 차이가 통계적으로 유의미한지를 파악하고자 두 독립표본 t 검정을 실시하였으며, 그 결과는 <표 17>에 제시하였다.

<표 17> 말하기 수행평가의 전문성 세 하위 요인들에 대한 중고등학교 집단 간의 인식 차이 검증

하위 요인	구분	N	평균	표준편차	표준오차	t	p
평가 방법과 도구의 선정 및 개발	중학교	128	3.67	.629	.056	3.513	.001
	고등학교	115	3.35	.772	.072		
평가의 실시 및 채점	중학교	128	4.62	.456	.040	3.230	.001
	고등학교	115	4.40	.569	.053		
평가 결과의 해석 및 활용	중학교	128	3.79	.683	.060	2.763	.006
	고등학교	115	3.54	.715	.067		

<표 17>에서 보듯이, 말하기 수행평가의 전문성 세 하위 요인들, 즉 평가 방법과 도구의 선정 및 개발($t=3.513, p<.05$), 평가의 실시 및 채점($t=3.230, p<.05$), 그리고 평가 결과의 해석 및 활용($t=2.763, p<.05$) 등 모든 영역에서 두 집단 간의 평균점의 차이는 통계적으로 유의미한 것으로 나타났다. 이 결과는 말하기 수행평가 시행의 모든 과정에서 고등학교 영어교사들보다는

중학교 영어교사들의 전문적인 능력에 대한 인식이 대체로 높음을 보여준다. 일선 고등학교에서도 관할 교육청에서 권고하는 대로 일정 비율의 수행평가를 내신 성적에 반영하고 있다. 그런데 실용영어 능력의 향상을 중시하는 중학교와는 달리 고등학교에서는 입시 위주의 학교 정책을 따르다 보니 다소 형식적이고 획일적인 수행평가를 시행하는 경우가 많은 것으로 조사되었다(이재호, 2012). 이는 말하기 수행평가의 시행과 관련된 구체적인 분야에서 고등학교 영어교사들의 전문성이 상대적으로 떨어질 수밖에 없는 이유로 해석될 수 있다.

아울러 말하기 수행평가의 전문성을 측정하는 각 문항에 대해서 중학교 영어교사들과 고등학교 영어교사들의 인식이 구체적으로 어떻게 차이가 나는지를 비교하였다. 우선 중학교 영어교사들과 고등학교 영어교사들 간에 말하기 수행평가의 전문성 세 하위 요인들 중 평가 방법과 도구의 선정 및 개발 요인의 각 문항에 대한 기술통계는 <표 18>과 같다.

<표 18> 평가 방법과 도구의 선정 및 개발 각 문항에 대한 기술통계

요인	개별 문항	구분	N	최소값	최대값	평균	표준편차	표준오차
방법과	1. 평가 목적과 내용에 적합한 방법과 도구 선택	중학교	128	1	5	3.65	.749	.066
		고등학교	115	1	5	3.37	.811	.076
도구의	2. 학생의 성취 수준에 적합한 평가 방법과 도구 선택	중학교	128	2	5	3.83	.744	.066
		고등학교	115	1	5	3.56	.808	.075
선정·	3. 적합한 평가 방법이나 도구의 부재 시 직접 개발	중학교	128	1	5	3.70	.847	.075
		고등학교	115	1	5	3.23	.965	.090
개발	4. 과제 및 채점 기준과 원리에 맞게 직접 개발	중학교	128	1	5	3.49	.869	.077
		고등학교	115	1	5	3.24	.961	.090

중학교 영어교사들과 고등학교 영어교사들의 평가 방법과 도구의 선정 및 개발 요인의 각 문항을 비교한 결과, <표 18>에서 보듯이 모든 문항에서 중

학교 영어교사들의 평균점이 고등학교 영어교사들의 평균점보다 높은 것으로 나타났다. 이러한 중학교 영어교사들 집단과 고등학교 영어교사들 집단 간에 말하기 수행평가의 방법과 도구의 선정 및 개발에 대한 인식 차이가 각 문항별로 통계적으로 유의미한지를 파악하고자 두 독립표본 t 검정을 실시하였으며, 그 결과는 <표 19>에 제시하였다.

<표 19> 평가 방법과 도구의 선정 및 개발 각 문항에 대한 중고등학교 집단 간의 인식 차이 검증

요인	개별 문항	구분	N	평균	표준편차	표준오차	t	p
방법과	1. 평가 목적과 내용에 적합한 방법과 도구 선택	중학교	128	3.65	.749	.066	2.744	.007
		고등학교	115	3.37	.811	.076		
도구의	2. 학생의 성취 수준에 적합한 평가 방법과 도구 선택	중학교	128	3.83	.744	.066	2.717	.007
		고등학교	115	3.56	.808	.075		
선정·개발	3. 적합한 평가 방법이나 도구의 부재 시 직접 개발	중학교	128	3.70	.847	.075	4.009	.000
		고등학교	115	3.23	.965	.090		
개발	4. 과제 및 채점 기준과 원리에 맞게 직접 개발	중학교	128	3.49	.869	.077	2.119	.035
		고등학교	115	3.24	.961	.090		

<표 19>에서 보듯이, 평가 방법과 도구의 선정 및 개발 요인에 있어서 두 집단 간의 평균점의 차이가 모든 문항에서 통계적으로 유의미한 것으로 나타났다(문항 1번: $t=2.744$, $p<.05$; 문항 2번: $t=2.717$, $p<.05$; 문항 3번: $t=4.009$, $p<.05$; 문항 4번: $t=2.119$, $p<.05$). 이 결과는 고등학교 영어교사들에 비해서 중학교 영어교사들이 학생들의 말하기 능력을 평가하기에 가장 적합한 평가 방법과 도구를 선택하고, 혹은 필요한 경우 그러한 평가 방법과 도구를 과제와 채점 기준 및 원리에 맞게 직접 개발하는 전문적인 능력에 대한 인식이 상대적으로 높음을 보여준다. 즉, 말하기 수행평가의 상대적인 시행 가능성이 사실 고등학교보다는 중학교에서 더 높기 때문에 말하기 수행평가의 방법이나 도구의 선택과 개발에 관한 모든 전문적인 능력도 역시 중학교 영어교사들에게서 더 높은 것으로 짐작된다.

덧붙여, 말하기 수행평가의 방법과 도구의 선정 및 개발 요인의 각 문항이 중학교 영어교사들과 고등학교 영어교사들 각 집단 내에서 유의미한 차이가 있는지를 파악하고자 집단별 모평균에 대한 95% 신뢰구간²⁾을 추정하였다. <그림 3>과 <그림 4>는 학교별로 평가 방법과 도구의 선정 및 개발의 개별 문항들에 대한 평균점 차이의 분포를 나타낸 결과이다.



<그림 3> 평가 방법과 도구의 선정 및 개발 각 문항별 신뢰구간(중학교)

<그림 3>은 중학교 영어교사들이 응답한 말하기 수행평가의 방법과 도구의 선정 및 개발 요인의 각 문항별 모평균의 상한값과 하한값의 분포를 상하 일직선으로 표시한 것이다. 이 네 문항 중 2번 문항인 학생의 성취 수준에 적합한 평가 방법과 도구 선택(신뢰구간: 3.701~3.959)과 4번 문항인 과제 및 채점 기준의 개발 원리에 맞게 직접 개발(신뢰구간: 3.339~3.641)이 서로 모평균의 95% 신뢰구간에서 뚜렷한 차이가 있는 것으로 나타났다. 구체적으로, 말하기 수행평가의 방법과 도구의 선정 및 개발에 관한 전문성 중에서 중학교 영어교

2) 집단별 모평균에 대한 95% 신뢰구간 계산법: $\bar{x}(\text{평균}) \pm 1.96 \times SE(\text{평균의 표준오차})$

사들의 학생의 성취 수준에 적합한 말하기 수행평가의 방법과 도구를 선택하는 능력에 대한 인식은 과제 및 채점 기준의 개발 원리에 맞게 말하기 수행평가의 방법이나 도구를 직접 개발하는 능력에 대한 인식보다 높은 것으로 나타났다.



<그림 4> 평가 방법과 도구의 선정 및 개발 각 문항별 신뢰구간(고등학교)

<그림 4>는 고등학교 영어교사들이 응답한 말하기 수행평가의 방법과 도구의 선정 및 개발 요인의 각 문항별 모평균의 상한값과 하한값의 분포를 상하 일직선으로 표시한 것이다. 이 네 문항 중 2번 문항인 학생의 성취 수준에 적합한 평가 방법과 도구 선택(신뢰구간: 3.413~3.707)과 3번 문항인 적합한 평가 방법이나 도구의 부재 시 직접 개발(신뢰구간: 3.054~3.406)이 서로 모평균의 95% 신뢰구간에서 뚜렷한 차이가 있는 것으로 나타났다. 구체적으로, 말하기 수행평가의 방법과 도구의 선정 및 개발과 관련된 전문성 중에서 고등학교 영어교사들의 학생의 성취 수준에 적합한 말하기 수행평가 방법과 도구를 선택하는 능력에 대한 인식은 적합한 말하기 수행평가 방법이나 도구가 없을 때 직접 개발하는 능력에 대한 인식보다 높은 것으로 밝혀졌다.

정리하면, 중학교 영어교사들과 고등학교 영어교사들 두 집단 모두 말하기 수행평가의 방법과 도구의 선정 및 개발이라는 전문성 분야에서 다른 능력보다 학생의 성취 수준에 적합한 평가 방법과 도구를 선택하는 능력에 대한 인식이 더 높았다. 그러므로 학교 현장에서 진정한 말하기 수행평가의 시행을 위해서는 시·도 교육청 차원에서 말하기 수행평가에 대한 다양한 평가 방법이나 도구를 개발하고 이와 관련된 교사 연수 프로그램을 자주 마련할 필요가 있을 것이다(김미은, 2002; 김정임, 2006; 문지형, 2007). 아무래도 말하기 수행평가에 대한 풍부한 자료와 정보를 갖춘 영어교사들은 말하기 수행평가에 대한 전문성이 높아질 것이고, 또한 말하기 수행평가의 시행에 이를 적극적으로 활용하여 학생들의 말하기 능력의 향상에 충분히 기여할 수 있을 것으로 예측된다.

다음으로 말하기 수행평가의 전문성을 측정하는 세 하위 요인들 중 평가의 실시 및 채점 요인의 각 문항에 대해서 중학교 영어교사들과 고등학교 영어교사들의 인식이 구체적으로 어떻게 차이가 나는지를 비교하였다. <표 20>은 평가의 실시 및 채점 요인의 각 문항에 대한 기술통계이다.

<표 20> 평가의 실시 및 채점 각 문항에 대한 기술통계

요인	개별 문항	구분	N	최소값	최대값	평균	표준편차	표준오차
평가의 실시· 채점	1. 평가 실시 전 학생들에게 평가 계획 공지	중학교	128	4	5	4.73	.443	.039
		고등학교	115	2	5	4.50	.654	.061
	2. 평가 실시 전 평가 방법과 도구의 숙지	중학교	128	3	5	4.62	.549	.049
		고등학교	115	1	5	4.29	.747	.070
	3. 평가 실시 전 학생들에게 과제 및 채점 기준 설명	중학교	128	2	5	4.56	.599	.053
		고등학교	115	3	5	4.43	.578	.054
	4. 채점 기준 숙지 후 채점 실시	중학교	128	3	5	4.55	.559	.049
		고등학교	115	2	5	4.39	.658	.061

중학교 영어교사들과 고등학교 영어교사들의 평가의 실시 및 채점 요인의 각 문항을 비교한 결과, <표 20>에서 보듯이 모든 문항에서 중학교 영어교사들의 평균점이 고등학교 영어교사들의 평균점보다 높은 것으로 나타났다. 이러한 중학교 영어교사들 집단과 고등학교 영어교사들 집단 간의 말하기 수행평가의 실시 및 채점에 대한 인식 차이가 각 문항별로 통계적으로 유의미한지를 파악하고자 두 독립표본 t 검정을 실시하였으며, 그 결과는 <표 21>에 제시하였다.

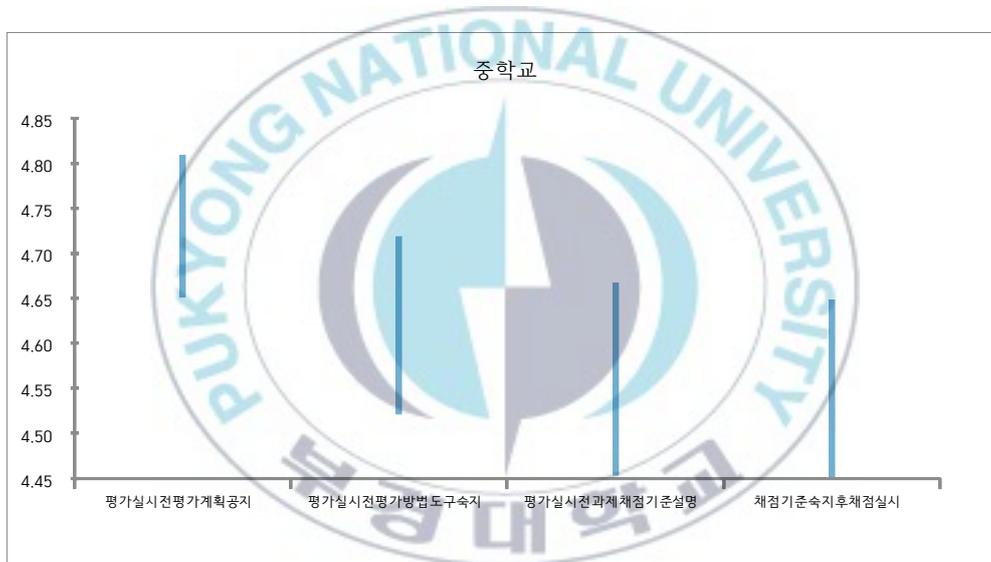
<표 21> 평가의 실시 및 채점 각 문항에 대한 중·고등학교 집단 간의 인식 차이 검증

요인	개별 문항	구분	N	평균	표준편차	표준오차	t	p
평가의 실시·채점	1. 평가 실시 전 학생들에게 평가 계획 공지	중학교	128	4.73	.443	.039	3.173	.002
		고등학교	115	4.50	.654	.061		
	2. 평가 실시 전 평가 방법과 도구의 숙지	중학교	128	4.62	.549	.049	3.892	.000
		고등학교	115	4.29	.747	.070		
	3. 평가 실시 전 학생들에게 과제 및 채점 기준 설명	중학교	128	4.56	.599	.053	1.802	.073
		고등학교	115	4.43	.578	.054		
	4. 채점 기준 숙지 후 채점 실시	중학교	128	4.55	.559	.049	2.092	.037
		고등학교	115	4.39	.658	.061		

<표 21>에서 보듯이, 평가의 실시 및 채점 요인의 경우 3번 문항($t=1.802$, $p>.05$)을 제외한 다른 나머지 문항에서 두 집단 간의 평균점의 차이가 통계적으로 유의미한 것으로 나타났다(문항 1번: $t=3.173$, $p<.05$; 문항 2번: $t=3.892$, $p<.05$; 문항 4번: $t=2.092$, $p<.05$). 이 결과는 고등학교 영어교사들보다 중학교 영어교사들이 말하기 수행평가를 시행하기 전에 평가 계획이나 평가 방법과 도구에 대한 사전 준비를 철저하게 할 뿐만 아니라, 말하기 수행평가를 시행한 후에도 채점 원리와 기준에 맞게 정확하고 공정하게 채점하는 전문적인 능력에 대한 인식을 더 높게 가지고 있음을 보여준다. 즉, 말하기 수행평가 시행의 기회가 사실 고등학교보다는 중학교에서 더 많기 때

문에 말하기 수행평가의 실제적인 준비와 전반적인 시행에 대한 전문적인 능력 또한 중학교 영어교사들에게서 대체로 더 높은 것으로 짐작된다.

덧붙여, 말하기 수행평가의 실시 및 채점 요인의 각 문항이 중학교 영어교사들과 고등학교 영어교사들 각 집단 내에서 유의미한 차이가 있는지를 파악하고자 집단별 모평균에 대한 95% 신뢰구간을 추정하였다. <그림 5>와 <그림 6>은 학교별로 평가의 실시 및 채점의 개별 문항들에 대한 평균점 차이의 분포를 나타낸 결과이다.



<그림 5> 평가의 실시 및 채점 각 문항별 신뢰구간(중학교)

<그림 5>는 중학교 영어교사들에 대한 말하기 수행평가의 실시 및 채점 요인의 각 문항별 모평균의 상한값과 하한값의 분포를 상하 일직선으로 표시한 것이다. 이 네 문항 중 1번 문항인 평가 실시 전 학생들에게 평가 계획 공지(신뢰구간: 4.654~4.806)와 4번 문항인 채점 기준 숙지 후 채점 실시(신뢰구간: 4.454~4.646)가 서로 모평균의 95% 신뢰구간에서 뚜렷한 차이가 있는 것으로 나타났다. 구체적으로, 말하기 수행평가의 실시 및 채점과 관련된

전문성 중에서 중학교 영어교사들은 말하기 수행평가를 실시하기 전에 학생들에게 전반적인 평가 계획을 공지하는 능력에 대한 인식이 말하기 수행평가의 채점 기준을 정확하게 숙지하고서 학생들의 수행평가 결과를 채점하는 능력에 대한 인식보다 높은 것으로 나타났다.

<그림 6>은 고등학교 영어교사들에 대한 말하기 수행평가의 실시 및 채점 요인의 각 문항별 모평균의 상한값과 하한값의 분포를 상하 일직선으로 표시한 것이다. 이 네 문항 중에서 모평균의 95% 신뢰구간에서 뚜렷한 차이가 있는 문항들은 하나도 없었다.



<그림 6> 평가의 실시 및 채점 각 문항별 신뢰구간(고등학교)

마지막으로, 말하기 수행평가의 전문성을 측정하는 세 하위 요인들 중 평가 결과의 해석 및 활용 요인의 각 문항에 대해서 중학교 영어교사들과 고등학교 영어교사들의 인식이 구체적으로 어떻게 다른지를 비교하였다. <표 22>는 평가 결과의 해석 및 활용 요인의 각 문항에 대한 기술통계이다.

<표 22> 평가 결과의 해석 및 활용 각 문항에 대한 기술통계

요인	개별 문항	구분	N	최소값	최대값	평균	표준편차	표준오차
평가	1. 평가 목적과 내용에 기초한 평가 결과 해석	중학교	128	1	5	3.95	.724	.064
		고등학교	115	1	5	3.74	.739	.069
결과의	2. 평가 결과 해석 시 여러 변수 등의 고려	중학교	128	1	5	3.78	.793	.070
		고등학교	115	1	5	3.52	.809	.075
해석·활용	3. 평가 결과에 근거한 학생의 성취적 특성 파악	중학교	128	1	5	3.80	.833	.074
		고등학교	115	1	5	3.57	.860	.080
4. 평가 결과의 종합과 교수·학습 자료로 활용	중학교	128	1	5	3.62	.888	.079	
	고등학교	115	1	5	3.33	.876	.082	

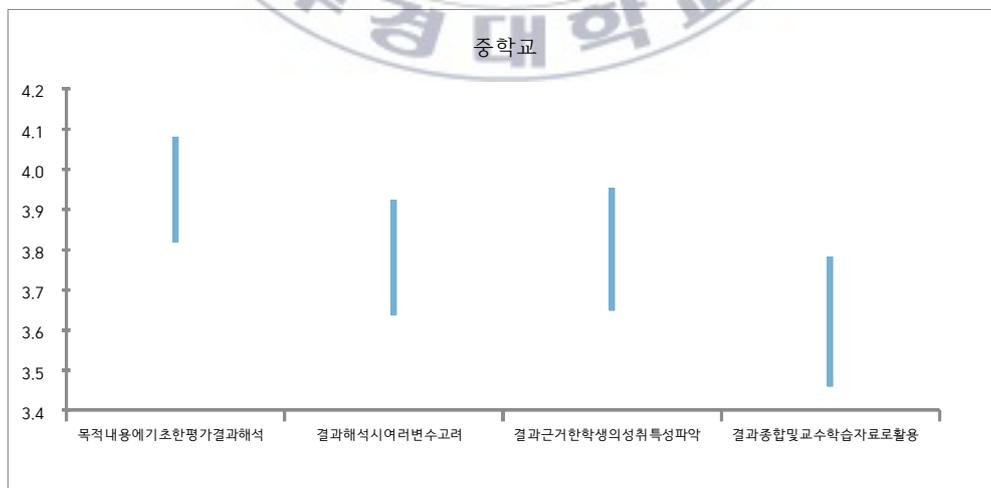
<표 22>에서 보듯이, 중학교 영어교사들과 고등학교 영어교사들의 평가 결과의 해석 및 활용 요인의 각 문항을 비교한 결과, 모든 문항에서 중학교 영어교사들의 평균점이 고등학교 영어교사들의 평균점보다 높은 것으로 나타났다. 이러한 중학교 영어교사들 집단과 고등학교 영어교사들 집단 간에 말하기 수행평가 결과의 해석 및 활용에 대한 인식 차이가 각 문항별로 통계적으로 유의미한지를 파악하고자 두 독립표본 t 검정을 실시하였으며, 그 결과는 <표 23>에 제시하였다.

<표 23> 평가 결과의 해석 및 활용 각 문항에 대한 중·고등학교 집단 간의 인식 차이 검증

요인	개별 문항	구분	N	평균	표준편차	표준오차	t	p
평가	1. 평가 목적과 내용에 기초한 평가 결과 해석	중학교	128	3.95	.724	.064	2.195	.029
		고등학교	115	3.74	.739	.069		
결과의	2. 평가 결과 해석 시 여러 변수 등의 고려	중학교	128	3.78	.793	.070	2.522	.012
		고등학교	115	3.52	.809	.075		
해석·활용	3. 평가 결과에 근거한 학생의 성취적 특성 파악	중학교	128	3.80	.833	.074	2.204	.028
		고등학교	115	3.57	.860	.080		
4. 평가 결과의 종합과 교수·학습 자료로 활용	중학교	128	3.62	.888	.079	2.529	.012	
	고등학교	115	3.33	.876	.082			

<표 23>에서 보듯이, 평가 결과의 해석 및 활용 요인에 있어서 두 집단 간의 평균점의 차이가 모든 문항에서 통계적으로 유의미한 것으로 나타났다(문항 1번: $t=2.195$, $p<.05$; 문항 2번: $t=2.522$, $p<.05$; 문항 3번: $t=2.204$, $p<.05$; 문항 4번: $t=2.529$, $p<.05$). 이 결과는 고등학교 영어교사들보다 중학교 영어교사들이 평가의 목적과 내용에 맞게 여러 변수들을 고려하여 평가 결과를 해석하고, 또한 그 평가 결과에 근거하여 학생들의 학습에 대한 정보 수집 및 차후 교수-학습에 대한 자료로 활용하는 전문성에 대한 인식이 높음을 보여준다. 즉, 말하기 수행평가의 실제적인 시행이 고등학교보다는 중학교에서 더 많기 때문에 말하기 수행평가의 결과 해석과 활용에 대한 모든 전문적인 능력도 역시 중학교 영어교사들에게서 더 높은 것으로 짐작된다.

덧붙여, 말하기 수행평가 결과의 해석 및 활용 요인의 각 문항이 중학교 영어교사들과 고등학교 영어교사들 각 집단 내에서 유의미한 차이가 있는지를 파악하고자 집단별 모평균에 대한 95% 신뢰구간을 추정하였다. <그림 7>과 <그림 8>은 학교별로 평가 결과의 해석 및 활용의 개별 문항들에 대한 평균점 차이의 분포를 나타낸 결과이다.



<그림 7> 평가 결과의 해석 및 활용 각 문항별 신뢰구간(중학교)

<그림 7>은 중학교 영어교사들이 응답한 말하기 수행평가 결과의 해석 및 활용 요인의 각 문항별 모평균의 상한값과 하한값의 분포를 상하 일직선으로 표시한 것이다. 이 네 문항 중 1번 문항인 평가 목적과 내용에 기초한 평가 결과의 해석(신뢰구간: 3.825~4.075)과 4번 문항인 평가 결과의 종합과 교수·학습 자료로의 활용(신뢰구간: 3.465~3.775)이 서로 모평균의 95% 신뢰구간에서 뚜렷한 차이가 있는 것으로 나타났다. 구체적으로, 말하기 수행평가 결과의 해석 및 활용에 관한 전문성 중에서 중학교 영어교사들의 말하기 수행평가의 목적과 내용에 기초하여 평가 결과를 해석하는 능력에 대한 인식이 말하기 수행평가의 결과를 종합하여 차후 교사의 교수·학습의 개선을 위한 참고 자료로 활용하는 능력에 대한 인식보다 높은 것으로 나타났다.



<그림 8> 평가 결과의 해석 및 활용 각 문항별 신뢰구간(고등학교)

<그림 8>은 고등학교 영어교사들이 응답한 말하기 수행평가 결과의 해석 및 활용 요인의 각 문항별 모평균의 상한값과 하한값의 분포를 상하 일직선으로 표시한 것이다. 이 네 문항 중 1번 문항인 평가 목적과 내용에 기초한 평

가 결과의 해석(신뢰구간: 3.605~3.875)과 4번 문항인 평가 결과의 종합과 교수·학습 자료로의 활용(신뢰구간: 3.169~3.491)이 서로 모평균의 95% 신뢰구간에서 뚜렷한 차이가 있는 것으로 나타났다. 구체적으로, 중학교 영어교사들과 마찬가지로 고등학교 영어교사들도 말하기 수행평가 결과의 해석 및 활용에 관한 전문성의 측면에서 말하기 수행평가의 목적과 내용에 근거하여 평가 결과를 해석하는 능력에 대한 인식이 말하기 수행평가의 결과를 종합하여 차후 교수·학습의 개선을 위해서 활용하는 능력에 대한 인식보다 높은 것으로 밝혀졌다.

정리하면, 중학교 영어교사들과 고등학교 영어교사들 두 집단 모두 말하기 수행평가 결과의 해석 및 활용이라는 전문성 분야에서 다른 능력보다 말하기 수행평가의 목적과 내용에 기초하여 평가 결과를 해석하는 능력에 대한 인식이 더 높았다. 학생들의 말하기 능력에 대한 성취 수준을 파악하기 위해서 영어교사들이 말하기 수행평가를 시행한 후 학생들의 말하기 수행평가의 결과를 제대로 해석하는 능력은 아주 중요하다. 이는 학생들의 성취 수준을 진단하여 그들의 말하기 능력의 향상을 위한 발전적인 조언을 하는 등 다양한 측면의 평가 정보를 적극적으로 활용하는 데 반드시 필요한 기초적인 자질이기 때문이다. 그리고 평가 결과에 대한 단순한 해석보다는 교사가 구체적이고 자세한 의견을 제공하는 일은 학습에 대한 학생들의 수행 능력을 향상시키는 계기가 될 수 있다(Brookhart, 1997).

5.1.2 학교급간 말하기 수행평가의 효용성에 대한 인식 차이

말하기 수행평가의 전반적인 효용성에 대해서 중학교 영어교사들과 고등학교 영어교사들 집단 간의 인식 차이를 비교하였고, 그에 대한 기술통계는 <표 24>

와 같다. 말하기 수행평가의 효용성 전체에 대한 중학교 영어교사들의 평균점은 3.48점으로 고등학교 영어교사들의 평균점인 3.36점보다 높은 것으로 나타났다.

<표 24> 말하기 수행평가의 효용성 전체에 대한 기술통계

구분	N	최소값	최대값	평균	표준편차	표준오차
중학교	128	2.00	5.00	3.48	.628	.056
고등학교	115	1.00	5.00	3.36	.688	.064

이러한 중학교 영어교사들과 고등학교 영어교사들 집단 간에 말하기 수행평가의 전반적인 효용성에 대한 인식 차이가 통계적으로 유의미한지를 파악하고자 두 독립표본 t 검정을 실시하였으며, 그 결과는 <표 25>에 제시하였다.

<표 25> 말하기 수행평가의 효용성 전체에 대한 중·고등학교 집단 간의 인식 차이 검증

구분	N	평균	표준편차	표준오차	t	p
중학교	128	3.48	.628	.056	1.455	.147
고등학교	115	3.36	.688	.064		

<표 25>에서 보듯이, 말하기 수행평가의 전반적인 효용성에 대한 인식에서 두 집단 간의 평균점의 차이는 통계적으로 유의미하지 않은 것으로 나타났다($t=1.455$, $p>.05$). 이는 학교 현장에서 말하기 수행평가의 시행이 학생들의 영어 말하기 능력뿐만 아니라 영어교사의 말하기 수업과 관련된 모든 점에서 도움이 될 수 있다는 인식이 중학교 영어교사들과 고등학교 영어교사들 사이에 크게 다르지 않다는 것을 보여준다. 이러한 결과는 말하기 수행평가의 시행을 통해서 중학교와 고등학교 학생들의 영어 수업과 평가에 대한 인식이 각각 긍정적으로 변하게 되었다는 윤소영(1999)과 최주경(2009)의 연구 결과와 연관이 있다.

그리고 말하기 수행평가의 효용성을 측정하는 각 문항에 대해서 중학교 영어교사들과 고등학교 영어교사들의 인식이 구체적으로 어떻게 차이가 나는지를 비교하였다. 중학교 영어교사들과 고등학교 영어교사들의 말하기 수행평가의 효용성 각 문항에 대한 기술통계는 <표 26>과 같다.

<표 26> 말하기 수행평가의 효용성 각 문항에 대한 기술통계

효용성 개별 문항	구분	N	최소값	최대값	평균	표준편차	표준오차
1. 학생의 영어 말하기 능력 향상에 도움이 됨.	중학교	128	2	5	3.48	.753	.067
	고등학교	115	1	5	3.38	.833	.078
2. 학생의 영어 말하기 능력 발달 과정의 파악에 도움이 됨.	중학교	128	2	5	3.54	.812	.072
	고등학교	115	1	5	3.41	.826	.077
3. 학생의 영어 말하기 성취 수준의 파악에 도움이 됨.	중학교	128	2	5	3.72	.763	.067
	고등학교	115	1	5	3.50	.862	.080
4. 학생에게 영어 말하기에 대한 학습 동기나 흥미 제공에 도움이 됨.	중학교	128	2	5	3.42	.790	.070
	고등학교	115	1	5	3.39	.835	.078
5. 교사의 영어 말하기 수업 방법의 개선에 도움이 됨.	중학교	128	2	5	3.25	.823	.073
	고등학교	115	1	5	3.12	.890	.083

중학교 영어교사들과 고등학교 영어교사들의 말하기 수행평가의 효용성 각 문항을 비교한 결과, <표 26>에서 보듯이 모든 문항에서 중학교 영어교사들의 평균점이 고등학교 영어교사들의 평균점보다 높은 것으로 나타났다. 이러한 중학교 영어교사들 집단과 고등학교 영어교사들 집단 간에 말하기 수행평가의 효용성에 대한 인식 차이가 각 문항별로 통계적으로 유의미한지를 파악하고자 두 독립표본 t 검정을 실시하였으며, 그 결과는 <표 27>에 제시하였다.

<표 27>에서 보듯이, 말하기 수행평가의 효용성 각 문항 중 3번 문항인 학생의 영어 말하기 성취 수준의 파악에 도움이 됨($t=2.126, p<.05$)에 있어서만 두 집단 간의 평균점의 차이가 통계적으로 유의미한 것으로 나타났다. 이 결과는 고등학교 영어교사들보다 중학교 영어교사들이 말하기 수행평가

의 효용성 측면에서, 특히 말하기 수행평가의 시행을 통해서 학생들의 영어 말하기 성취 수준을 파악하는 데 도움을 얻는다는 인식이 더 높음을 보여준다. 즉, 수행평가를 통해서 학생들의 강점과 약점 및 성취도에 대한 종합적인 평가를 할 수 있는데(O'Brien, 1997), 입시를 중시하는 고등학교 영어교사들보다 실용영어를 중시하는 중학교 영어교사들이 아무래도 말하기 수행평가를 더 자주 시행하게 될 기회가 많다(고현우, 2014). 따라서 학생들의 말하기 성취 수준을 파악하는 데 도움이 된다는 인식은 아무래도 고등학교 영어교사들보다 중학교 영어교사들에게서 더 높은 것으로 추측된다.

<표 27> 말하기 수행평가의 효용성 각 문항에 대한 중·고등학교 집단 간의 인식 차이 검증

효용성 개별 문항	구분	N	평균	표준편차	표준오차	t	p
1. 학생의 영어 말하기 능력 향상에 도움이 됨.	중학교	128	3.48	.753	.067	1.000	.318
	고등학교	115	3.38	.833	.078		
2. 학생의 영어 말하기 능력 발달 과정의 파악에 도움이 됨.	중학교	128	3.54	.812	.072	1.239	.217
	고등학교	115	3.41	.826	.077		
3. 학생의 영어 말하기 성취 수준의 파악에 도움이 됨.	중학교	128	3.72	.763	.067	2.126	.035
	고등학교	115	3.50	.862	.080		
4. 학생에게 영어 말하기에 대한 학습 동기나 흥미 제공에 도움이 됨.	중학교	128	3.42	.790	.070	.293	.770
	고등학교	115	3.39	.835	.078		
5. 교사의 영어 말하기 수업 방법의 개선에 도움이 됨.	중학교	128	3.25	.823	.073	1.167	.244
	고등학교	115	3.12	.890	.083		

덧붙여, 말하기 수행평가의 효용성 각 문항이 중학교 영어교사들과 고등학교 영어교사들 각 집단 내에서 유의미한 차이가 있는지를 파악하고자 집단별 모평균에 대한 95% 신뢰구간을 추정하였다. <그림 9>와 <그림 10>은 학교별로 말하기 수행평가의 효용성의 개별 문항들에 대한 평균점 차이의 분포를 나타낸 결과이다.

<그림 9>는 중학교 영어교사들에 대한 말하기 수행평가의 효용성 각 문항별 모평균의 상한값과 하한값의 분포를 상하 일직선으로 표시한 것이다. 이

다섯 문항 중 3번 문항인 학생의 영어 말하기 성취 수준의 파악에 도움(신뢰구간: 3.589~3.851)과 4번 문항인 학생에게 영어 말하기에 대한 학습 동기나 흥미 제공에 도움(신뢰구간: 3.283~3.557)이 서로 모평균의 95% 신뢰구간에서 뚜렷한 차이가 있는 것으로 나타났다. 또한, 3번 문항은 5번 문항인 교사의 영어 말하기 수업 방법 개선에 도움(신뢰구간: 3.107~3.393)과도 서로 모평균의 95% 신뢰구간에서 뚜렷한 차이가 있었다. 즉, 말하기 수행평가의 시행으로 인해 중학교 영어교사들은 학교 현장에서 학생들의 영어 말하기에 대한 학습 동기나 흥미가 높아지거나 혹은 교사의 영어 말하기 수업 방법이 개선될 수 있다고 생각하기보다는, 오히려 학생들의 영어 말하기 성취 수준을 파악하는 데 도움이 된다는 생각을 더 많이 가지고 있는 것으로 나타났다.



<그림 9> 말하기 수행평가의 효용성 각 문항별 신뢰구간(중학교)

<그림 10>은 고등학교 영어교사들에 대한 말하기 수행평가의 효용성 각 문항별 모평균의 상한값과 하한값의 분포를 상하 일직선으로 표시한 것이다. 이 다섯 문항 중 3번 문항인 학생의 영어 말하기 성취 수준의 파악에 도움

(신뢰구간: 3.343~3.657)과 5번 문항인 교사의 영어 말하기 수업 방법 개선에 도움(신뢰구간: 2.957~3.283)이 서로 모평균의 95% 신뢰구간에서 뚜렷한 차이가 있는 것으로 나타났다. 즉, 고등학교 영어교사들도 역시 말하기 수행평가를 통해서 교사의 영어 말하기 수업 방법을 개선하는 데 도움을 얻는다는 인식보다는, 오히려 학생들의 영어 말하기 성취 수준을 파악하는 데 도움이 된다는 인식을 더 많이 가지고 있는 것으로 나타났다.



<그림 10> 말하기 수행평가의 효용성 각 문항별 신뢰구간(고등학교)

정리하면, 중학교 영어교사들과 고등학교 영어교사들 두 집단 모두 말하기 수행평가의 효용성에 대한 다른 인식들보다는 말하기 수행평가는 학생들의 영어 말하기에 대한 성취 수준을 파악하는 데 도움을 준다는 인식을 보다 더 높게 가지고 있었다. 이는 앞서 언급된 바와 같이 학교 수업에서 말하기 수행평가의 시행과 학생들의 학업 성취도와와의 긍정적인 관계에 관한 선행 연구들(김인철, 1998; 고현우, 2014; Lim, 1992; Wudong, 1994; Zhou, 1991)의 결과를 다시 한 번 강조하고 있다.

5.1.3 학교급간 말하기 수행평가의 시행한계성에 대한 인식 차이

말하기 수행평가의 전반적인 시행한계성에 대해서 중학교 영어교사들과 고등학교 영어교사들 집단 간의 인식 차이를 비교하였고, 그에 대한 기술통계는 <표 28>과 같다. 표에서 보듯이, 말하기 수행평가의 시행한계성 전체에 대한 고등학교 영어교사들의 평균점은 3.64점으로 중학교 영어교사들의 평균점인 3.34점보다 높은 것으로 나타났다.

<표 28> 말하기 수행평가의 시행한계성 전체에 대한 기술통계

구분	N	최소값	최대값	평균	표준편차	표준오차
중학교	128	1.00	4.80	3.34	.727	.064
고등학교	115	2.20	5.00	3.64	.650	.061

이러한 중학교 영어교사들과 고등학교 영어교사들 집단 간에 말하기 수행평가의 전반적인 시행한계성에 대한 인식 차이가 통계적으로 유의미한지를 파악하고자 두 독립표본 t 검정을 실시하였으며, 그 결과는 <표 29>에 제시하였다.

<표 29> 말하기 수행평가의 시행한계성 전체에 대한 중·고등학교 집단 간의 인식 차이 검증

구분	N	평균	표준편차	표준오차	t	p
중학교	128	3.34	.727	.064	-3.345	.001
고등학교	115	3.64	.650	.061		

<표 29>에서 보듯이, 말하기 수행평가의 전반적인 시행한계성에 대한 인식에서 두 집단 간의 평균점의 차이는 통계적으로 유의미한 것으로 나타났다($t=-3.345$, $p<.05$). 이 결과는 중학교 영어교사들과 고등학교 영어교사들 모두 말하기 수행평가에 대한 긍정적인 측면을 공감하고 있더라도, 말하기

수행평가의 실질적인 시행을 위한 교육적 여건들이 아무래도 고등학교보다는 중학교에 더 많이 조성되어 있기 때문일 것으로 추측된다. 이는 방인자와 천세영(2011) 및 장인순(2004)이 주장하듯이, 대학입시의 준비로 인해 많은 고등학교에서는 진정한 의사소통 중심의 말하기 수업과 말하기 수행평가를 시행하기가 어려운 실정임을 설명하고 있다.

그리고 말하기 수행평가의 시행한계성을 측정하는 각 문항에 대해서 중학교 영어교사들과 고등학교 영어교사들의 인식이 구체적으로 어떻게 차이가 나는지를 비교하였다. <표 30>은 중학교 영어교사들과 고등학교 영어교사들의 말하기 수행평가의 시행한계성 각 문항에 대한 기술통계이다.

<표 30> 말하기 수행평가의 시행한계성 각 문항에 대한 기술통계

시행한계성 개별 문항	구분	N	최소값	최대값	평균	표준편차	표준오차
1. 학급 규모가 크거나 학생 수가 많아 시행하기가 어려움.	중학교	128	1	5	3.77	.941	.083
	고등학교	115	2	5	3.94	.841	.078
2. 채점의 객관성 확보가 어려워 시행하기가 어려움.	중학교	128	1	5	3.50	.914	.081
	고등학교	115	1	5	3.75	.944	.088
3. 적절한 평가 방법이나 도구가 없어 시행하기가 어려움.	중학교	128	1	5	3.29	.949	.084
	고등학교	115	1	5	3.57	.900	.084
4. 도움이 되는 자료나 교사 연수가 부족해 시행하기가 어려움.	중학교	128	1	5	3.14	.839	.074
	고등학교	115	1	5	3.41	.917	.085
5. 학생들에게 또 다른 학습 부담이 될 수 있어서 시행하기가 어려움.	중학교	128	1	5	2.98	.964	.085
	고등학교	115	2	5	3.51	.820	.077

중학교 영어교사들과 고등학교 영어교사들의 말하기 수행평가의 시행한계성 각 문항을 비교한 결과, <표 30>에서 보듯이 모든 문항에서 고등학교 영어교사들의 평균점이 중학교 영어교사들의 평균점보다 높은 것으로 나타났다. 이러한 중학교 영어교사들 집단과 고등학교 영어교사들 집단 간에 말하기 수행평가의 시행한계성에 대한 인식 차이가 각 문항별로 통계적으로 유

의미한지를 파악하고자 두 독립표본 t 검정을 실시하였으며, 그 결과는 <표 31>에 제시하였다.

<표 31> 말하기 수행평가의 시행한계성 각 문항에 대한 중·고등학교 집단 간의 인식 차이 검증

시행한계성 개별 문항	구분	N	평균	표준 편차	표준 오차	t	p
1. 학급 규모가 크거나 학생 수가 많아 시행하기가 어려움.	중학교	128	3.77	.941	.083	-1.450	.148
	고등학교	115	3.94	.841	.078		
2. 채점의 객관성 확보가 어려워 시행하기가 어려움.	중학교	128	3.50	.914	.081	-2.078	.039
	고등학교	115	3.75	.944	.088		
3. 적절한 평가 방법이나 도구가 없어 시행하기가 어려움.	중학교	128	3.29	.949	.084	-2.322	.021
	고등학교	115	3.57	.900	.084		
4. 도움이 되는 자료나 교사 연수가 부족해 시행하기가 어려움.	중학교	128	3.14	.839	.074	-2.380	.018
	고등학교	115	3.41	.917	.085		
5. 학생들에게 또 다른 학습 부담이 될 수 있어서 시행하기가 어려움.	중학교	128	2.98	.964	.085	-4.578	.000
	고등학교	115	3.51	.820	.077		

<표 31>에서 보듯이, 말하기 수행평가의 시행한계성 각 문항 중 1번 문항 ($t=-1.450$, $p>.05$)을 제외한 다른 나머지 문항에서 두 집단 간의 평균점의 차이가 통계적으로 유의미한 것으로 나타났다(문항 2번: $t=-2.078$, $p<.05$; 문항 3번: $t=-2.322$, $p<.05$; 문항 4번: $t=-2.380$, $p<.05$; 문항 5번: $t=-4.578$, $p<.05$). 이 결과는 채점의 객관성 확보의 어려움, 적절한 말하기 수행평가 방법이나 도구의 부재, 말하기 수행평가와 관련된 도움 자료나 교사 연수의 부족 및 학생들에 대한 또 다른 학습 부담의 우려 등의 이유로 중학교 영어교사들보다 고등학교 영어교사들이 말하기 수행평가를 시행하는데 있어서 더 많은 어려움을 인식하고 있음을 보여준다. 비록 중학교에서도 말하기 수행평가의 시행이 쉽지는 않겠지만, 특히 우리나라의 전반적인 고등학교 교육과정이 입시 지향적인 관계로 진정한 말하기 수행평가의 시행을 둘러싼 현실적인 장애들을 무시한 채 영어 말하기 교육을 시행하기는 어렵

다(이문복 외, 2012)는 점을 고려해 볼 때, 오히려 고등학교 영어교사들이 느끼는 말하기 수행평가의 시행한계성이 당연히 더 높을 지도 모른다.

덧붙여, 말하기 수행평가의 시행한계성 각 문항이 중학교 영어교사들과 고등학교 영어교사들 각 집단 내에서 유의미한 차이가 있는지를 파악하고자 집단별 모평균에 대한 95% 신뢰구간을 추정하였다. <그림 11>과 <그림 12>는 학교별로 말하기 수행평가의 시행한계성의 개별 문항들에 대한 평균점 차이의 분포를 나타낸 결과이다.



<그림 11> 말하기 수행평가의 시행한계성 각 문항별 신뢰구간(중학교)

<그림 11>은 중학교 영어교사들에 대한 말하기 수행평가의 시행한계성 각 문항별 모평균의 상한값과 하한값의 분포를 상하 일직선으로 표시한 것이다. 이 다섯 문항 중 1번 문항인 학급 규모의 크기나 학생 수의 과다(신뢰구간: 3.607~3.933)는 3번 문항인 적절한 평가 방법이나 도구의 부재(신뢰구간: 3.125~3.455), 4번 문항인 도움이 되는 자료나 교사 연수의 부족(신뢰구간: 2.995~3.285) 및 5번 문항인 학생들에게 또 다른 학습 부담 제공(신뢰구간: 2.813~3.147)과 각각 서로 모평균의 95% 신뢰구간에서 뚜렷한 차이가 있는

것으로 나타났다. 또한, 2번 문항인 채점의 객관성 확보의 어려움(신뢰구간: 3.341~3.659)도 역시 4번 문항 및 5번 문항과 각각 서로 모평균의 95% 신뢰구간에서 뚜렷한 차이를 보였다. 즉, 중학교 영어교사들은 적절한 평가 방법이나 도구가 없거나, 도움이 되는 자료나 교사 연수가 부족하거나 혹은 학생들에게 추가적인 학습 부담이 될 수 있다는 우려로 인해 말하기 수행평가의 시행에 어려움을 인식하기보다는, 오히려 학급 규모가 크고 학생 수가 너무 많아서 진정한 말하기 수행평가의 시행이 더 어렵다고 인식하고 있는 것으로 나타났다. 뿐만 아니라, 중학교 영어교사들은 도움이 되는 자료나 교사 연수가 부족하거나 혹은 학생들에게 추가적인 학습 부담이 될 수 있다는 우려보다는, 오히려 말하기 수행평가의 채점 시 객관성 확보가 어렵기 때문에 말하기 수행평가를 시행하는 데 있어서 더 많은 한계성을 인식하고 있는 것으로 나타났다.

<그림 12>는 고등학교 영어교사들에 대한 말하기 수행평가의 시행한계성 각 문항별 모평균의 상한값과 하한값의 분포를 상하 일직선으로 표시한 것이다. 이 다섯 문항 중 1번 문항인 학급 규모의 크기나 학생 수의 과다(신뢰구간: 3.783~4.097)는 3번 문항인 적절한 평가 방법이나 도구의 부재(신뢰구간: 3.405~3.735), 4번 문항인 도움이 되는 자료나 교사 연수의 부족(신뢰구간: 3.243~3.577) 및 5번 문항인 학생들에게 또 다른 학습 부담 제공(신뢰구간: 3.359~3.661)과 각각 서로 모평균의 95% 신뢰구간에서 뚜렷한 차이가 있는 것으로 나타났다. 또한, 2번 문항인 채점의 객관성 확보의 어려움(신뢰구간: 3.578~3.923)도 역시 4번 문항과 서로 모평균의 95% 신뢰구간에서 뚜렷한 차이를 보였다. 즉, 고등학교 영어교사들도 역시 적절한 평가 방법이나 도구가 없거나, 도움이 되는 자료나 교사 연수가 부족해서, 혹은 학생들에게 추가적인 학습 부담이 될 수 있다는 우려로 인해서 말하기 수행평가의 시행을 어렵게 인식하기보다는, 오히려 학급 규모가 크고 학생 수가 너무 많아서 진정한 말하기 수행평가의 시행을 더 어렵게 인식하고 있는 것으로 나타났다.

다. 뿐만 아니라, 고등학교 영어교사들은 도움이 되는 자료나 교사 연수가 부족한 이유보다는, 오히려 말하기 수행평가의 채점 시 객관성 확보가 어렵다는 이유로 인해서 말하기 수행평가의 시행에 있어서 더 많은 한계성을 인식하고 있는 것으로 나타났다.



<그림 12> 말하기 수행평가의 시행한계성 각 문항별 신뢰구간(고등학교)

정리하면, 중학교 영어교사들과 고등학교 영어교사들 두 집단 모두 말하기 수행평가의 시행한계성에 대한 인식에 있어서 다른 이유들보다 학급 규모가 크고 학생 수가 너무 많거나 혹은 채점 시 객관성 확보가 어려워 말하기 수행평가를 시행하기가 어렵다는 인식이 더 높은 것으로 나타났다. 이러한 결과는 중·고등학교에서 말하기 수행평가의 시행의 한계점들에 대한 영어교사들의 인식 연구의 결과(고현우, 2014; 이소영, 2008; 이은주, 2003)와도 일치한다.

본 연구는 설문조사로 얻어진 연구 결과의 타당성을 보완하기 위해서 질적 연구의 한 방법인 자유서술식 설문 자료를 추가하였다. 보통 자유서술식 의견 조사는 설문 응답자의 익명성을 보장하기 때문에 자기 방어적인 성향이

낮아서 솔직한 답변을 할 가능성이 높아(Kearney et al., 1984) 설문 자료에서 얻어진 통계 결과의 정당성을 높여줄 수 있다. 이에 말하기 수행평가를 시행하는 중·고등학교 영어교사들의 자유서술식 문항에 대한 응답 내용을 중학교와 고등학교로 구분하여 범주화한 결과, 주로 말하기 수행평가의 시행한계성과 개선 방안에 관한 의견들로 구성되었다. 우선 중학교 영어교사들이 진술한 말하기 수행평가의 시행한계성에 관한 내용을 빈도순에 따라 정리한 결과는 <표 32>에 제시하였다.

<표 32> 자유서술식 응답에 나타난 말하기 수행평가의 시행상의 한계점들(중학교)

시행상의 한계점	빈도
· 채점 시 객관성 확보의 어려움	15
· 적절한 평가 방법이나 도구 및 기준의 부재	10
· 학급당 학생 수의 과다	7
· 원어민 영어교사의 부재	6
· 관련 연수나 유용한 자료의 부족	3
· 수업 진도로 인한 시간적 여유의 부족	3
· 영어교사의 전문성 부족	2
· 동료 교사들 간의 교류나 협력의 부족	2
· 업무 과다로 인한 시간적 여유의 부족	2
· 물리적·환경적 여건의 미비	1
· 말하기에 대한 동기 부여의 어려움	1
· 사회·경제적 능력에 따른 학생의 영어능력 차이의 심화	1
· 입시 위주의 교육과정과의 연계성 부족	1
총 빈도 수	54

<표 32>에서 보듯이, 말하기 수행평가를 시행함에 있어서 현재 우리나라의 중학교 영어교사들이 언급한 가장 큰 한계점은 채점 시 객관성 확보의 어려움이었다. 이는 중학교 영어교사들의 말하기 수행평가의 시행한계성에 대한 각 문항별 모평균의 차이를 비교한 결과와 연결된다. 즉, 다른 몇 가지

이유보다는 채점 시 객관성 확보가 어렵다는 점이 중학교 영어교사들의 말하기 수행평가의 시행에 상대적으로 더 큰 한계점으로 작용한다는 점을 시사하고 있다. 이 범주에 속한 응답 내용을 일부 제시하면 다음과 같다.

말하기 평가는 객관성 확보가 가장 문제가 된다고 생각합니다. 채점자 개인의 판단이 포함되는 부분이기 때문에 누가 채점하느냐에 따라 채점결과가 달라질 수 있는 여지가 다분합니다. 하지만 현실적으로는 여러 교사의 공동 채점이 매우 어렵기 때문에 모든 학생들을 같은 조건과 같은 환경에서 평가하기가 너무 힘듭니다. (교사 A)

평가 결과의 객관성 및 공정성을 최우선 과제로 하는 현재의 평가 환경에서 채점자의 주관적인 판단이 개입될 가능성이 높은 말하기 평가를 시행하기가 매우 어렵습니다. (교사 B)

말하기 평가 채점의 기준을 명확하게 정하기 어려운 것이 점수의 공정성을 확보하기 매우 힘들게 만든다. 말하기 내용을 우위에 둘 것이냐 유창성 및 정확성을 우위에 둘 것이냐 하는 문제도 있다. 그리고 평가 후 적절한 피드백이 이루어진다면 학습에 많은 도움이 될 것이라 생각한다. (교사 C)

평가자의 기준과 감정에 따라 점수가 달라질 우려가 많아 부담감이 크다. 학급마다 평가자가 다르니까 확실한 기준을 정할 수 없으니 불공평할 우려가 있다. (교사 D)

중학교 영어교사들이 두 번째로 지적한 말하기 수행평가의 시행상의 어려운 점은 적절한 평가 방법이나 도구 및 기준의 부재로 나타났다. 이는 첫 번째의 문제점으로 언급된 채점 시 객관성 확보의 어려움과 직접적인 관련성이 높을 것이다. 즉, 아무래도 말하기 수행평가를 공정하게 할 수 있는 기준이나 도구가 없다 보니 채점을 할 때도 자연스럽게 채점자의 임의적인 주관

성이 다소 개입될 여지가 있기 때문이라 여겨진다. 이 범주에 포함되는 응답 내용을 일부 제시하면 다음과 같다.

진정한 말하기 수행평가가 현실상 이루어지지 않고 있는 실정입니다. 말하기 수행평가 기준과 방법 모두 담당 교사들이 직접 고안해서 해야 하는 부담도 있습니다. 아이들의 수준에 맞는 진정한 평가가 아니라 단순한 문장 암기식 평가가 진정 학생들의 말하기 수행평가인지 의문이 앞섭니다. 말하기 수행평가의 연구가 더욱 필요하다고 생각합니다. (교사 E)

원어민 교사가 말하기 평가를 하든, 영어교사가 말하기 평가를 하든 평가자의 정확한 말하기 평가에 대한 척도를 세워놓지 않으면 객관 타당한 평가를 하기 힘든 것이 말하기 평가라고 생각함. 필기고사와 같이 정확한 정답이 있지 않기 때문에 평가자의 주관적 평가 태도를 받아들일 수밖에 없는 상황임. (교사 F)

학교별 학생의 수준에 따른 성취기준 작성과 성취기준에 따른 수행평가 문항 작성이 어려운 것이 현실이다. 말하기 수행평가에 대한 운영 매뉴얼과 수준별 성취기준 및 평가문항의 개발이 절실히 필요하다. (교사 C)

말하기를 수행평가에 반영하는 일이 필요하다고 생각은 하나, 학생의 환경이나 흥미, 수행 능력 등을 파악해서 평가 기준과 도구를 개발하기가 현 교육적 환경으로는 무척 힘들고, 기존에 개발되어 있는 것이 없어 이용할 수도 없고 단지 얼마나 영어 표현을 외워서 하느냐의 능력만 묻는 경우가 많아서 학생의 실제 의사소통능력을 객관적으로 판단하고 채점하는 데는 한계가 있다는 아쉬움이 있음. (교사 G)

세 번째로 중학교 영어교사들이 언급한 말하기 수행평가의 시행을 힘들게 하는 원인으로는 학급당 학생 수의 과다라는 한계점인데, 이도 마찬가지로 중학교 영어교사들의 말하기 수행평가의 시행한계성에 대한 각 문항별 모평균의 차이를 비교한 결과와 결부된다. 즉, 중학교 영어교사들의 말하기 수행

평가의 시행을 어렵게 만드는 몇 가지 이유들 중 학급 규모가 크고 학생 수가 너무 많다는 점이 상대적으로 더 큰 이유임을 시사한다. 더구나 이러한 이유는 앞서 진술된 두 가지 한계점들과 높은 연관성이 있다. 다시 말해, 영어교사가 혼자서 많은 학생들을 대상으로 말하기 수행평가를 시행하다 보면 채점을 할 때 객관성을 유지하기가 힘들 뿐더러, 또한 다양한 평가 방법을 활용하지 못하고 일률적인 평가 기준을 적용하기가 쉬워 타당한 말하기 수행평가를 시행하기가 어려워질 가능성이 높다. 이를 보여주는 몇몇 응답 내용을 제시하면 다음과 같다.

학급당 학생 수가 많고 한 선생님마다 담당해야 하는 학급 수가 많아서 모든 학생의 authentic한 말하기 수준을 평가하기란 거의 불가능하고, 단지 몇 개의 문장을 주고 그것을 암기해서 말할 수 있냐에 시험이 그치는 것 같습니다. (교사 H)

학급 사이즈가 너무 크다보니 주어진 수업 시간에 모든 학생들을 다 상대하기가 현실적으로 힘들다. 수업 진도도 신경이 쓰이고 대충 대화문 외워서 발표하는 식의 말하기 수행평가를 실시하고 있다. 학급마다 거의 비슷하게 평균을 맞추기 위한 상대평가로 치우치는 경향이 있어 올바른 평가가 될 수 없고 다른 교사들과 함께 평가하지 못해 임의로 점수를 줄 가능성이 생기는 것 같다. (교사 D)

평가대상자가 다수이므로 채점의 객관성 확보가 힘들. (교사 F)

학급당 학생 인원수가 15명 이내인 경우 말하기 평가가 원활하다고 생각합니다. (교사 I)

이와 같은 내용 외에도 원어민 영어교사의 부재와 관련 연수나 유용한 자료의 부족 및 수업 진도로 인한 시간적 여유의 부족 등 말하기 수행평가의 시행한계성에 영향을 주는 여러 가지 다양한 의견들이 보고되었다. 이처럼

학교 현장에서 중학교 영어교사들이 말하기 수행평가를 시행하는 데 있어서 많은 어려움을 겪고 있음을 알 수 있다.

아울러 말하기 수행평가의 시행한계성에 대해서 중학교에서 제기된 내용을 고등학교에서 제기된 내용과 서로 비교해 보고자 한다. 고등학교 영어교사들이 진술한 말하기 수행평가의 시행한계성에 관한 의견들을 빈도순에 따라 정리하였고, 그 결과는 <표 33>에 제시하였다.

<표 33> 자유서술식 응답에 나타난 말하기 수행평가의 시행상의 한계점들(고등학교)

시행상의 한계점	빈도
· 채점 시 객관성 확보의 어려움	11
· 입시 위주의 교육과정과의 연계성 부족	9
· 적절한 평가 방법이나 도구 및 기준의 부재	7
· 학급당 학생 수의 과다	5
· 학생의 학업 부담 가중	4
· 사회·경제적 능력에 따른 학생의 영어능력 차이의 심화	2
· 물리적·환경적 여건의 미비	2
· 수업 진도로 인한 시간적 여유의 부족	2
· 말하기에 대한 동기 부여의 어려움	2
· 관련 연수나 유용한 자료의 부족	1
· 업무 과다로 인한 시간적 여유의 부족	1
총 빈도 수	46

<표 33>에서 보듯이, 말하기 수행평가를 시행함에 있어서 현재 우리나라의 고등학교 영어교사들이 언급한 가장 큰 한계점도 중학교와 마찬가지로 채점 시 객관성 확보의 어려움인 것으로 나타났다. 이는 고등학교 영어교사들의 말하기 수행평가의 시행한계성에 대한 각 문항별 모평균의 차이를 비교한 결과와 연관된다. 즉, 다른 이유보다는 채점 시 객관성 확보가 어렵다는 점이 고등학교 영어교사들의 말하기 수행평가의 시행에 상대적으로 더

큰 한계점으로 작용한다는 점을 시사하고 있다. 이 범주에 속한 응답 내용을 일부 제시하면 다음과 같다.

현 수능체제의 영어교육에서 말하기 평가는 현실적으로 시행하기가 힘들다. 여러 교사가 같이 평가를 하고 있어 합의된 평가 방법을 끌어내기가 쉽지 않다. (교사 A)

채점자간 편차(신뢰도 확보): 충분한 사전 협의를 거친다 하더라도 채점자간 편차 발생. 객관성 확보: 명확한 기준을 수립하더라도 학생에 대한 선입견이 채점 시 영향을 미칠 수 있음. (교사 B)

말하기 평가 자체가 워낙 주관적인 요소가 많이 작용하는 거라 채점하는 데 어려움이 많을 수밖에 없습니다. (교사 C)

평가기준이 명확하게 확립되지 않아 채점상의 어려움이 존재하며, 학생의 직접적인 영어 말하기 능력보다는 암기식 말하기 평가여서 다소 부족하다. (교사 D)

고등학교 영어교사들이 두 번째로 자주 제기한 말하기 수행평가의 시행상의 어려운 점은 입시 위주의 교육과정과의 연계성 부족으로 나타났는데, 특히 이는 말하기 수행평가의 시행한계점으로 단 하나의 사례만 언급되었던 중학교 영어교사들의 경우와 상반되고 있다. 즉, 고등학교 교육과정에서도 효용성의 측면에서 영어 말하기 교육이 중요함을 강조하고 있지만, 실제로 대학수학능력시험에서는 학생들의 영어 능력의 측정 기능들 중 듣기와 읽기만 거의 반영하고 있기 때문일 것이다. 따라서 대부분의 고등학교에서는 이러한 현실에 부응하여 학생들의 영어 말하기 능력을 향상시키고자 굳이 말하기 수행평가를 시행할 필요성을 그다지 느끼지 못하고 있는 상황이라 여겨진다. 이 범주에 포함되는 응답 내용을 일부 제시하면 다음과 같다.

말하기 수행평가를 통하여 학생들의 의사소통 능력을 확인할 수 있다고 생각 하지만, 고등학교에서는 입시와 다소 무관한 영역이라 말하기 수행평가를 시행하는 것이 망설여집니다. (교사 E)

수능에서 말하기 직접 평가 영역이 없다면 말하기 수행평가의 정상적인 실행은 어려울 것 같다. (교사 A)

현재 수능영어 체제의 말하기 수행평가는 요식적일 수밖에 없다. 읽기, 쓰기, 말하기, 듣기의 네 가지 능력을 측정하는 언어시험일 때는 실제적인 말하기 수행평가가 가능하다고 생각된다. (교사 F)

말하기가 의사소통 능력 향상에 중요하나 인문계 고교의 특성상 대입에 말하기가 평가영역으로 도입되지 않는 한 학교에서 평가의 빈도나 양을 늘리게 되는 데는 한계가 있는 것으로 여겨진다. 대입에 활용되는 방안이 말하기 수업 활성화에 기여하게 될 것으로 보인다. (교사 G)

세 번째로 고등학교 영어교사들의 말하기 수행평가의 시행을 힘들게 하는 원인으로서는 적절한 평가 방법이나 도구 및 기준의 부재라는 한계점이었다. 이도 역시 첫 번째의 한계점으로 언급된 채점 시 객관성 확보의 어려움과 높은 상관성을 지니고 있다. 즉, 타당한 말하기 수행평가의 방법이나 도구 및 기준이 없이는 당연히 말하기 수행평가의 채점도 객관적으로 이루어지기 힘들기 때문에 이 문제 또한 말하기 수행평가를 시행하는 데 있어서 큰 걸림돌이 되는 것으로 여겨진다. 이 범주에 포함되는 응답 내용을 일부 제시하면 다음과 같다.

국가적 차원에서 도구 필요하다고 생각함. 세분화된 학생들의 능력 수준을 토대로 말하기 교육과정과 말하기 평가가 연계되어야 하지만, 현재 교육 현실은 평가를 위한 평가인 상황이 많다. NEAT 등의 평가 도구도 중요하지만, 우선 학교단위 수행평가를 실행하라고 하기 전에 체계적인 교육과정과 평가기준이 정립되는 것이 필요함. (교사 H)

현실적으로 말하기 수행 능력을 정확히 판별할 수 있는 평가 도구 마련이 쉽지 않다. (교사 A)

형식적인 말하기 평가기준이 아닌 한국의 실제적인 학급 규모(40명)에서 실행할 수 있는 평가 도구와 자료들의 개발이 필수적이다. (교사 I)

말하기 수행평가의 경우 교과서 내용과 관련된 주제를 통해서 실시하고 있으나 대부분 개인적인 발표 중심이 되는 경향이 있다. 그룹이나 2명 이상의 말하기 평가 도구를, 그것도 실용적이며 관심과 흥미가 있는 topic으로 실시하는 평가 도구를 개발할 필요가 있다. (교사 J)

또한, 학급당 학생 수의 과다라는 한계점이 부각되었는데, 이는 고등학교 영어교사들의 말하기 수행평가의 시행한계성에 대한 각 문항별 모평균의 차이를 비교한 결과와 직접 연결된다. 그리고 앞서 언급된 채점 시 객관성 확보가 어렵고 적절한 평가 방법이나 도구 및 기준이 없어 말하기 수행평가를 시행하기가 어려운 이유와도 연관성이 있어 보인다. 이 외에도 학생의 학업 부담 가중을 비롯한 많은 다양한 의견들이 말하기 수행평가의 시행한계성에 영향을 주는 것으로 보고되었다. 이처럼 학교 현장에서 고등학교 영어교사들도 말하기 수행평가를 시행하는 데 있어서 많은 어려움을 겪고 있음을 알 수 있다.

참고로, <표 34>는 중학교 영어교사들과 고등학교 영어교사들이 각각 언급한 말하기 수행평가의 시행한계점들 중에서 빈도순에 따라 서로 공통적인 부분을 우선순위로 정리한 것이다.

<표 34> 중·고등학교 말하기 수행평가의 시행상의 한계점들 순위 비교

중·고등학교의 공통적인 시행상의 한계점	중학교 순위	고등학교 순위
· 채점 시 객관성 확보의 어려움	1	1
· 적절한 평가 방법이나 도구 및 기준의 부재	2	3
· 학급당 학생 수의 과다	3	4
· 관련 연수나 유용한 자료의 부족	4	6
· 수업 진도로 인한 시간적 여유의 부족	4	5
· 업무 과다로 인한 시간적 여유의 부족	5	6
· 물리적·환경적 여건의 미비	6	5
· 말하기에 대한 동기 부여의 어려움	6	5
· 사회·경제적 능력에 따른 학생의 영어능력 차이의 심화	6	5
· 입시 위주의 교육과정과의 연계성 부족	6	2

<표 34>에서 보듯이, 고등학교 영어교사들이 지적한 입시 위주의 교육과정과의 연계성 부족을 제외하고, 채점 시 객관성 확보의 어려움(중학교 1위, 고등학교 1위), 적절한 평가 방법이나 도구 및 기준의 부재(중학교 2위, 고등학교 3위), 그리고 학급당 학생 수의 과다(중학교 3위, 고등학교 4위) 등이 말하기 수행평가를 시행할 때 중학교 영어교사들과 고등학교 영어교사들이 공통적으로 인식하고 있는 상위 세 가지의 한계점들이었으므로 나타났다. 이 결과는 중·고등학교에서 운영되고 있는 수행평가의 문제점들을 개방형 설문을 통해서 조사한 이소영(2008)의 결과와도 거의 일치한다. 따라서 이러한 한계점들은 현재 우리나라의 중·고등학교 영어교사들 모두가 말하기 수행평가를 제대로 시행하기 위해서 가장 우선적으로 해결되어야 할 과제임을 시사하고 있다.

5.1.4 학교급간 말하기 수행평가의 시행한계성에 대한 개선 방안

본 연구의 자유서술식 설문에 대한 응답 내용에서 조사된 말하기 수행평가의 시행한계점들에 대해서 중학교와 고등학교의 영어교사들이 제안한 개선 방안을 살펴보고자 한다. 먼저 중학교 영어교사들이 제안한 개선 방안들을 빈도순에 따라 정리하였고, 그 결과는 <표 35>에 제시하였다.

<표 35> 말하기 수행평가의 시행한계성에 대한 개선 방안들(중학교)

개선 방안	빈도
· 객관적인 평가 방법과 도구 및 기준의 개발	12
· 다양한 교사 연수의 마련	7
· 원어민 영어교사의 필요	5
· 학급당 학생 수의 축소	3
· 실질적인 말하기 중심의 수업 및 평가	2
· 교과서 내용과 연계한 말하기 평가 방법 개발	1
· 말하기 능력 향상을 위한 평가 체제의 변환	1
총 빈도 수	31

<표 35>에서 보듯이, 말하기 수행평가의 시행한계성을 극복하기 위해서 중학교 영어교사들이 제안한 최우선의 개선 방안은 객관적인 평가 방법과 도구 및 기준의 개발이었다. 이는 말하기 수행평가 시행에 있어서 중학교 영어교사들이 인식하고 있는 결정적인 한계점인 채점 시 객관성 확보의 어려움과 적절한 평가 방법이나 도구 및 기준의 부재와 직접적으로 연결된다. 즉, 말하기 수행평가를 가장 타당하게 시행하는 일은 객관적인 평가 방법이나 도구를 가지고 객관적인 평가 기준에 맞춰 공정하게 채점이 이루어질 때만 가능하기 때문이다. 이 범주에 속한 응답 내용을 일부 제시하면 다음과 같다.

진정한 말하기 평가가 되기 위해서는 선생님들과의 의견 조율 및 좋은 평가 도구 개발, 채점 기준 등이 잘 만들어져야 될 것 같다. (교사 A)

교육청에서 말하기 수행평가에 대한 적절한 수준별 평가 도구를 학교급별로 개발해서 보급해주면 좋겠음. (교사 B)

말하기 영역의 평가는 교사의 전문성에 따라 크게 영향을 받으므로 객관적인 평가 도구가 개발·제공되어야 하며, 실질적인 말하기 능력 신장을 위한 수업이 이루어질 때 수행평가로서 유의미하리라 생각합니다. (교사 C)

말하기 수행평가의 평가의 공정성 및 객관성 확보를 위해서 교육청 차원에서 원 어민을 대체할만한 적절한 평가 도구를 개발해서 배포해 주면 좋겠다. (교사 D)

중학교 영어교사들이 두 번째로 많이 제시한 말하기 수행평가의 시행한계성에 대한 개선 방안은 다양한 교사 연수의 마련으로 나타났다. 이는 일부 중학교 영어교사들이 언급한 관련 연수나 유용한 자료의 부족이라는 한계점과 연결된다. 즉, 영어교사의 말하기 수행평가에 대한 전문성을 높이기 위해서는 각 시·도 교육청이나 정부 차원에서 말하기 수행평가의 객관적인 평가 기준과 평가 도구를 개발해서 다양한 연수나 프로그램을 제공하는 일이 필요하다는 의견이다. 이 범주에 해당되는 응답 내용을 일부 언급하면 다음과 같다.

한국인 영어교사가 말하기 평가를 하기 위해서는 원어민 수준의 말하기 능력을 가지도록 부단히 연수를 해야 할 필요가 있음. (교사 E)

객관적인 criteria의 구축과 그에 맞는 교사 연수를 확대하여야 한다. (교사 D)

말하기 수행평가를 시행하기에 한국인 영어교사의 역량이 부족하고 객관성을 유지하기가 쉽지 않다. 객관적 평가를 할 수 있는 연수 프로그램이나 시스템 등이 있으면 가능할 듯하다. (교사 F)

세 번째로 중학교 영어교사들이 제시한 말하기 수행평가의 시행한계성에 대한 개선 방안으로는 원어민 영어교사의 필요성이었다. 이는 말하기 수행평가를 시행하는 데 있어서 일부 중학교 영어교사들이 제기한 원어민 영어교사의 부재라는 문제와 직접적인 연관이 있다. 즉, 많은 중학교에서 영어회화 수업을 담당하던 원어민 영어교사의 수를 정책적으로 대폭 축소함으로써 그들의 역할을 한국인 영어교사가 맡게 되었고, 결국 원어민 영어교사가 있을 때보다 말하기 수행평가의 신뢰성과 객관성 확보가 어렵게 되었다는 불평이 생겨난 것으로 보인다. 이를 보여주는 몇몇 응답 내용을 제시하면 다음과 같다.

원어민 영어교사를 활용한 말하기 수행평가는 학생들에게 적절한 동기부여와 평가의 신뢰성 측면에서 많은 도움이 된다고 생각함. (교사 E)

원어민이 학생 전체의 말하기 평가 채점을 해서 공평한 평가가 될 수 있었던 과거와 달리, 지금은 채점의 공평성에 대한 의논이나 의견이 다양하게 제시되고 있습니다. (교사 G)

원어민이 평가를 하면 학생들도 평가 결과에 큰 불만이 없을 듯합니다. (교사 H)

이러한 의견들 외에도 학급당 학생 수의 감축이나 실질적인 말하기 중심의 수업 및 평가의 필요성 등 말하기 수행평가의 시행을 어렵게 하는 요인들을 극복할 수 있는 다양한 방안들이 제시되었다. 또한, 말하기 수행평가의 개선 방안으로서 늘 제안되어 왔던 국가 수준의 영어 평가에 말하기를 포함해야 한다는 의견도 있었다.

아울러 말하기 수행평가의 시행한계성에 대해서 중학교 영어교사들이 제안한 개선 방안들을 고등학교 영어교사들이 제안한 개선 방안들과 서로 비교하고자 한다. <표 36>은 고등학교 영어교사들이 말하기 수행평가의 시행한계성을 극복하기 위해서 제시한 방안들을 빈도순에 따라 정리한 결과이다.

<표 36> 말하기 수행평가의 시행한계성에 대한 개선 방안들(고등학교)

개선 방안	빈도
· 객관적인 평가 방법과 도구 및 기준의 개발	9
· 실질적인 말하기 중심의 수업 및 평가	6
· 원어민 영어교사의 필요	5
· 교과서 내용과 연계한 말하기 평가 방법 개발	3
· 다양한 교사 연수의 마련	1
· 말하기 능력 향상을 위한 평가 체제의 변환	1
· 말하기 평가 횟수의 증가	1
총 빈도 수	26

<표 36>에서 보듯이, 말하기 수행평가의 시행한계성을 해결하기 위해서 고등학교 영어교사들이 제안한 최우선적인 개선 방안으로서 객관적인 평가 방법과 도구 및 기준의 개발이 제안되었다. 이는 중학교 영어교사들이 제안한 가장 우선적인 개선 방안과도 일치하는데, 말하기 수행평가의 시행에 있어서 고등학교 영어교사들이 인식하고 있는 주요한 한계점인 채점 시 객관성 확보의 어려움과 적절한 평가 방법이나 도구 및 기준의 부재와 직접적으로 연결된다. 즉, 말하기를 수행평가에 반영할 것을 권장하는 교육 정책에 따라 말하기 수행평가를 시행하고는 있지만, 실제로 학교 현장의 많은 고등학교 영어교사들은 구체적이고 세부적인 기준과 그에 맞게 활용할 만한 마땅한 평가 방법이나 도구가 없어서 이 방안을 가장 시급한 과제로 간주하고 있는 것으로 보인다. 이 범주에 해당하는 몇몇 응답 내용을 제시하면 다음과 같다.

서로 다른 평가자가 평가하므로 객관적인 말하기 수행평가 척도가 있어야 함. (교사 A)

실제로 말하기 평가를 할 때 해마다 손쉽고 편하게 할 수 있는 방법을 모색합니다. 보다 객관적으로 말하기 평가를 할 수 있는 평가 방법과 채점 기준을 마련하는 것이 가장 중요하다고 생각합니다. (교사 B)

말하기 수행평가는 객관성 확보가 어려운 것이 가장 문제점이라고 봅니다. 만일 객관성이 보장된 도구만 개발된다면(힘들겠지만ㅠ.ㅠ) 적극적으로 실시할 자신이 있습니다. (교사 C)

고등학교 영어교사들이 두 번째로 많이 제시한 말하기 수행평가의 시행한 계성을 해결하기 위한 개선 방안은 실질적인 말하기 중심의 수업 및 평가의 필요성으로 나타났다. 이는 중학교와는 사뭇 다르게 고등학교에서만 지적된 말하기 수행평가 시행의 어려움으로서, 아마도 입시 위주의 교육과정과의 연계성 부족이라는 한계점과 밀접한 관련이 있을 것이다. 대학입학시험에 독해와 듣기 위주의 평가만 반영되고 말하기 평가가 빠져있는 한, 일선 고등학교의 수업 방식도 그러한 방향에 맞춰서 움직일 수밖에 없는 현실이다. 그러므로 말하기 수행평가의 효용성을 인정하고 있고 또한 그 효용성의 가치를 진정 실천하고자 한다면, 고등학교에서도 대학입시의 준비보다는 말하기 중심의 수업이 선행되어야 함을 역설하는 것으로 보인다. 이 범주에 속한 응답 내용을 일부 제시하면 다음과 같다.

영어 말하기가 진정으로 우리 학생들의 말하기에 도움이 될 수 있는 현 실생활의 말하기 위주의 수업과 수행평가가 이루어져야 되지 않을까 하는 생각이 듭니다. (교사 D)

분명 대화문을 작성하여 role play 해 보는 등의 speaking test가 영어에 대한 흥미와 실질적 언어사용에 대한 연습으로서 최소한의 긍정적 역할을 하는 것 같습니다. Reading이 주를 이루는 입시 위주의 영어 학습을 하는 고2, 특히 고3 학생들에게는 “말하기 수행평가가 과연 수능위주의 학습에 도움이 되나, 연계가 될까?”라는 의문을 남기는 것이 현실인 것 같습니다. 고학년이 될 수록 교사나 학습자 모두 필요성을 체감할 수 있는 의미 있는 수행평가가 이루어진다면 더욱 좋겠습니다. (교사 E)

딱딱한 영어독해보다는 흥미와 관심을 불러오는 말하기 수업이 정말 필요한 것 같아요. 그러게 되면 학생들의 말하기 능력도 많이 향상되지 않겠어요? (교사 F)

세 번째로 말하기 수행평가의 시행한계성에 대한 개선 방안으로서 중학교 영어교사들처럼 고등학교 영어교사들도 똑같이 원어민 영어교사의 필요성을 주장하였다. 이는 고등학교 영어교사들의 말하기 수행평가의 시행을 힘들게 하는 이유로서 직접적으로 언급되지는 않았지만, 아마도 채점 시 객관성 확보의 어려움이나 혹은 적절한 평가 방법이나 도구 및 기준의 부재라는 이유와 어느 정도 관련성이 있을 것으로 보인다. 왜냐하면 말하기 수행평가의 시행에서 가장 중요한 문제가 객관성과 신뢰성의 확보인데, 아무래도 영어를 모국어로 사용하는 원어민 영어교사들이 말하기 수행평가를 시행하게 되면 한국인 영어교사들보다는 채점 기준이나 방법적인 측면에서 보다 더 객관적이어서 믿음이 간다고 생각하는 경향이 있기 때문일 것이다. 이 범주에 해당하는 응답 내용을 일부 제시하면 다음과 같다.

교사들의 채점 기준에 대한 숙지가 부족하니 원어민 교사가 있다면 협력을 통해서 온전한 평가가 이루어지고, 학생들 또한 점수에 대한 신뢰성을 가질 것이라 봅니다. (교사 B)

원어민을 활용하여 일정 기간에 다른 학교나 교육청에 근무하시는 분이 우리 학교 등 원어민이 없는 학교를 방문하여 말하기 시험을 실시해주고 가시는 방안을 건의합니다. (교사 G)

교사 입장에서도 원어민 영어교사가 직접 하는 것이 더 신뢰가 가긴 함. (교사 H)

이러한 의견들 외에도 교과서 내용과 연계한 말하기 평가 방법의 개발, 다양한 교사 연수의 마련 및 말하기 평가 횟수의 증가 등 현 말하기 수행평가

가 당연한 한계점들을 극복할 수 있는 다양한 방안들이 제시되었다. 그리고 중학교 영어교사들처럼 고등학교 영어교사들도 국가 수준에서 시행되는 말하기 평가 체제로의 변환을 강조하기도 하였다.

참고로, <표 37>은 중학교 영어교사들과 고등학교 영어교사들이 각각 언급한 말하기 수행평가의 시행한계성을 해결하기 위한 개선 방안들 중에서 빈도순에 따라 서로 공통적인 부분을 우선순위로 정리한 것이다. 표에서 보듯이, 고등학교 영어교사들이 지적한 다양한 교사 연수의 마련은 제외하더라도, 객관적인 평가 방법과 도구 및 기준의 개발(중학교 1위, 고등학교 1위)과 원어민 영어교사의 필요(중학교 3위, 고등학교 3위)는 진정한 말하기 수행평가의 시행을 위해서 우선적으로 실천해야 할 중요한 두 가지 방안으로서 중학교 영어교사들과 고등학교 영어교사들이 공통적으로 주장하는 내용이였다. 박순경 외(2003)는 학교 현장의 교사들이 수행평가에 참고할 만한 문항이나 채점 기준에 대한 범례가 거의 없다고 지적하였는데, 이는 객관적인 평가 방법과 도구 및 기준의 개발에 대한 필요성과 연결된다. 또한, 이러한 객관적인 평가 방법과 도구 및 기준을 가장 우선적으로 개발해야 한다는 주장은 기존의 선행 연구들(김정임, 2006; 염지연, 2011; 이소영, 2008)과도 일치한다.

<표 37> 중·고등학교 말하기 수행평가의 시행한계성에 대한 개선 방안들 순위 비교

중·고등학교의 공통적인 개선 방안	중학교 순위	고등학교 순위
· 객관적인 평가 방법과 도구 및 기준의 개발	1	1
· 다양한 교사 연수의 마련	2	5
· 원어민 영어교사의 필요	3	3
· 실질적인 말하기 중심의 수업 및 평가	4	2
· 교과서 내용과 연계한 말하기 평가 방법 개발	5	4
· 말하기 능력 향상을 위한 평가 체제의 변환	5	5

한편, 말하기 수행평가의 바람직한 시행을 위해서 중·고등학교 영어교사들이 언급한 원어민 영어교사의 필요성은 본 연구에서 새롭게 발견된 주요 개선 방안으로서, 원어민 영어교사보다는 한국인 영어교사가 시행하는 말하기 수행평가를 과연 얼마나 신뢰할 수 있을 것인가에 관한 평가의 객관성과 연관이 있는 것으로 보인다. 사실 이러한 의견은 현재 말하기 수행평가에 대한 객관적이고 타당한 평가 도구나 기준이 부족함을 시사하고 있음과 동시에 원어민 영어교사에 못지않도록 한국인 영어교사들의 말하기 수행평가에 대한 전문성 또한 보다 신장시킬 필요가 있음을 암시한다.

정리하면, 말하기 수행평가의 시행에 있어서 주된 한계점은 한 학급당 학생 수의 과다로 인한 채점 시 객관성 확보의 어려움과 적절한 평가 방법이나 도구 및 기준이 없다는 것이었다. 그리고 우리나라의 중·고등학교에서 말하기 수행평가의 시행이 정착되기 위해서는 무엇보다도 객관적인 평가 방법과 도구 및 기준을 개발하여 각 학교에 보급하는 일이 가장 시급한 문제로 드러났다.

5.1.5 요약

이상에서 살펴본 것처럼 말하기 수행평가를 시행하고 있는 중학교 영어교사들과 고등학교 영어교사들 사이에 말하기 수행평가의 전문성, 전문성의 세 하위 요인들, 효용성 및 시행한계성 등 여섯 가지 요인들에 대한 그들의 인식 차이를 분석한 결과를 요약하자면 다음과 같다. 첫째, 말하기 수행평가의 전문성 요인 전체와 전문성을 구성하는 세 하위 요인들, 그리고 전문성을 구성하는 12개의 각 문항에 대해서 두 집단 간의 차이를 비교하고 분석하였다. 그 결과 전문성 전체와 세 하위 요인들에서 고등학교 영어교사들보다 중

학교 영어교사들의 평균점이 더 높았고 두 집단 간에 통계적으로 유의미한 차이가 있는 것으로 나타났다. 또한, 말하기 수행평가의 전문성의 하위 요인 별 각 문항을 서로 비교한 결과, 중학교 영어교사들은 학생의 성취 수준에 적합한 평가 방법과 도구의 선택, 평가의 실시 전 학생들에게 전반적인 평가 계획의 공지 및 평가의 목적과 내용에 근거한 평가 결과의 해석 등과 같은 전문적인 능력에 대한 인식이 상대적으로 더 높은 것으로 밝혀졌다. 그리고 고등학교 영어교사들은 학생의 성취 수준에 적합한 평가 방법과 도구의 선택과 평가의 목적과 내용에 근거한 평가 결과의 해석에 대한 능력 등과 같은 전문적인 능력에 대한 인식이 상대적으로 더 높은 것으로 밝혀졌다.

둘째, 말하기 수행평가의 효용성 요인 전체와 효용성을 구성하는 5개의 각 문항에 대해서 두 집단 간의 차이를 비교하고 분석하였다. 그 결과 말하기 수행평가의 전반적인 효용성에 대한 인식에서 고등학교 영어교사들보다 중학교 영어교사들의 평균점이 모두 높았으나, 3번 문항(학생의 영어 말하기 성취 수준의 파악에 도움이 됨)을 제외하고는 두 집단 간에 통계적으로 유의미한 차이는 없는 것으로 나타났다. 또한, 말하기 수행평가의 효용성 각 문항을 서로 비교한 결과, 중학교 영어교사들과 고등학교 영어교사들 두 집단 모두 말하기 수행평가를 통해서 학생들의 영어 말하기 성취 수준을 파악하는 데 도움이 된다는 인식이 말하기 수행평가의 효용성에 대한 다른 인식들보다 상대적으로 높은 것으로 밝혀졌다.

셋째, 말하기 수행평가의 시행한계성 요인 전체와 시행한계성을 구성하는 5개의 각 문항에 대해서 두 집단 간의 차이를 비교하고 분석하였다. 그 결과 말하기 수행평가의 전반적인 시행한계성에 대한 인식에서 중학교 영어교사들보다 고등학교 영어교사들의 평균점이 모두 높았고, 1번 문항(학급 규모가 크거나 학생 수가 많아 시행하기가 어려움)을 제외하고는 두 집단 간에 통계적으로도 유의미한 차이가 있는 것으로 나타났다. 또한, 말하기 수행평가의

시행한계성 각 문항을 서로 비교한 결과, 말하기 수행평가의 시행을 어렵게 하는 원인들 중에서 중학교와 고등학교 영어교사들 두 집단 모두 다른 시행상의 한계점들보다 특히 학급 규모가 너무 크고 학생 수가 너무 많다거나 혹은 수행평가의 채점 시 객관성 확보가 어렵다는 점을 상대적으로 더 높게 인식하는 것으로 밝혀졌다.

마지막으로, 말하기 수행평가의 시행에 대한 자유서술식 응답의 내용을 살펴본 결과, 중학교와 고등학교 영어교사들 두 집단 모두 말하기 수행평가의 필요성과 효용성을 인지하고 있음에도 불구하고 말하기 수행평가의 채점 시 객관성 확보가 어렵다는 점을 가장 큰 한계점으로 보고하였다. 이러한 말하기 수행평가의 시행한계성을 극복하기 위해서 중학교와 고등학교 영어교사들이 공통적으로 제안한 최우선적인 개선 방안은 객관적인 평가 방법과 도구 및 기준의 개발이었다.

이러한 연구 결과들을 종합해 보면, 말하기 수행평가의 시행과 관련된 모든 단계에서 중학교 영어교사들의 전문성에 대한 인식이 고등학교 영어교사들의 전문성에 대한 인식보다 상대적으로 더 뛰어났다. 그리고 말하기 수행평가를 시행함으로써 얻게 되는 교육적인 효용성의 측면에 있어서는 중학교 영어교사들과 고등학교 영어교사들 두 집단 모두 대체적으로 긍정적인 인식을 가지고 있었다. 또한, 중학교 영어교사들보다 고등학교 영어교사들이 말하기 수행평가를 시행하는 데 있어서 여러 가지 한계점들로 인해서 더 많은 어려움을 인식하고 있었다. 아울러 말하기 수행평가의 시행을 어렵게 하는 주된 이유로서 중학교와 고등학교 영어교사들 두 집단 모두 학급 규모가 너무 크고 학생 수가 너무 많거나 혹은 채점 시 객관성을 확보하기가 어렵다는 점을 들었다. 이에 대한 해결책으로서 중학교와 고등학교 영어교사들이 공통적으로 제시한 최우선적인 과제는 객관적인 말하기 수행평가의 시행을 위한 평가 방법과 도구 및 기준을 개발해야 한다는 것이었다.

5.2 배경 변인에 따른 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 인식

말하기 수행평가를 시행하는 중·고등학교 영어교사들의 배경 변인인 교직 경력, 관련 연수 이수 및 말하기 능력에 따라 말하기 수행평가의 전문성, 전문성의 세 하위 요인들, 효용성 및 시행한계성 등 여섯 가지 요인들에 대한 그들의 인식 차이가 어떠한지를 알아보기 위하여 일원배치분산분석과 Scheffe 사후 검정을 실시하였다. 요인별 평균값을 각각의 요인 지수로 사용하였는데, 영어교사들의 배경 변인을 독립변수로, 그리고 각 요인 지수를 종속변수로 두고 분석하였다.

5.2.1 배경 변인에 따른 말하기 수행평가의 전문성에 대한 인식 차이

우선 연구자는 말하기 수행평가의 전문성 전체에 대해서 중·고등학교 영어교사들의 배경 변인에 따른 인식 차이를 비교하였다. <표 38>은 영어교사들의 배경 변인에 따른 말하기 수행평가의 전문성 전체에 대한 기술통계이다. 먼저 교직 경력의 경우 5개 집단 중 6~10년의 교직 경력을 가진 영어교사들의 평균점이 4.05점으로 가장 높았고, 11~15년의 교직 경력을 가진 영어교사들의 평균점은 3.98점으로 두 번째로 높았다. 관련 연수 이수에 있어서는 3개 집단 중 2회 이상의 말하기 수행평가에 관한 연수를 받은 영어교사들의 평균점이 4.09점으로 가장 높았고, 1회의 말하기 수행평가에 관한 연수를 받은 영어교사들의 평균점은 3.94점으로 그 다음으로 높았다. 그리고 말하기 능력은 3개 집단 중 본인의 영어 말하기에 대한 능력이 우수하다고 생각하는 영어교사들의 평균점이 4.02점으로 가장 높았고, 그 다음으로 높은 집단

은 본인의 영어 말하기에 대한 능력이 보통이라고 생각하는 영어교사들로서 그 평균점은 3.88점이었다.

<표 38> 말하기 수행평가의 전문성 전체에 대한 배경 변인별 기술통계

배경 변인	구분	N	최소값	최대값	평균	표준편차	표준오차
교직 경력	5년 이하	69	2.42	5.00	3.88	.500	.060
	6~10년	39	2.67	5.00	4.05	.481	.077
	11~15년	43	2.67	4.83	3.98	.506	.077
	16~20년	18	1.42	4.75	3.70	.696	.164
	21년 이상	74	2.17	5.00	3.85	.546	.063
관련 연수 이수	전혀 없음	97	1.42	5.00	3.74	.594	.060
	1회	81	2.92	5.00	3.94	.433	.048
	2회 이상	65	3.00	5.00	4.09	.481	.060
말하기 능력	부족함	21	2.42	4.42	3.61	.592	.129
	보통임	141	1.42	5.00	3.88	.521	.044
	우수함	81	2.67	5.00	4.02	.506	.056

이러한 말하기 수행평가의 전문성 전체에 대해서 중·고등학교 영어교사들의 배경 변인에 따른 인식 차이가 통계적으로 유의미한지를 파악하고자 일원배치분산분석을 실시하였다. 또한, 일원배치분산분석의 결과에서 통계적으로 유의미한 차이가 있는 것으로 밝혀진 배경 변인이 있다면, 그 배경 변인의 세부 집단들 중 어느 집단들 간에 유의미한 차이가 있는지를 확인하기 위해서 Scheffe 사후검정을 실시하였고, 그 결과는 <표 39>에 제시하였다.

<표 39>에서 보듯이, 영어교사들의 말하기 수행평가의 전문성 전체에 대한 인식은 교직 경력이라는 배경 변인에 있어서 통계적으로 유의미한 차이를 보이지 않았다($F=1.837, p>.05$). 그러나 관련 연수 이수와 말하기 능력이라는 배경 변인에 따라서는 통계적으로 유의미한 차이가 나타났다(관련 연수 이수: $F=9.187, p<.05$; 말하기 능력: $F=5.352, p<.05$).

<표 39> 말하기 수행평가의 전문성 전체에 대한 배경 변인별 인식 차이 검증

배경 변인	구분	N	평균	F	p	Scheffe
교직 경력	5년 이하	69	3.88	1.837	.122	-
	6~10년	39	4.05			
	11~15년	43	3.98			
	16~20년	18	3.70			
	21년 이상	74	3.85			
관련 연수 이수	전혀 없음	97	3.74	9.187	.000	전혀 없음<1회 전혀 없음<2회 이상
	1회	81	3.94			
	2회 이상	65	4.09			
말하기 능력	부족함	21	3.61	5.352	.005	부족함<우수함
	보통임	141	3.88			
	우수함	81	4.02			

관련 연수 이수에 있어서 영어교사들이 말하기 수행평가와 관련된 연수를 자주 받으면 받을수록 말하기 수행평가의 전문성 전체에 대한 인식이 대체로 높아지는 것으로 나타났다. 그리고 Scheffe 사후검정을 실시한 결과, 특히 말하기 수행평가와 관련된 연수를 전혀 받지 않은 영어교사들($\bar{x}=3.74$)보다 말하기 수행평가와 관련된 연수를 1회 받은 영어교사들($\bar{x}=3.94$)과 말하기 수행평가와 관련된 연수를 2회 이상 받은 영어교사들($\bar{x}=4.09$)의 말하기 수행평가의 전문성 전체에 대한 인식이 더 높은 것으로 밝혀졌다. 교사의 전문성이란 끊임없이 교육 정보에 노출되기 위해서 스스로 부단한 노력을 해가는 진화의 연속된 과정임을 생각해 볼 때(이화자, 2006), 영어교사들이 말하기 수행평가와 관련된 온·오프라인 상의 다양한 연수를 자주 수강할수록 말하기 수행평가에 대한 전문성 역시 향상됨은 당연할 것이다. 또한, Zhang과 Burry-Stock(2003)의 연구에서도 평가 훈련을 받지 않은 교사들보다 평가 훈련을 받은 교사들은 수행평가를 시행할 때 자신의 평가 기술(assessment skills)에 대한 인식이 더 높은 것으로 나타났다.

말하기 능력의 경우 자신의 영어 말하기 능력이 상대적으로 높다고 생각하는 영어교사들일수록 말하기 수행평가의 전문성 전체에 대한 인식이 높아지는 것으로 나타났다. 그리고 Scheffe 사후검정을 실시해 본 결과, 특히 자신의 영어 말하기 능력이 부족하다고 생각하는 영어교사들($\bar{x}=3.61$)보다 자신의 영어 말하기 능력이 우수하다고 생각하는 영어교사들($\bar{x}=4.02$)의 말하기 수행평가의 전문성 전체에 대한 인식이 더 높은 것으로 밝혀졌다. Bandura (1977)는 교사의 평가 시행 능력에 대한 자기 자신감이 교사의 평가 시행에 영향을 주는 결정적인 요인이라고 주장하였다. 이는 곧 자신의 영어 말하기 능력이 뛰어나다고 인지하고 있는 영어교사들은 말하기 수행평가의 시행에 대한 전문성도 또한 높음을 설명해 준다.

이어서 말하기 수행평가의 전문성을 구성하는 세 하위 요인들, 즉 평가 방법과 도구의 선정 및 개발, 평가의 실시 및 채점, 그리고 평가 결과의 해석 및 활용에 대해서 중·고등학교 영어교사들의 배경 변인별 기술통계분석을 하였다. 그리고 말하기 수행평가의 전문성을 구성하는 세 하위 요인들이 배경 변인에 따라 통계적으로 유의미한 차이가 있는지를 파악하고자 일원배치분산분석을 실시하였다. 또한, 일원배치분산분석의 결과에서 통계적으로 유의미한 차이가 있는 것으로 밝혀진 배경 변인이 있다면, 그 배경 변인의 세부 집단들 중 어느 집단들 간에 유의미한 차이가 있는지를 확인하기 위해서 Scheffe 사후검정을 실시하였다.

먼저 평가 방법과 도구의 선정 및 개발이라는 전문성의 하위 요인에 대한 중·고등학교 영어교사들의 배경 변인별 기술통계는 <표 40>과 같다. 표에서 보듯이, 교직 경력의 경우 5개 집단 중 6~10년의 교직 경력을 가진 영어교사들의 평균점이 3.76점으로 가장 높았고, 11~15년의 교직 경력을 가진 영어교사들의 평균점은 3.62점으로 두 번째로 높았다. 관련 연수 이수 여부에 있어서는 3개 집단 중 2회 이상의 말하기 수행평가에 관한 연수를 받은 영어교사들의 평

균점이 3.76점으로 가장 높았고, 1회의 말하기 수행평가에 관한 연수를 받은 영어교사들의 평균점은 3.58점으로 그 다음으로 높았다. 그리고 말하기 능력은 3개 집단 중 본인의 영어 말하기에 대한 능력이 우수하다고 생각하는 영어교사들의 평균점이 3.77점으로 가장 높았고, 그 다음 순은 본인의 영어 말하기에 대한 능력이 보통이라고 생각하는 영어교사들로서 그 평균점은 3.43점이었다.

<표 40> 평가 방법과 도구의 선정 및 개발에 대한 배경 변인별 기술통계

요인	배경 변인	구분	N	최소값	최대값	평균	표준 편차	표준 오차	
평가 방법과 도구의 선정 및 개발	교직 경력	5년 이하	69	1.75	5.00	3.42	.699	.084	
		6~10년	39	2.25	5.00	3.76	.616	.099	
		11~15년	43	2.00	5.00	3.62	.761	.116	
		16~20년	18	1.00	4.50	3.42	.831	.196	
		21년 이상	74	1.50	5.00	3.44	.708	.082	
	관련 연수 이수	말하기 능력	전혀 없음	97	1.00	5.00	3.30	.751	.076
			1회	81	2.00	5.00	3.58	.613	.068
			2회 이상	65	2.00	5.00	3.76	.700	.087
			부족함	21	1.50	4.50	3.17	.792	.173
			보통임	141	1.00	5.00	3.43	.676	.057
우수함			81	1.75	5.00	3.77	.698	.078	

이러한 말하기 수행평가의 방법과 도구의 선정 및 개발이라는 전문성에 대해서 중·고등학교 영어교사들의 배경 변인에 따른 인식 차이가 통계적으로 유의미한지를 파악하고자 일원배치분산분석을 실시하였다. 그리고 그 결과에 따른 Scheffe 사후검정을 실시하였고, 그 결과는 <표 41>에 제시하였다.

<표 41>에서 보듯이, 영어교사들의 말하기 수행평가의 전문성 하위 요인들 중 평가 방법과 도구의 선정 및 개발에 대한 인식은 교직 경력이라는 배경 변인에 따라 통계적으로 유의미한 차이를 보이지 않았다($F=1.953, p>.05$). 그러나 관련 연수 이수와 말하기 능력이라는 배경 변인에 따라서는 통계적으로 유의미한 차이가 나타났다(관련 연수 이수: $F=8.920, p<.05$; 말하기 능력: $F=9.103, p<.05$).

<표 41> 평가 방법과 도구의 선정 및 개발에 대한 배경 변인별 인식 차이 검증

요인	배경 변인	구분	N	평균	F	p	Scheffe
평가 방법과 도구의 선정 및 개발	교직 경력	5년 이하	69	3.42	1.953	.102	-
		6~10년	39	3.76			
		11~15년	43	3.62			
		16~20년	18	3.42			
		21년 이상	74	3.44			
관련 연수 이수	관련 연수	전혀 없음	97	3.30	8.920	.000	전혀 없음<1회 전혀 없음<2회 이상
		1회	81	3.58			
		2회 이상	65	3.76			
말하기 능력	말하기 능력	부족함	21	3.17	9.103	.000	부족함<우수함 보통임<우수함
		보통임	141	3.43			
		우수함	81	3.77			

관련 연수 이수에 있어서 영어교사들이 말하기 수행평가와 관련된 연수를 많이 받을수록 말하기 수행평가의 방법과 도구를 선정하고 개발하는 전문성에 대한 인식이 대체로 높아지는 것으로 나타났다. 그리고 Scheffe 사후검정을 실시해 본 결과, 특히 말하기 수행평가와 관련된 연수를 전혀 받지 않은 영어교사들($\bar{x}=3.30$)보다 말하기 수행평가와 관련된 연수를 1회 받은 영어교사들($\bar{x}=3.58$)과 말하기 수행평가와 관련된 연수를 2회 이상 받은 영어교사들($\bar{x}=3.76$)에게 있어서 말하기 수행평가의 방법과 도구를 선정하고 개발하는 전문적인 능력에 대한 인식이 훨씬 높은 것으로 밝혀졌다.

말하기 능력의 경우 자신의 영어 말하기 능력이 상대적으로 뛰어나다고 생각하는 영어교사들일수록 말하기 수행평가의 방법과 도구를 선정하고 개발하는 전문성에 대한 인식이 높은 것으로 나타났다. 그리고 Scheffe 사후검정을 실시한 결과, 특히 자신의 영어 말하기 능력이 부족하다고 생각하는 영어교사들($\bar{x}=3.17$)과 자신의 영어 말하기 능력이 보통이라고 생각하는 영어교사들($\bar{x}=3.43$)보다 자신의 영어 말하기 능력이 우수하다고 생각하는 영어교사들($\bar{x}=3.77$)이 말하기 수행평가의 방법과 도구를 선정하고 개발하는 전문적인 능력에 대한 인식이 더 높은 것으로 밝혀졌다.

다음으로 평가의 실시 및 채점이라는 전문성의 하위 요인에 대한 중·고등학교 영어교사들의 배경 변인별 기술통계는 <표 42>와 같다. 표에서 보듯이, 교직 경력의 5개 집단 중 11년~15년의 교직 경력이 있는 영어교사들의 평균점이 4.66점으로 가장 높았고, 6~10년의 교직 경력을 가진 영어교사들의 평균점은 4.62점으로 두 번째로 높았다. 관련 연수 이수의 경우 3개 집단 중 2회 이상의 말하기 수행평가에 관한 연수를 받은 영어교사들의 평균점이 4.65점으로 가장 높았고, 1회의 말하기 수행평가에 관한 연수를 받은 영어교사들의 평균점은 4.55점으로 그 다음으로 높았다. 그리고 말하기 능력은 3개 집단 중 본인의 영어 말하기에 대한 능력이 우수하다고 생각하는 영어교사들의 평균점이 4.54점으로 가장 높았고, 그 다음으로 높은 집단은 본인의 영어 말하기에 대한 능력이 보통이라고 생각하는 영어교사들로서 그 평균점은 4.53점이었다.

<표 42> 평가의 실시 및 채점에 대한 배경 변인별 기술통계

요인	배경 변인	구분	N	최소값	최대값	평균	표준편차	표준오차
평가의 실시 및 채점	교직 경력	5년 이하	69	2.75	5.00	4.49	.533	.064
		6~10년	39	3.25	5.00	4.62	.473	.076
		11~15년	43	3.75	5.00	4.66	.412	.063
		16~20년	18	2.25	5.00	4.21	.626	.148
		21년 이상	74	3.00	5.00	4.48	.540	.063
	관련 연수 이수	전혀 없음	97	2.25	5.00	4.40	.585	.059
		1회	81	3.25	5.00	4.55	.461	.051
		2회 이상	65	3.00	5.00	4.65	.461	.057
	말하기 능력	부족함	21	2.75	5.00	4.27	.656	.143
		보통임	141	2.25	5.00	4.53	.509	.043
		우수함	81	3.00	5.00	4.54	.498	.055

이러한 말하기 수행평가의 실시 및 채점이라는 전문성에 대해서 중·고등학교 영어교사들의 배경 변인에 따른 인식 차이가 통계적으로 유의미한지를 파악하고자 일원배치분산분석을 실시하였다. 그리고 그 결과에 따른 Scheffe 사후검정을 실시하였고, 그 결과는 <표 43>에 제시하였다.

<표 43> 평가의 실시 및 채점에 대한 배경 변인별 인식 차이 검증

요인	배경 변인	구분	N	평균	F	p	Scheffe
평가의 실시 및 채점	교직 경력	5년 이하	69	4.49	2.993	.019	16~20년<11~15년
		6~10년	39	4.62			
		11~15년	43	4.66			
		16~20년	18	4.21			
		21년 이상	74	4.48			
	관련 연수 이수	전혀 없음	97	4.40	4.839	.009	전혀 없음<2회 이상
		1회	81	4.55			
		2회 이상	65	4.65			
	말하기 능력	부족함	21	4.27	2.491	.085	-
		보통임	141	4.53			
우수함		81	4.54				

<표 43>에서 보듯이, 영어교사들의 말하기 수행평가의 전문성 하위 요인들 중 평가의 실시 및 채점에 대한 인식은 영어교사들의 교직 경력과 관련 연수 이수라는 배경 변인에 따라 통계적으로 유의미한 차이가 있는 것으로 나타났다(교직 경력: $F=2.993$, $p<.05$; 관련 연수 이수: $F=4.839$, $p<.05$). 그러나 말하기 능력이라는 배경 변인에 따라서는 통계적으로 유의미한 차이를 보이지 않았다($F=2.491$, $p>.05$).

교직 경력의 경우 말하기 수행평가를 실시하고 채점하는 전문성에 대한 인식은 영어교사들의 교직 경력이 15년이 될 때까지는 점점 높아지다가 교직 경력이 16년을 넘어서면서부터 다소 낮아지는 것으로 나타났다. 그리고 Scheffe 사후검정을 실시해 본 결과, 특히 16~20년의 교직 경력을 가진 영

어교사들($\bar{x}=4.21$)보다 11~15년의 교직 경력을 가진 영어교사들($\bar{x}=4.66$)에게 있어서 말하기 수행평가를 실시하고 채점하는 전문적인 능력에 대한 인식이 보다 높은 것으로 밝혀졌다. 이 결과는 교직 경력이 적을수록 교사의 전문성 함양에 관한 연수의 효과가 높으며, 교사의 교직 경력이 증가할수록 연수의 효과가 점점 감소한다는 Shih와 Wang(2010)의 연구 결과와 연관성이 있다.

관련 연수 이수에 있어서 말하기 수행평가를 실시하고 채점하는 전문성에 대한 인식은 영어교사들이 말하기 수행평가와 관련된 연수를 이수한 횟수가 많아질수록 점차적으로 높아지는 것으로 나타났다. 그리고 Scheffe 사후검정을 실시한 결과, 특히 말하기 수행평가와 관련된 연수를 전혀 받지 않은 영어교사들($\bar{x}=4.40$)보다 말하기 수행평가와 관련된 연수를 2회 이상 받은 영어교사들($\bar{x}=4.65$)에게 있어서 말하기 수행평가를 실시하고 채점하는 전문적인 능력에 대한 인식이 더 높은 것으로 밝혀졌다.

마지막으로, 평가 결과의 해석 및 활용이라는 전문성 하위 요인에 대한 중·고등학교 영어교사들의 배경 변인별 기술통계는 <표 44>와 같다. 표에서 보듯이, 교직 경력의 경우 5개 집단 중 6~10년의 교직 경력을 가진 영어교사들의 평균점이 3.77점으로 가장 높았고, 5년 이하의 교직 경력을 가진 영어교사들의 평균점은 3.74점으로 두 번째로 높았다. 관련 연수 이수에 있어서는 3개 집단 중 2회 이상의 말하기 수행평가에 관한 연수를 받은 영어교사들의 평균점이 3.86점으로 가장 높았고, 1회의 말하기 수행평가에 관한 연수를 받은 영어교사들의 평균점은 3.69점으로 그 다음으로 높았다. 그리고 말하기 능력은 3개 집단 중 본인의 영어 말하기에 대한 능력이 우수하다고 생각하는 영어교사들의 평균점이 3.74점으로 가장 높았고, 그 다음 순은 본인의 영어 말하기에 대한 능력이 보통이라고 생각하는 영어교사들로서 그 평균점은 3.67점이었다.

<표 44> 평가 결과의 해석 및 활용에 대한 배경 변인별 기술통계

요인	배경 변인	구분	N	최소값	최대값	평균	표준편차	표준오차
평가 결과의 해석 및 활용	교직 경력	5년 이하	69	2.50	5.00	3.74	.649	.078
		6~10년	39	2.50	5.00	3.77	.635	.102
		11~15년	43	2.00	4.75	3.65	.718	.110
		16~20년	18	1.00	4.75	3.49	.864	.204
		21년 이상	74	1.00	5.00	3.62	.754	.088
	관련 연수 이수	전혀 없음	97	1.00	5.00	3.53	.774	.079
		1회	81	2.00	5.00	3.69	.637	.071
		2회 이상	65	2.25	5.00	3.86	.650	.081
	말하기 능력	부족함	21	2.25	4.25	3.39	.640	.140
		보통임	141	1.00	5.00	3.67	.727	.061
우수함		81	2.00	5.00	3.74	.681	.076	

이러한 말하기 수행평가 결과의 해석 및 활용이라는 전문성에 대해서 중·고등학교 영어교사들의 배경 변인에 따른 인식 차이가 통계적으로 유의미한지를 파악하고자 일원배치분산분석을 실시하였다. 그리고 그 결과에 따른 Scheffe 사후검정을 실시하였고, 그 결과는 <표 45>에 제시하였다.

<표 45> 평가 결과의 해석 및 활용에 대한 배경 변인별 인식 차이 검증

요인	배경 변인	구분	N	평균	F	p	Scheffe
평가 결과의 해석 및 활용	교직 경력	5년 이하	69	3.74	.765	.549	-
		6~10년	39	3.77			
		11~15년	43	3.65			
		16~20년	18	3.49			
		21년 이상	74	3.62			
관련 연수 이수	전혀 없음	전혀 없음	97	3.53	4.548	.012	전혀 없음<2회 이상
		1회	81	3.69			
		2회 이상	65	3.86			
말하기 능력	부족함	부족함	21	3.39	2.068	.129	-
		보통임	141	3.67			
		우수함	81	3.74			

<표 45>에서 보듯이, 영어교사들의 말하기 수행평가의 전문성 하위 요인들 중 평가 결과의 해석 및 활용에 대한 인식은 교직 경력과 말하기 능력이라는 배경 변인에 따라 통계적으로 유의미한 차이를 보이지 않았다(교직 경력: $F=.765, p>.05$; 말하기 능력: $F=2.068, p>.05$). 그러나 관련 연수 이수여부에 있어서는 통계적으로 유의미한 차이가 있는 것으로 나타났다($F=4.548, p<.05$).

관련 연수 이수여부에 있어서 영어교사들이 말하기 수행평가와 관련된 연수를 자주 받으면 받을수록 말하기 수행평가의 결과를 해석하고 이를 활용하는 전문성에 대한 인식이 점차 높아지는 것으로 나타났다. 그리고 Scheffe 사후검정을 실시한 결과, 특히 말하기 수행평가에 관한 연수를 전혀 받지 않는 영어교사들($\bar{x}=3.53$)보다 말하기 수행평가에 관한 연수를 2회 이상 받은 영어교사들($\bar{x}=3.86$)이 말하기 수행평가의 결과를 해석하고 이를 활용하는 전문적인 능력에 대한 인식이 훨씬 높은 것으로 밝혀졌다.

정리하면, 말하기 수행평가의 전문성에 대한 중·고등학교 영어교사들의 인식 차이는 교직 경력이라는 배경 변인의 경우, 특히 16~20년의 교직 경력을 지닌 영어교사들보다는 11~15년의 교직 경력을 지닌 영어교사들에게서 말하기 수행평가를 실시하고 채점하는 전문적인 능력에 대한 인식이 더 높은 것으로 나타났다. 그리고 말하기 수행평가에 관한 연수를 전혀 이수하지 않는 영어교사들보다 말하기 수행평가에 관한 연수를 적어도 1회나 혹은 2회 이상 이수한 영어교사들이 말하기 수행평가의 시행과 관련된 모든 전문적인 능력에 대한 인식을 훨씬 높게 가진 것으로 밝혀졌다. 또한, 자신의 영어 말하기 능력이 부족하거나 보통이라고 생각하는 영어교사들보다는 자신의 영어 말하기 능력이 우수하다고 생각하는 영어교사들에게서 말하기 수행평가에 대한 전반적인 전문성이나, 특히 말하기 수행평가의 방법과 도구를 선정하고 개발하는 전문성에 대한 인식이 더 높은 것으로 나타났다.

5.2.2 배경 변인에 따른 말하기 수행평가의 효용성에 대한 인식 차이

말하기 수행평가의 효용성에 대한 중·고등학교 영어교사들의 교직 경력, 관련 연수 이수 및 말하기 능력이라는 배경 변인별 기술통계는 <표 46>과 같다.

<표 46> 말하기 수행평가의 효용성에 대한 배경 변인별 기술통계

배경 변인	구분	N	최소값	최대값	평균	표준편차	표준오차
교직 경력	5년 이하	69	1.60	5.00	3.39	.688	.083
	6~10년	39	2.40	5.00	3.42	.620	.099
	11~15년	43	2.00	5.00	3.43	.722	.110
	16~20년	18	1.00	4.40	3.47	.758	.179
	21년 이상	74	2.00	5.00	3.45	.600	.070
관련 연수 이수	전혀 없음	97	1.00	5.00	3.29	.734	.075
	1회	81	2.00	5.00	3.46	.598	.067
	2회 이상	65	2.40	5.00	3.59	.572	.071
말하기 능력	부족함	21	2.00	4.20	3.13	.691	.151
	보통임	141	1.00	5.00	3.38	.638	.054
	우수함	81	2.00	5.00	3.58	.657	.073

<표 46>에서 보듯이, 교직 경력의 경우 5개 집단 중 16~20년의 교직 경력을 가진 영어교사들의 평균점이 3.47점으로 가장 높았고, 21년 이상의 교직 경력을 가진 영어교사들의 평균점은 3.45점으로 두 번째로 높았다. 관련 연수 이수에 있어서는 3개 집단 중 2회 이상의 말하기 수행평가에 관한 연수를 받은 영어교사들의 평균점이 3.59점으로 가장 높았고, 1회의 말하기 수행평가에 관한 연수를 받은 영어교사들의 평균점은 3.46점으로 그 다음으로 높았다. 그리고 말하기 능력은 3개 집단 중 자신의 영어 말하기에 대한 능력이 우수하다고 생각하는 영어교사들의 평균점이 3.58점으로 가장 높았고, 그 다음으로 높은 집단은 자신의 영어 말하기에 대한 능력이 보통이라고 생각하는 영어교사들로서 그 평균점은 3.38점이었다.

이러한 말하기 수행평가의 효용성에 대해서 중·고등학교 영어교사들의 배경 변인에 따른 인식 차이가 통계적으로 유의미한지를 파악하고자 일원배치 분산분석을 실시하였다. 그리고 그 결과에 따른 Scheffe 사후검정을 실시하였고, 그 결과는 <표 47>에 제시하였다.

<표 47> 말하기 수행평가의 효용성에 대한 배경 변인별 인식 차이 검증

배경 변인	구분	N	평균	F	p	Scheffe
교직 경력	5년 이하	69	3.39	.117	.977	-
	6~10년	39	3.42			
	11~15년	43	3.43			
	16~20년	18	3.47			
	21년 이상	74	3.45			
관련 연수 이수	전혀 없음	97	3.29	4.381	.014	전혀 없음<2회 이상
	1회	81	3.46			
	2회 이상	65	3.56			
말하기 능력	부족함	21	3.13	4.612	.011	부족함<우수함
	보통임	141	3.38			
	우수함	81	3.58			

<표 47>에서 보듯이, 영어교사들의 말하기 수행평가의 효용성에 대한 인식은 교직 경력이라는 배경 변인에 따라 통계적으로 유의미한 차이를 보이지 않았다($F=.117, p>.05$). 그러나 관련 연수 이수와 말하기 능력에 있어서는 통계적으로 유의미한 차이가 있는 것으로 나타났다(관련 연수 이수: $F=4.381, p<.05$; 말하기 능력: $F=4.612, p<.05$).

관련 연수 이수의 경우 영어교사들이 말하기 수행평가와 관련된 연수를 받는 횟수가 많아질수록 말하기 수행평가의 효용성에 대한 인식이 대체로 높아지는 것으로 나타났다. 그리고 Scheffe 사후검정을 실시한 결과, 특히 말하기 수행평가에 관한 연수를 전혀 받지 않은 영어교사들($\bar{x}=3.29$)보다 말하기 수행평가에 관한 연수를 2회 이상 받은 영어교사들($\bar{x}=3.56$)이 말하기 수

행평가의 효용성에 대해서 보다 높은 인식을 가진 것으로 밝혀졌다. 즉, 평가에 대한 교사 연수가 평가의 교육적 장점에 대한 교사의 믿음에 영향을 주는 경향이 있다(Borg, 2003)는 주장에서 알 수 있듯이, 말하기 수행평가와 관련한 적절한 연수를 이수함으로써 영어교사들은 말하기 수행평가에 대한 긍정적인 인식을 더 가지게 될 것이다.

말하기 능력의 경우에 있어서는 자신의 영어 말하기 능력이 상대적으로 뛰어나다고 생각하는 영어교사들일수록 말하기 수행평가의 효용성에 대한 인식이 차츰 높아지는 것으로 나타났다. 그리고 Scheffe 사후검정을 실시해 본 결과, 특히 자신의 영어 말하기 능력이 부족하다고 생각하는 영어교사들($\bar{x}=3.13$)보다 자신의 영어 말하기 능력이 우수하다고 생각하는 영어교사들($\bar{x}=3.58$)이 말하기 수행평가의 효용성에 대해서 훨씬 더 긍정적으로 인식하고 있는 것으로 밝혀졌다. Woods(1996)는 교사의 신념이란 자신의 사고(thinking)와 해석(interpretation) 및 계획된 행동(planned action)에 대해서 무엇이 영향을 미치는지를 아는 것이라고 하였다. 즉, 영어교사가 스스로 자신의 말하기 능력이 우수하다는 믿음을 가지고 있으면 말하기 평가의 중요성을 인식하게 되고, 그럼으로써 영어교사는 말하기 수행평가의 시행이 긍정적인 효과를 불러올 것이라는 믿음을 가지게 될 것이다.

정리하면, 말하기 수행평가와 관련된 연수를 전혀 받지 않은 중·고등학교 영어교사들보다는 말하기 수행평가에 관한 연수를 2회 이상 받은 영어교사들에게서 말하기 수행평가의 효용성에 대한 긍정적인 인식이 더 높은 것으로 나타났다. 그리고 자신의 영어 말하기 능력이 부족하다고 생각하는 영어교사들보다 자신의 영어 말하기 능력이 우수하다고 생각하는 영어교사들이 말하기 수행평가의 효용성에 대해서 훨씬 더 긍정적으로 인식하고 있는 것으로 밝혀졌다.

5.2.3 배경 변인에 따른 말하기 수행평가의 시행한계성에 대한 인식 차이

말하기 수행평가의 시행한계성에 대한 중·고등학교 영어교사들의 교직 경력, 관련 연수 이수 및 말하기 능력이라는 배경 변인별 기술통계는 <표 48>과 같다. 교직 경력의 경우 5개 집단 중 6~10년의 교직 경력을 가진 영어교사들의 평균점이 3.70점으로 가장 높았고, 16~20년의 교직 경력을 가진 영어교사들의 평균점은 3.61점으로 두 번째로 높았다. 관련 연수 이수에는 3개 집단 중 말하기 수행평가와 관련된 연수를 전혀 이수하지 않은 영어교사들의 평균점이 3.59점으로 가장 높았고, 1회의 말하기 수행평가와 관련된 연수를 이수한 영어교사들의 평균점은 3.46점으로 그 다음으로 높았다. 그리고 말하기 능력은 3개 집단 중 본인의 영어 말하기에 대한 능력이 부족하다고 생각하는 영어교사들의 평균점이 3.66점으로 가장 높았고, 그 다음으로 높은 집단은 본인의 영어 말하기에 대한 능력이 보통이라고 생각하는 영어교사들로서 그 평균점은 3.52점이었다.

<표 48> 말하기 수행평가의 시행한계성에 대한 배경 변인별 기술통계

배경 변인	구분	N	최소값	최대값	평균	표준편차	표준오차
교직 경력	5년 이하	69	1.00	5.00	3.39	.663	.080
	6~10년	39	2.00	4.80	3.70	.684	.109
	11~15년	43	1.40	5.00	3.47	.767	.117
	16~20년	18	2.40	5.00	3.61	.836	.197
	21년 이상	74	1.00	4.80	3.42	.676	.079
관련 연수 이수	전혀 없음	97	1.00	5.00	3.59	.723	.073
	1회	81	2.00	5.00	3.46	.597	.066
	2회 이상	65	1.00	4.80	3.33	.784	.097
말하기 능력	부족함	21	2.20	5.00	3.66	.617	.135
	보통임	141	1.00	5.00	3.52	.700	.059
	우수함	81	1.00	4.80	3.36	.729	.081

이러한 말하기 수행평가의 시행한계성에 대해서 중·고등학교 영어교사들의 배경 변인에 따른 인식 차이가 통계적으로 유의미한지를 파악하고자 일원배치분산분석을 실시하였고, 그 결과는 <표 49>에 제시하였다.

<표 49> 말하기 수행평가의 시행한계성에 대한 배경 변인별 인식 차이 검증

배경 변인	구분	N	평균	F	p	Scheffe
교직 경력	5년 이하	69	3.39	1.535	.193	-
	6~10년	39	3.70			
	11~15년	43	3.47			
	16~20년	18	3.61			
	21년 이상	74	3.42			
관련 연수 이수	전혀 없음	97	3.59	2.710	.069	-
	1회	81	3.46			
	2회 이상	65	3.33			
말하기 능력	부족함	21	3.66	1.991	.139	-
	보통임	141	3.52			
	우수함	81	3.36			

<표 49>에서 보듯이, 말하기 수행평가의 시행한계성에 대한 영어교사들의 인식은 교직 경력과 관련 연수 이수 및 말하기 능력 등 모든 배경 변인에 따라 통계적으로 유의미한 차이를 보이지 않았다(교직 경력: $F=1.535$, $p>.05$; 관련 연수 이수: $F=2.710$, $p>.05$; 말하기 능력: $F=1.991$, $p>.05$).

5.2.4 요약

이상과 같이 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 우리나라 중·고등학교 영어교사들의 인식이 교직 경력, 관련 연수 이수 및 말하기 능력이라는 배경 변인에 따라 어떤 차이가 있는지를 분석한 결과를 요

약하면 다음과 같다. 첫째, 영어교사들의 말하기 수행평가의 전반적인 전문성에 대한 인식은 교직 경력에 따라 통계적으로 유의미한 차이가 없는 것으로 나타났다. 그러나 관련 연수 이수와 말하기 능력에 따라서는 통계적으로 유의미한 차이가 있는 것으로 나타났다. 즉, 말하기 수행평가에 관한 연수를 전혀 받지 않은 영어교사들보다 말하기 수행평가에 관한 연수를 1회나 혹은 2회 이상 받은 영어교사들에게 있어서 말하기 수행평가의 전반적인 전문성에 대한 인식이 더 높은 것으로 밝혀졌다. 또한, 자신의 영어 말하기 능력이 부족하다고 생각하는 영어교사들보다 자신의 영어 말하기 능력이 우수하다고 생각하는 영어교사들에게 있어서 말하기 수행평가의 전반적인 전문성에 대한 인식이 훨씬 높은 것으로 밝혀졌다.

둘째, 영어교사들의 말하기 수행평가의 전문성 세 하위 요인들 중에서 우선 평가 방법과 도구의 선정 및 개발이라는 전문성에 대한 인식은 교직 경력에 따라 통계적으로 유의미한 차이를 보이지 않았지만, 관련 연수 이수와 말하기 능력에 따라서는 통계적으로 유의미한 차이를 보였다. 즉, 말하기 수행평가와 관련된 연수를 전혀 받지 않은 영어교사들보다 말하기 수행평가와 관련된 연수를 1회나 혹은 2회 이상 받은 영어교사들이 말하기 수행평가의 방법과 도구의 선정 및 개발이라는 전문성에 대한 인식을 더 높게 가진 것으로 밝혀졌다. 또한, 자신의 영어 말하기 능력이 우수하다고 생각하는 영어교사들은 자신의 영어 말하기 능력이 부족하거나 보통이라고 생각하는 영어교사들보다 말하기 수행평가의 방법과 도구의 선정 및 개발이라는 전문성에 대한 인식을 훨씬 높게 가진 것으로 밝혀졌다.

셋째, 평가의 실시 및 채점이라는 전문성에 대한 인식은 말하기 능력에 따라 통계적으로 유의미한 차이를 보이지 않았으나, 교직 경력과 관련 연수 이수에 따라서는 통계적으로 유의미한 차이를 보였다. 즉, 교직 경력이 16~20년인 영어교사들보다 교직 경력이 11~15년인 영어교사들에게 있어서 말하기

수행평가의 실시 및 채점이라는 전문성에 대한 인식이 더 높은 것으로 밝혀졌다. 또한, 말하기 수행평가에 관한 연수를 전혀 이수하지 않은 영어교사들보다 말하기 수행평가에 관한 연수를 2회 이상 이수한 영어교사들에게 있어서 말하기 수행평가를 실시하고 채점하는 전문적인 능력에 대한 인식이 더 높은 것으로 밝혀졌다.

넷째, 평가 결과의 해석 및 활용이라는 전문성에 대한 인식은 교직 경력과 말하기 능력에 따라 통계적으로 유의미한 차이가 나타나지 않았지만, 관련 연수 이수에 따라서는 통계적으로 유의미한 차이가 있는 것으로 나타났다. 즉, 말하기 수행평가에 관한 연수를 전혀 받지 않은 영어교사들보다 말하기 수행평가에 관한 연수를 2회 이상 받은 영어교사들에게 있어서 말하기 수행평가의 결과를 해석하고 활용하는 전문적인 능력에 대한 인식이 더 높은 것으로 밝혀졌다.

다섯째, 영어교사들의 말하기 수행평가의 효용성에 대한 인식은 교직 경력에 따라 통계적으로 유의미한 차이가 없는 것으로 나타났다. 그러나 관련 연수 이수와 말하기 능력에 따라서는 통계적으로 유의미한 차이가 있는 것으로 나타났다. 즉, 말하기 수행평가에 관한 연수를 전혀 받지 않은 영어교사들에 비해서 말하기 수행평가에 관한 연수를 2회 이상 받은 영어교사들이 말하기 수행평가의 효용성에 대해서 더 긍정적인 인식을 가지고 있는 것으로 밝혀졌다. 또한, 자신의 영어 말하기 능력이 부족하다고 생각하는 영어교사들보다 자신의 영어 말하기 능력이 우수하다고 생각하는 영어교사들이 말하기 수행평가의 효용성에 대해서 보다 긍정적으로 인식하고 있는 것으로 밝혀졌다.

마지막으로, 영어교사들의 말하기 수행평가의 시행한계성에 대한 인식은 교직 경력에 따라 통계적으로 유의미한 차이가 없었다. 뿐만 아니라, 관련 연수 이수와 말하기 능력이라는 변인에 따라서도 통계적으로 유의미한 차이가 발견되지 않았다.

이러한 결과들을 종합해서 정리해 보면, 우선 교직 경력이 16~20년인 영어교사들보다 교직 경력이 11~15년인 영어교사들의 말하기 수행평가를 실시하고 채점하는 전문적인 능력에 대한 인식이 훨씬 높았다. 그리고 관련 연수 이수에 있어서 말하기 수행평가에 관한 연수를 전혀 받지 않은 영어교사들보다 말하기 수행평가에 관한 연수를 최소 1회나 2회 이상 받은 영어교사들의 말하기 수행평가의 전문성에 대한 인식이 여러모로 높았을 뿐만 아니라, 말하기 수행평가의 효용성에 대한 긍정적인 인식 또한 높았다. 더구나 본인의 영어 말하기 능력이 상대적으로 우수하다고 생각하는 영어교사들일수록 말하기 수행평가의 효용성에 대한 긍정적인 인식이 높을 뿐만 아니라 말하기 수행평가의 전반적인 전문성에 대한 인식도 높았는데, 특히 말하기 수행평가의 방법과 도구를 선정하거나 개발하는 전문적인 능력에 대한 인식이 훨씬 높았다.

5.3 말하기 수행평가의 전문성, 효용성 및 시행한계성 간의 상관관계

연구자는 우리나라의 중·고등학교 영어교사들이 인식하는 말하기 수행평가의 전문성, 효용성 및 시행한계성이라는 요인들이 서로 간에 어떠한 상관성을 가지는지를 파악하기 위해서 영어교사들 전체 집단을 기준으로 상관분석을 실시하였다. 그리고 중학교와 고등학교 집단별로 이 세 가지 요인들의 상관관계가 구체적으로 어떠한지도 살펴보았다.

5.3.1 전체 영어교사들을 대상으로 한 세 주요 요인들 간의 상관관계 분석

우선 중·고등학교 영어교사들 전체를 대상으로 말하기 수행평가의 전문성, 효용성 및 시행한계성이라는 세 주요 요인들에 대한 기술통계분석을 실시하였고, 그 결과는 <표 50>과 같다.

<표 50> 세 주요 요인들에 대한 기술통계(전체 영어교사들)

요인	N	최소값	최대값	평균	표준편차	표준오차
전문성	243	1.42	5.00	3.90	.532	.034
효용성	243	1.00	5.00	3.43	.658	.042
시행한계성	243	1.00	5.00	3.48	.706	.045

<표 50>에서 보듯이, 전문성과 효용성 및 시행한계성에 대한 영어교사들의 평균점들 중 전문성에 대한 평균점이 3.90점으로 가장 높았다. 그리고 효용성의 평균점은 3.43점으로 가장 낮았고, 시행한계성의 평균점은 3.48점이었다. 이러한 전체 중·고등학교 영어교사들이 인식하는 말하기 수행평가의 전문성, 효용성 및 시행한계성이 서로 어떠한 상관관계에 있는지를 알아보기 위해서 상관분석을 실시하였으며, 그 결과는 <표 51>에 제시하였다.

<표 51> 세 주요 요인들 간의 상관관계(전체 영어교사들)

요인	전문성	효용성
전문성		
효용성	.596 (.000)	
시행한계성	-.188 (.003)	-.180 (.005)

<표 51>에서 보듯이, 전체 중·고등학교 영어교사들의 말하기 수행평가의 효용성은 말하기 수행평가의 전문성과 서로 높은 정의 상관관계($r=.596$, $p<.05$)를 가진 것으로 나타났다. 이 결과는 말하기 수행평가의 효용성에 대한 긍정적인 인식이 높은 중·고등학교 영어교사들일수록 말하기 수행평가의 전문성에 대한 인식이 높으며, 또한 그 역의 관계도 성립할 수 있음을 보여준다. 즉, 말하기 수행평가의 효용성에 대해서 긍정적으로 생각하는 영어교사들은 말하기 수행평가에 대한 자신들의 평가 전문성 또한 높다고 생각하고 있음을 의미한다. 이와 관련하여 McMillan(2007)은 평가 시행 시 교사가 지닌 평가에 대한 신념은 자신의 평가 시행을 결정하는 데 있어서 중요한 역할을 한다고 하였다. 그러므로 말하기 수행평가에 대한 긍정적인 효용성과 높은 전문성을 지닌 것으로 믿는 영어교사들은 아마도 말하기 수행평가의 시행에 있어서도 보다 적극적인 의지를 보일 것으로 예측된다.

한편, 전체 중·고등학교 영어교사들의 말하기 수행평가의 시행한계성은 말하기 수행평가의 전문성과 서로 음의 상관관계($r=-.188$, $p<.05$)에 있지만 그 상관성은 그다지 높은 편이 아닌 것으로 나타났다. 이 결과는 중·고등학교 영어교사들의 말하기 수행평가의 시행한계성에 대한 인식이 높을수록 말하기 수행평가의 전문성에 대한 인식이 낮음을 보여주며, 또한 그 역의 관계도 성립할 수 있음을 보여준다. 현재 중학교와 고등학교에서 진행되고 있는 수행평가는 주로 듣기나 어휘 평가, 따라 읽기, 혹은 암기식 발표가 위주인데(김성연, 이호경, 2013; 방인자, 천세영, 2011; 이소영, 2000; 장인순, 2004), 특히 말하기 수행평가는 학생들의 말하기 능력의 측정이라는 본질에서 벗어난 형식적인 평가 방법을 사용하는 경우가 많아 그 진정성이 낮다는 문제가 발생하고 있다. 이러한 실태는 객관적인 말하기 수행평가 도구나 기준이 없고 채점의 신뢰성조차 장담하기 힘든 현실(고현우, 2014; 이소영, 2008)에서 비롯된 결과이며, 더구나 말하기라는 복잡한 특성을 가진 언어 기

능을 수행평가로서 직접적으로 측정하는 일은 아주 힘이 들기 때문이다 (Heaton, 1975). 그러다보니 영어교사들은 진정성이 있는 말하기 수행평가를 시행하지 않거나 혹은 꺼리게 되고, 결과적으로 말하기 수행평가에 대한 전문성도 그에 따라서 낮아질 것으로 추측된다.

아울러 전체 중·고등학교 영어교사들의 말하기 수행평가의 시행한계성은 말하기 수행평가의 효용성과 서로 음의 상관관계($r=-.180, p<.05$)를 가지지만 그 상관성은 높지 않은 것으로 나타났다. 이 결과는 말하기 수행평가의 시행한계성이 높다고 인식하는 중·고등학교 영어교사들일수록 말하기 수행평가의 효용성에 대한 긍정적인 인식이 낮음을 보여주며, 또한 그 역의 관계도 성립할 수 있음을 보여준다. 앞 절에서 밝혀진 대로 영어교사들이 말하기 수행평가의 시행이 어렵다고 생각하는 가장 큰 이유는 학급 규모가 너무 크고 학생 수가 많으며, 또한 채점 시의 객관성 확보가 어렵다는 점이었다. 즉, 말하기 수행평가를 위한 타당한 도구나 방법 및 객관적인 채점 기준조차 제대로 정립되어 있지 않은 상황에서 영어교사들이 큰 규모의 학급과 학생들을 대상으로 학생들의 진정한 말하기 능력을 측정하는 일은 어려울 것이라는 회의적인 생각을 가지게 된다. 이는 곧 말하기 수행평가의 효용성에 대한 영어교사들의 긍정적인 인식마저 떨어뜨리는 주요한 원인인 것으로 추측된다.

정리하면, 전체 중·고등학교 영어교사들은 말하기 수행평가의 효용성에 대한 긍정적인 인식이 높을수록 말하기 수행평가의 전문성에 대한 인식은 높은 반면, 그들의 말하기 수행평가의 시행한계성에 대한 인식이 높을수록 말하기 수행평가의 전문성에 대한 인식은 낮은 것으로 나타났다. 그리고 말하기 수행평가의 시행한계성을 높게 인식하는 중·고등학교 영어교사들일수록 말하기 수행평가의 효용성에 대한 긍정적인 인식은 낮은 것으로 밝혀졌다. 더구나 이러한 세 요인들은 각각 서로 역의 관계도 성립될 수 있음을 보여주었다.

5.3.2 중학교 영어교사들을 대상으로 한 세 주요 요인들 간의 상관관계 분석

중학교 영어교사들을 대상으로 말하기 수행평가의 전문성, 효용성 및 시행한계성이라는 세 주요 요인들에 대한 기술통계분석을 실시하였고, 그 결과는 <표 52>에 제시하였다. 전문성과 효용성 및 시행한계성에 대한 영어교사들의 평균점들 중 전문성에 대한 평균점이 4.02점으로 가장 높았다. 그리고 효용성의 평균점은 3.48점이고, 시행한계성의 평균점은 3.34점으로 가장 낮았다.

<표 52> 세 주요 요인들에 대한 기술통계(중학교 영어교사들)

요인	N	최소값	최대값	평균	표준편차	표준오차
전문성	128	2.17	5.00	4.02	.464	.041
효용성	128	2.00	5.00	3.48	.628	.056
시행한계성	128	1.00	4.80	3.34	.727	.064

이러한 중학교 영어교사들이 인식하는 말하기 수행평가의 전문성, 효용성 및 시행한계성이 서로 어떠한 상관관계에 있는지를 알아보기 위해서 상관분석을 실시하였으며, 그 결과는 <표 53>에 제시하였다.

<표 53> 세 주요 요인들 간의 상관관계(중학교 영어교사들)

요인	전문성	효용성
전문성		
효용성	.520 (.000)	
시행한계성	-.128 (.150)	-.135 (.129)

<표 53>에서 보듯이, 중학교 영어교사들의 말하기 수행평가의 효용성은 말하기 수행평가의 전문성과 서로 높은 정의 상관관계($r=.520, p<.05$)를 가진 것으로 나타났다. 이 결과는 말하기 수행평가의 효용성에 대한 중학교 영어교사들의 긍정적인 인식이 높으면 높을수록 말하기 수행평가의 전문성에 대한 인식이 높으며, 또한 그 역의 관계도 성립할 수 있음을 보여준다.

한편, 중학교 영어교사들의 말하기 수행평가의 시행한계성과 말하기 수행평가의 전문성은 서로 음의 상관관계($r=-.128, p>.05$)가 있는 것으로 나타났지만 통계적으로 유의미하지 않았다. 또한, 중학교 영어교사들의 말하기 수행평가의 시행한계성과 말하기 수행평가의 효용성과의 상관관계에서도 두 요인들이 서로 음의 상관관계($r=-.135, p>.05$)에 있는 것으로 나타났으나 통계적으로 유의미하지는 않았다.

5.3.3 고등학교 영어교사들을 대상으로 한 세 주요 요인들 간의 상관관계 분석

고등학교 영어교사들을 대상으로 말하기 수행평가의 전문성, 효용성 및 시행한계성이라는 세 주요 요인들에 대한 기술통계분석을 실시하였고, 그 결과는 <표 54>와 같다. 전문성과 효용성 및 시행한계성에 대한 영어교사들의 평균점들 중 전문성에 대한 평균점이 3.76점으로 가장 높았다. 그리고 효용성의 평균점은 3.36점이고, 시행한계성의 평균점은 3.64점이었다.

<표 54> 세 주요 요인들에 대한 기술통계(고등학교 영어교사들)

요인	N	최소값	최대값	평균	표준편차	표준오차
전문성	115	1.42	5.00	3.76	.571	.053
효용성	115	1.00	5.00	3.36	.688	.064
시행한계성	115	2.20	5.00	3.64	.650	.061

이러한 고등학교 영어교사들이 인식하는 말하기 수행평가의 전문성, 효용성 및 시행한계성이 서로 어떠한 상관관계에 있는지를 알아보기 위해서 상관분석을 실시하였으며, 그 결과는 <표 55>에 제시하였다. 표에서 보듯이, 고등학교 영어교사들이 인식하는 말하기 수행평가의 효용성은 말하기 수행평가의 전문성과 서로 높은 정의 상관관계($r=.657, p<.05$)를 가진 것으로 나타났다. 이 결과는 말하기 수행평가의 효용성에 대한 고등학교 영어교사들의 인식이 긍정적일수록 말하기 수행평가의 전문성에 대한 인식이 높으며, 또한 그 역의 관계도 성립할 수 있음을 보여준다.

<표 55> 세 주요 요인들 간의 상관관계(고등학교 영어교사들)

요인	전문성	효용성
전문성		
효용성	.657 (.000)	
시행한계성	-.164 (.079)	-.200 (.032)

한편, 고등학교 영어교사들의 말하기 수행평가의 시행한계성과 말하기 수행평가의 전문성은 서로 음의 상관관계($r=-.164, p>.05$)가 있는 것으로 나타났지만 통계적으로 유의미하지 않았다. 그러나 고등학교 영어교사들의 말하기 수행평가의 시행한계성과 말하기 수행평가의 효용성과 서로 음의 상관관계($r=-.200, p<.05$)에 있지만 그 상관성은 그리 높지 않은 것으로 나타났다. 이 결과는 말하기 수행평가의 시행상의 어려움이 많다고 생각하는 고등학교 영어교사들일수록 말하기 수행평가의 효용성에 대한 긍정적인 인식이 낮으며, 또한 그 역의 관계도 성립할 수 있음을 보여준다.

5.3.4 요약

이상으로 말하기 수행평가를 시행하는 우리나라 중·고등학교 영어교사들의 말하기 수행평가의 전문성, 효용성 및 시행한계성이라는 세 주요 요인들 간의 상관관계를 분석한 결과를 요약하자면 다음과 같다. 첫째, 전체 중·고등학교 영어교사들의 말하기 수행평가의 효용성은 말하기 수행평가의 전문성과 서로 높은 정의 상관관계를 가지는 것으로 나타났다. 특히, 학교급별로 두 요인들의 상관관계를 분석해 본 결과, 중학교와 고등학교 영어교사들 두 집단 모두 말하기 수행평가의 효용성은 말하기 수행평가의 전문성과 서로 높은 정의 상관관계를 가지는 것으로 밝혀졌다.

둘째, 전체 중·고등학교 영어교사들의 말하기 수행평가의 시행한계성은 말하기 수행평가의 전문성과 서로 음의 상관관계를 가지는 것으로 나타났다. 특히, 학교급별로 두 요인들의 상관관계를 분석했을 때, 중학교와 고등학교 영어교사들 두 집단 모두 말하기 수행평가의 시행한계성과 말하기 수행평가의 전문성은 서로 음의 상관관계를 가지지만 통계적으로 유의미하지는 않았다.

셋째, 전체 중·고등학교 영어교사들의 말하기 수행평가의 시행한계성과 효용성은 서로 음의 상관관계에 있는 것으로 나타났다. 특히, 학교급별로 두 요인들의 상관관계를 분석해 본 결과, 고등학교 영어교사들에게 있어서 말하기 수행평가의 시행한계성과 말하기 수행평가의 효용성은 서로 음의 상관관계를 가지면서 통계적으로 유의미한 것으로 밝혀졌다. 그러나 중학교 영어교사들의 경우 말하기 수행평가의 시행한계성과 말하기 수행평가의 효용성은 서로 음의 상관관계에 있으나 통계적으로 유의미하지는 않았다.

이러한 결과들을 종합해 보면, 전체 중·고등학교 영어교사들 뿐만 아니라 중학교와 고등학교 영어교사들 두 집단 모두 말하기 수행평가에 대한 긍정적인 효용성이 높다고 생각할수록 말하기 수행평가의 전문성에 대한 인식이

높았고, 또한 말하기 수행평가의 전문성에 대한 인식이 높을수록 말하기 수행평가가 지닌 긍정적인 효용성 역시 높다고 생각하였다. 그리고 전체 중·고등학교 영어교사들은 말하기 수행평가의 시행에 있어서 어려움이 많다고 생각할수록 말하기 수행평가의 전문성에 대한 인식이 낮았고, 또한 말하기 수행평가의 전문성에 대한 인식이 높을수록 말하기 수행평가의 시행이 어렵다는 생각을 적게 하였다. 게다가 전체 중·고등학교 영어교사들은 말하기 수행평가의 시행에 있어서 어려움이 많다고 생각할수록 말하기 수행평가가 지닌 효용성이 낮은 것으로 인식하였고, 또한 말하기 수행평가의 효용성이 높다고 인식할수록 말하기 수행평가의 시행이 어렵다는 생각을 적게 하였다. 특히, 이러한 말하기 수행평가의 효용성과 시행한계성이 서로 음의 상관관계를 가지고 있다는 인식은 고등학교 영어교사들에게 있어서 뚜렷하였다.

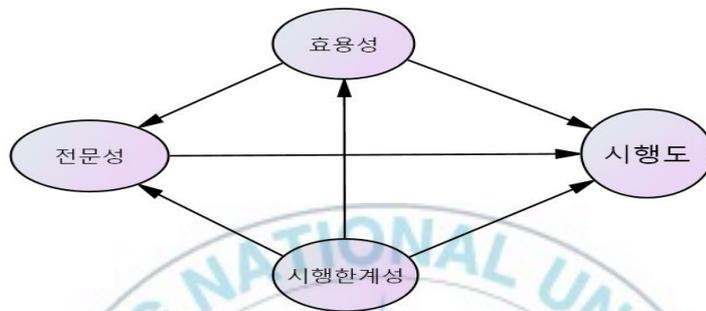
5.4 말하기 수행평가의 시행에 대한 구조방정식모형

구조방정식모형은 요인 분석과 경로 분석을 결합한 다차원적인 분석 방법으로서, 여러 개의 종속 변인들을 하나의 차원에서 분석하기 위해서 측정 지표를 통해서 잠재 변인을 형성하고 잠재 변인들 간의 인과 관계를 파악하는 방법이다(문수백, 2009). 본 절에서는 중·고등학교 영어교사들이 인식하는 말하기 수행평가의 전문성, 효용성, 시행한계성 및 시행도라는 네 가지 변인들이 서로 어떠한 구조적 관계를 가지는지를 파악하고자 한다. 연구자는 이러한 네 가지 잠재 변인들의 이론적 관계를 증명해 주는 구조방정식모형에 대한 정규성 검정과 적합도 검증을 한 후 이들 간의 인과 관계와 직·간접 효과를 분석하였다.

5.4.1 이론적 모형

앞 절에서 말하기 수행평가의 전문성, 효용성 및 시행한계성이라는 주요 변인들은 서로 유의미한 상관관계에 있는 것으로 밝혀졌다. 말하기 수행평가의 전문성은 말하기 수행평가의 효용성과 서로 정의 상관관계를 가졌으나, 말하기 수행평가의 시행한계성과는 서로 음의 상관관계를 가지고 있었다. 그리고 말하기 수행평가의 효용성은 말하기 수행평가의 시행한계성과 서로 음의 상관관계를 가지고 있었다. 이 결과는 EFL 상황에서 교사의 평가 시행에 영향을 주는 요인들과 교사의 실제적 평가 시행과의 상관관계를 다룬 Yang(2008)의 연구 결과와 거의 일치하고 있다.

이러한 연구 결과를 바탕으로 연구자가 세운 본 연구 과제의 기본 가설은 영어교사들의 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 인식은 말하기 수행평가의 시행도에 영향을 미친다는 것이다. 이를 검증하기 위해서 측정 오차를 통제할 수 있고 매개 변인의 사용이 용이하며, 또한 이론적 모형에 대한 통계적 평가가 가능하다는 장점을 가진 구조방정식모형을 이용하였다. 말하기 수행평가의 전문성, 효용성, 시행한계성 및 시행도가 각각 잠재 변인이고, 이 잠재 변인들은 각 하위 설문 문항인 측정 변인들에 의해 구성되었다. 따라서 말하기 수행평가와 관련된 이론적 배경에 기초하여 연구자가 설정한 말하기 수행평가의 전문성, 효용성, 시행한계성 및 시행도라는 잠재 변인들 간의 인과 관계를 나타낸 이론적 모형은 <그림 13>과 같다.



<그림 13> 잠재 변인들 간의 이론적 모형

<그림 13>에서 보듯이, 본 연구 과제의 기본 가설에 따른 영어교사들의 말하기 수행평가의 전문성, 효용성, 시행한계성 및 시행도라는 잠재 변인들 간의 인과 관계를 설명하는 경로는 다음과 같다. 첫째, 영어교사들의 말하기 수행평가의 전문성에 대한 인식은 말하기 수행평가의 시행에 직접적인 영향을 미칠 것이다. 이는 자신의 평가 능력에 대한 인식과 자신감은 교사의 수업 평가에 아주 큰 영향을 미친다는 연구(Bandura, 1977; Yang, 2008)와 교사의 평가 전문성 관련 연수는 평가의 시행과 긍정적으로 유의미한 관계를 지닌다는 연구(Borko, Flory & Cumbo, 1993; Green & Stager, 1987)의 결과에 근거한다. 둘째, 말하기 수행평가의 효용성에 대한 영어교사들의 긍정적인 인식은 말하기 수행평가의 시행에 직접적인 영향을 미치거나, 혹은

말하기 수행평가의 전문성에 대한 인식을 매개로 하여 말하기 수행평가의 시행에 간접적으로 영향을 미칠 것이다. 이는 평가의 교육적 장점에 대한 교사의 신념과 교사의 평가 시행은 긍정적으로 유의미한 관계가 있다는 연구(Chan, 2007; Vitali, 1993; Yang, 2008)의 결과에 근거한다. 셋째, 말하기 수행평가의 시행한계성에 대한 영어교사들의 인식은 말하기 수행평가의 시행에 직접적인 영향을 미치거나, 혹은 말하기 수행평가의 전문성과 효용성에 대한 인식을 매개로 하여 말하기 수행평가의 시행에 간접적으로 영향을 미칠 것이다. 이는 교사의 과도한 업무, 시간적 제약, 평가를 위한 교실 상황 및 학부모의 압력 등과 같은 많은 교육적 상황 요소들로 인하여 평가 과제의 수행이 어렵다는 교사의 인식이 평가의 실제적인 시행과 부정적인 상관관계를 가진다는 연구(Shih & Wang, 2010; Yang, 2008)의 결과에 근거한다. 따라서 연구자는 이러한 말하기 수행평가의 시행과 관련한 선행 연구들의 결과와 앞 절에서 논의된 잠재 변인들 간의 상관관계에 대한 연구의 결과를 토대로 하여 주요 변인들 간의 이론적 관계를 모색한 후 최종적인 구조방정식모형을 설정하였다.

5.4.2 구조방정식모형의 정규성 검정

배병렬(2011)은 구조방정식모형과 같은 다변량 분석 방법에서 각 변인의 정상분포조건³⁾이 충족되지 않을 경우 왜곡된 결과가 도출될 수 있다고 하였다. <표 56>은 연구자가 설정한 말하기 수행평가에 대한 구조방정식모형의 정규성을 검정하기 위해서 각 변인별 평균, 표준편차, 왜도(skewness) 및 첨도(kurtosis)를 나타낸 것이다. 모든 변인들의 왜도 기준값이 2를 넘지 않고

3) 변인들의 정상분포조건: 왜도<2, 첨도<4

침도 기준값이 4를 초과하지 않기 때문에 본 연구의 구조방정식모형을 적용하는 데 필요한 정상분포조건을 충족시키고 있음을 알 수 있다.

<표 56> 각 변인별 평균, 표준편차, 왜도 및 침도(N=243)

잠재변인	측정변인	평균	표준편차	왜도	침도
평가 방법과 도구의 선정 및 개발	선정개발1	3.52	.789	-.341	.141
	선정개발2	3.70	.785	-.500	.250
	선정개발3	3.47	.933	-.522	-.249
	선정개발4	3.37	.920	-.267	-.174
전문성	실시채점1	4.63	.564	-1.484	2.744
	실시채점2	4.46	.669	-1.358	3.022
	실시채점3	4.50	.592	-.836	.376
	실시채점4	4.48	.612	-.840	.277
평가 결과의 해석 및 활용	해석활용1	3.85	.737	-.564	1.126
	해석활용2	3.66	.810	-.477	.290
	해석활용3	3.69	.852	-.491	.092
	해석활용4	3.48	.892	-.155	-.265
효용성	효용1	3.44	.792	-.141	-.224
	효용2	3.48	.820	-.358	-.098
	효용3	3.61	.817	-.277	-.159
	효용4	3.41	.810	-.168	-.110
	효용5	3.19	.856	-.016	-.060
시행한계성	시행한계1	3.85	.897	-.674	.164
	시행한계2	3.62	.935	-.575	.179
	시행한계3	3.42	.934	-.286	-.317
	시행한계4	3.27	.885	-.191	-.289
	시행한계5	3.23	.935	-.209	-.307
시행도	시행횟수	1.23	.475	1.991	3.239
	시행방법	1.40	.637	1.369	.683

5.4.3 구조방정식모형의 적합도 검증

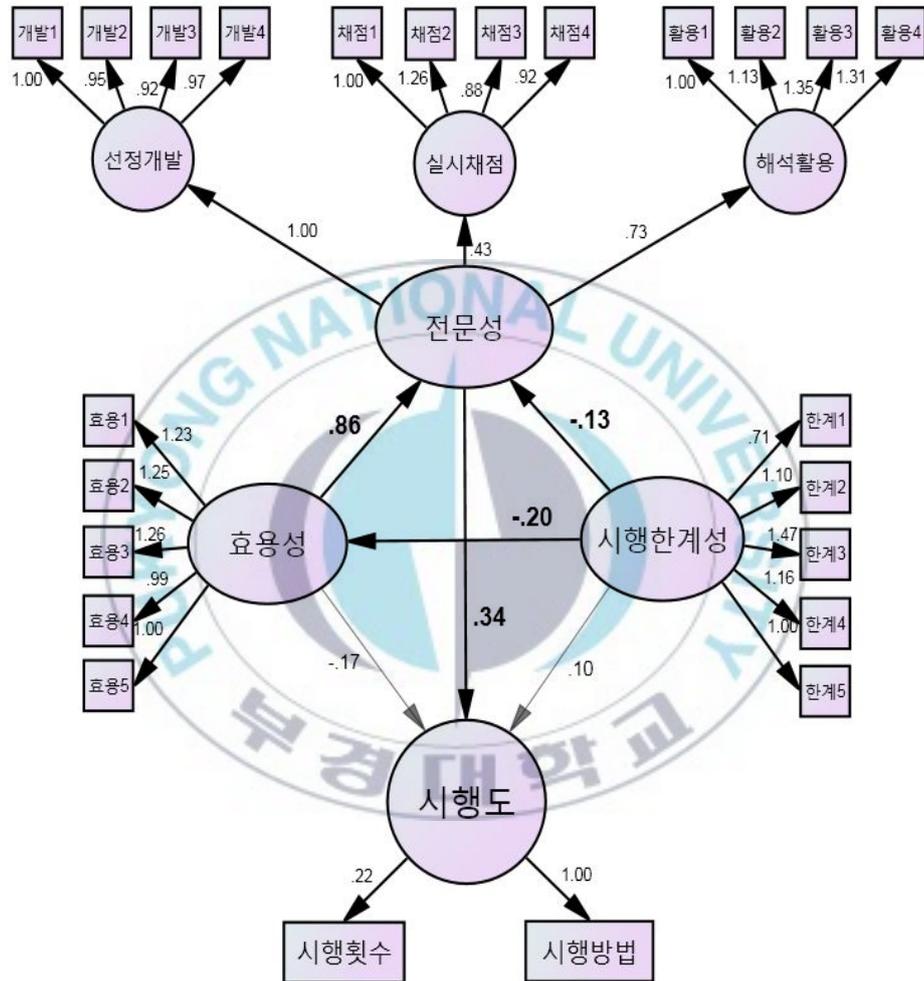
일반적으로 구조방정식모형의 적합도를 평가하는 방법에는 χ^2 검증을 이용하는 방법과 기타 적합도 지수를 이용하는 방법이 있다. 그러나 χ^2 검증은 표본 크기에 민감하고 영가설이 상당히 엄격하므로, χ^2 검증에 전적으로 의존하여 모형을 평가하지는 않는다(김주환 외, 2009). 본 연구에서는 표본 크기에 민감하지 않고 모형의 간명성을 고려하였으며, 또한 적합도 평가 지수의 기준이 확립된 CFI(Comparative Fit Index), TLI(Tucker-Lewis Index) 및 RMSEA(Root Mean Square Error of Approximation)를 통해서 구조방정식모형의 적합도를 평가하였다. Browne과 Cudeck(1993)에 의하면, 모형의 RMSEA 값이 .05이하이면 좋은 적합도이고 .05에서 .08사이이면 적합한 적합도이며, .10이상이면 부적절한 적합도이다. 또한, CFI와 TLI의 경우 그 값이 .90이상이면 적합도가 좋다고 할 수 있다(Bentler, 1990; Tucker & Lewis, 1973). 따라서 <표 57>에서 나타난 바와 같이, 본 연구의 구조방정식모형의 적합도는 $\chi^2=338.3$, $df=240$, $CFI=.966$, $TLI=.961$, 그리고 $RMSEA=.041$ 로서 꽤 만족할 만한 수준임을 알 수 있다.

<표 57> 구조방정식모형의 적합도

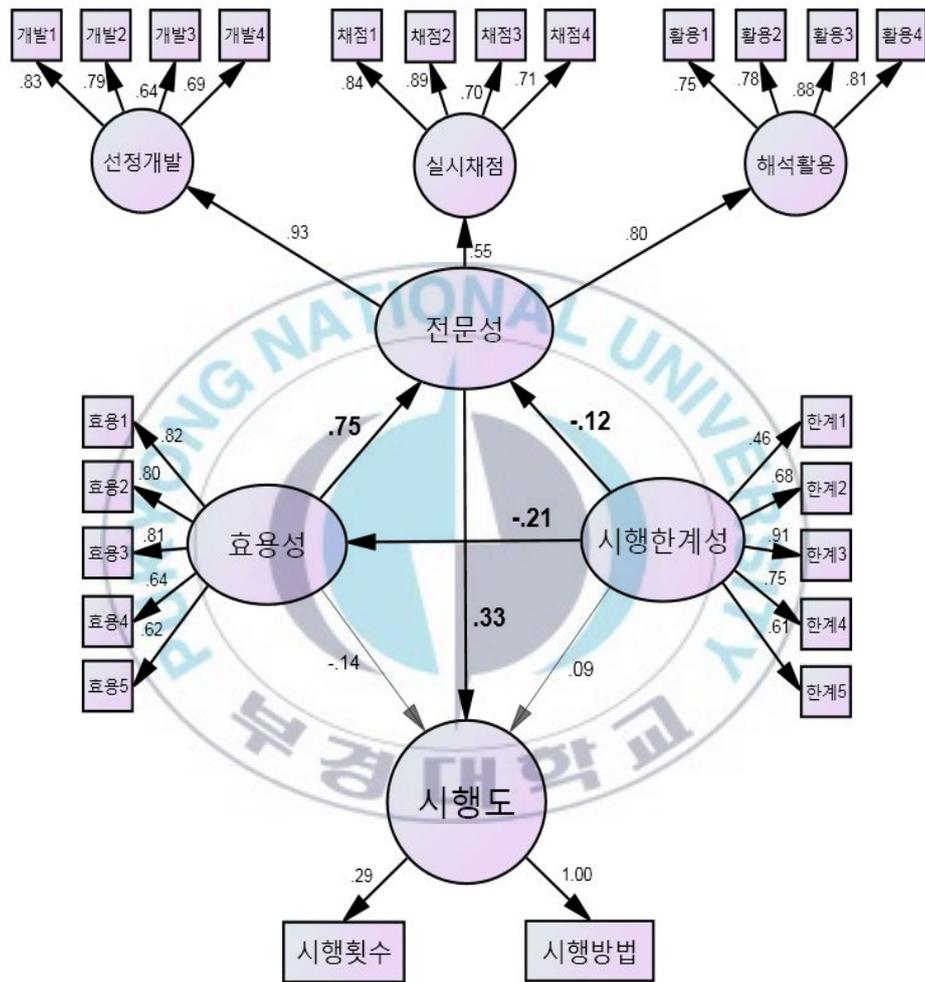
χ^2	df	CFI	TLI	RMSEA
338.3	240	.966	.961	.041

5.4.4 구조방정식모형의 해석

<그림 14>는 구조방정식모형을 통해서 추정된 모든 변인들의 비표준화된 경로계수를 나타낸 것이다. 이어서 <그림 15>는 구조방정식모형을 통해서 추정된 모든 변인들의 표준화된 경로계수를 보여주고 있다. <그림 15>에서 알 수 있듯이, 말하기 수행평가의 전문성에 대한 인식이 말하기 수행평가의 시행도에 미치는 영향은 통계적으로 유의미하며 그 직접 효과(.33)는 중간 정도를 나타내고 있다. 그리고 말하기 수행평가의 효용성에 대한 인식이 말하기 수행평가의 시행도에 미치는 영향은 통계적으로 유의미하지 않지만, 말하기 수행평가의 전문성에 대한 인식에 미치는 영향은 통계적으로 유의미하며 그 직접 효과(.75)는 아주 큰 정도를 나타내고 있다. 아울러 말하기 수행평가의 시행한계성에 대한 인식이 말하기 수행평가의 시행도에 미치는 영향은 통계적으로 유의미하지 않지만, 말하기 수행평가의 전문성에 대한 인식에 미치는 영향은 통계적으로 유의미하며 그 직접 효과(-.12)는 작은 정도를 나타내고 있다. 또한, 말하기 수행평가의 시행한계성에 대한 인식이 말하기 수행평가의 효용성에 대한 인식에 미치는 영향은 통계적으로 유의미하며 그 직접 효과(-.21)가 어느 정도 있음을 나타내고 있다.



<그림 14> 구조방정식모형(비표준화)



<그림 15> 구조방정식모형(표준화)

말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 인식이라는 잠재 변인들이 말하기 수행평가의 시행도라는 잠재 변인에 영향을 미친다는 기본 가설에 입각하여 구조방정식모형을 통한 잠재 변인들 사이의 인과 관계를 규명하였다. 그 결과에 따른 잠재 변인들에 대한 경로계수의 추정치는 <표 58>에 제시하였다.

<표 58> 잠재 변인들 간의 경로 분석 결과

경로 및 방향		비표준화 계수	표준화 계수	유의확률
전문성	→ 시행도	.34	.33	.021
효용성	→ 전문성	.86	.75	.000
	→ 시행도	-.17	-.14	.287
시행한계성	→ 전문성	-.13	-.12	.041
	→ 효용성	-.20	-.21	.005
	→ 시행도	.10	.09	.203

<표 58>을 보면, 우선 말하기 수행평가의 전문성은 말하기 수행평가의 시행도에 영향을 미치며 그 정도가 통계적으로 유의미한 것으로 나타났다. 이는 말하기 수행평가의 전문성에 대한 인식이 높은 중·고등학교 영어교사들일수록 말하기 수행평가를 시행하는 정도가 높음을 보여준다.

말하기 수행평가의 효용성은 말하기 수행평가의 전문성에 영향을 미치며 그 정도가 통계적으로 유의미한 것으로 나타났다. 이는 중·고등학교 영어교사들의 말하기 수행평가의 효용성에 대한 긍정적인 인식이 높을수록 말하기 수행평가의 전문성에 대한 인식도 높음을 보여준다. 한편, 말하기 수행평가의 효용성은 말하기 수행평가의 시행도에 부적인 영향을 미치는 것처럼 보이지만 통계적으로 유의미하지는 않았다. 이는 말하기 수행평가의 효용성에 대한 영어교사들의 인식이 긍정적이라고 해서 반드시 말하기 수행평가를 시행하는 것이 아닐 수도 있음을 암시한다.

말하기 수행평가의 시행한계성은 말하기 수행평가의 전문성과 효용성에 부적 인 영향을 미치며 그 정도가 통계적으로 유의미한 것으로 나타났다. 이는 말하기 수행평가를 시행하는 데 있어서 어려움이 많다고 생각하는 중·고등학교 영어교사들일수록 말하기 수행평가의 전문성에 대한 인식이 낮고, 또한 말하기 수행평가의 효용성에 대한 긍정적인 인식도 낮음을 보여준다. 한편, 말하기 수행평가의 시행한계성은 말하기 수행평가의 시행도에 영향을 미치는 것처럼 보이지만 통계적으로 유의미하지는 않았다. 즉, 중·고등학교 영어교사들은 말하기 수행평가의 시행이 어렵다고 인식은 하지만, 각 관할 교육청에서 말하기 수행평가를 내신 성적에 반영할 것을 권고함으로 인해서 아마도 말하기 수행평가를 형식적으로 시행하고 있는 이유와 연관이 있을 것으로 추측된다.

다음으로 구조방정식모형을 통해서 중·고등학교 영어교사들의 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 인식이 그들의 말하기 수행평가의 시행도에 미치는 영향을 직접 효과와 간접 효과 및 전체 효과를 분석하였고, 그 결과는 <표 59>에 제시하였다. 보통 변인들 간의 상대적 효과 크기(effect size)를 비교할 수 있는 표준화된 회귀계수는 그 절대값의 크기가 .10보다 작으면 '작은' 효과를 나타내고, .30은 '중간' 정도의 효과를, 그리고 .50 이상은 '큰' 효과를 나타내는 것으로 볼 수 있다(Kline, 1998).

<표 59> 잠재 변인들 간의 직·간접 효과 및 전체 효과

경로 및 방향		직접 효과	간접 효과	전체 효과
전문성	→ 시행도	.33	.00	.33
효용성	→ 전문성 → 시행도	-.14	.24	.10
	→ 전문성		-.04	
시행 한계성	→ 효용성 → 시행도	.09	.03	.03
	→ 효용성 → 전문성		-.05	

<표 59>는 말하기 수행평가의 전문성, 효용성 및 시행한계성이라는 잠재 변인들이 말하기 수행평가의 시행도라는 잠재 변인에 미치는 직·간접적인 영향에 대한 효과의 크기를 보여주고 있다. 첫째, 말하기 수행평가의 전문성이 말하기 수행평가의 시행도에 미치는 전체 효과(.33)는 중간 정도인 것으로 나타났다. 즉, 말하기 수행평가의 전문성에 대한 인식이 높은 중·고등학교 영어교사들은 말하기 수행평가를 적극적으로 시행하려는 의지가 높음을 알 수 있다.

둘째, 말하기 수행평가의 효용성이 말하기 수행평가의 시행도에 미치는 전체 효과(.10)는 직접 효과(-.14)에 말하기 수행평가의 전문성을 매개로 하는 간접 효과(.24)⁴⁾를 합친 것으로서, 그 효과의 크기는 작은 것으로 나타났다. 이 결과를 통해서 중·고등학교 영어교사들은 말하기 수행평가의 효용성에 대해서 긍정적으로 생각하더라도, 말하기 수행평가의 전문성에 대한 인식이 높을 때만 말하기 수행평가를 시행하려는 의지가 다소 있음을 보여주고 있다.

셋째, 말하기 수행평가의 시행한계성이 말하기 수행평가의 시행도에 미치는 전체 효과(.03)는 직접 효과(.09)에 말하기 수행평가의 전문성을 매개로 하는 간접 효과(-.04)와 말하기 수행평가의 효용성을 매개로 하는 간접 효과(.03) 및 말하기 수행평가의 효용성과 전문성 둘 다 매개로 하는 간접 효과(-.05)를 모두 합친 것으로서, 그 효과의 크기는 아주 작은 것으로 나타났다. 이 결과를 통해서 현재 중·고등학교 영어교사들은 말하기 수행평가의 시행에 있어서 많은 어려움이 있다고 인식하더라도, 말하기 수행평가의 전문성에 대한 인식이 높으면서 동시에 말하기 수행평가의 효용성에 대한 긍정적인 인식 또한 높을 경우에만 말하기 수행평가를 시행하려는 최소한의 의지가 있음을 보여주고 있다.

4) 간접 효과(.24) = [효용성→전문성(.75)] × [전문성→시행도(.33)]

이상과 같이 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 중·고등학교 영어교사들의 인식이 말하기 수행평가의 시행도에 미치는 영향과 그 효과를 알아보았다. 그 결과 말하기 수행평가의 전문성에 대한 인식이 높은 중·고등학교 영어교사들은 말하기 수행평가를 적극적으로 시행하려는 의지가 높았다. 그리고 말하기 수행평가의 효용성에 대한 긍정적인 인식이 높은 중·고등학교 영어교사들일지라도 말하기 수행평가의 전문성에 대한 인식이 높을 때만 말하기 수행평가를 시행하려는 의지가 있는 것으로 나타났다. 더구나 중·고등학교 영어교사들은 말하기 수행평가의 시행을 가로막는 어떠한 어려움을 인식하더라도, 말하기 수행평가의 전문성에 대한 인식이 높을 뿐만 아니라 그 효용성에 대한 긍정적인 인식 또한 높을 경우에만 말하기 수행평가를 시행하려는 최소한의 노력을 보이고 있음을 알 수 있었다. 이러한 본 연구의 결과는 자신의 평가 능력에 대한 인식, 평가의 교육적 장점에 대한 교사의 신념 및 평가 시행의 어려움에 대한 교사의 인식이라는 요인들과 교사의 실제적인 평가 시행과의 관계에 관한 Yang(2008)의 연구 결과와 아주 상응한 것으로 나타났다.

VI. 결론 및 제언

지금까지 말하기 수행평가를 시행하고 있는 우리나라 중·고등학교 영어교사들의 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 인식 차이를 살펴보았다. 그리고 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 중·고등학교 영어교사들의 인식이 말하기 수행평가의 시행도에 미치는 영향에 대해서도 알아보았다. 이번 장에서는 본 연구의 결과를 정리하고 이를 바탕으로 우리나라 영어교육의 현장에 적용할 수 있는 몇 가지 제언을 하고자 한다.

6.1 연구의 결론

본 연구는 말하기 수행평가를 시행하는 우리나라 중·고등학교 영어교사들이 지닌 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 인식 차이를 비교하고, 또한 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 영어교사들의 인식이 말하기 수행평가를 시행하는 데 있어서 어떠한 영향을 미치는지를 연구하는 것이 목적이었다. 이에 따라 문헌 연구와 설문조사 및 질적 자료를 바탕으로 수행된 본 연구의 결과는 다음과 같다.

첫째, 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 중학교 영어교사들과 고등학교 영어교사들의 인식 차이를 비교한 결과, 중학교 영어교사들과 고등학교 영어교사들 두 집단 모두 말하기 수행평가의 교육적 효용성에 대해서 긍정적으로 인식함에도 불구하고 중학교 영어교사들보다 고등학교 영어교사들의 말하기 수행평가의 전문성에 대한 인식이 더 낮았다. 더

구나 고등학교 영어교사들은 말하기 수행평가를 시행하는 일에 있어서 중학교 영어교사들보다 더 많은 어려움을 인식하였는데, 그 이유는 학급 규모가 크고 학생 수가 너무 많거나 또는 채점 시 객관성 확보가 어렵다는 점 때문이었다.

둘째, 말하기 수행평가를 시행하는 중·고등학교 영어교사들의 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 인식이 교직 경력, 관련 연수 이수 및 말하기 능력이라는 배경 변인에 따라 상이한 것으로 나타났다. 그 결과 교직 경력이 11~15년인 영어교사들은 교직 경력이 16~20년인 영어교사들보다 말하기 수행평가를 실시하고 채점하는 전문적인 능력에 대한 인식이 훨씬 높았다. 그리고 말하기 수행평가에 관한 연수를 전혀 받지 않은 영어교사들보다 말하기 수행평가에 관한 연수를 최소 1회나 2회 이상 받은 영어교사들의 말하기 수행평가의 전문성에 대한 인식이 높을 뿐만 아니라 말하기 수행평가의 효용성에 대한 긍정적인 인식 또한 높았다. 또한, 자신의 영어 말하기 능력이 상대적으로 우수하다고 생각하는 영어교사들은 말하기 수행평가의 전반적인 전문성에 대한 인식도 높았고, 말하기 수행평가의 효용성에 대해서도 긍정적인 인식을 가지고 있었다.

셋째, 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 중·고등학교 영어교사들의 인식이 서로 어떠한 상관관계를 가지는지를 살펴본 결과, 중·고등학교 영어교사들은 말하기 수행평가의 긍정적인 효용성이 높다고 생각할수록 말하기 수행평가의 전문성에 대한 인식이 높았고, 또한 말하기 수행평가의 전문성에 대한 인식이 높을수록 말하기 수행평가의 긍정적인 효용성도 높다고 생각하였다. 그리고 말하기 수행평가의 시행한계성이 높다고 인식하는 중·고등학교 영어교사들일수록 말하기 수행평가의 전문성에 대한 인식이 낮았고, 또한 말하기 수행평가의 전문성에 대한 인식이 높은 중·고등학교 영어교사들일수록 그들이 인식하는 말하기 수행평가의 시행한계성은 낮

았다. 게다가 말하기 수행평가의 시행한계성이 높다고 인식하는 중·고등학교 영어교사들일수록 말하기 수행평가의 긍정적인 효용성이 낮은 것으로 생각하였고, 또한 말하기 수행평가의 긍정적인 효용성이 높다고 생각하는 영어교사들일수록 그들이 인식하는 말하기 수행평가의 시행한계성은 낮았다.

마지막으로, 말하기 수행평가의 전문성, 효용성 및 시행한계성에 대한 중·고등학교 영어교사들의 인식이 말하기 수행평가의 시행도에 미치는 직·간접적인 영향과 그 효과를 확인할 수 있었다. 말하기 수행평가의 전문성에 대한 인식이 높은 중·고등학교 영어교사들은 말하기 수행평가를 적극적으로 시행하려는 의지가 강하였다. 그리고 중·고등학교 영어교사들은 말하기 수행평가의 효용성에 대한 인식이 긍정적일지라도 그들의 말하기 수행평가의 전문성에 대한 인식이 높을 경우에만 말하기 수행평가에 대한 의지가 있음을 보여주었다. 또한, 중·고등학교 영어교사들은 말하기 수행평가를 시행하는 데 있어서 어느 정도의 시행한계성을 인식하고 있더라도, 그들의 말하기 수행평가의 긍정적인 효용성에 대한 인식이 높으면서 말하기 수행평가의 전문성에 대한 인식 또한 높을 경우에만 말하기 수행평가를 시행하려는 최소한의 의지가 있음을 보여주었다.

6.2 제언

본 연구의 결과에 근거하여 연구자는 우리나라의 영어교육 현장에 다음과 같이 몇 가지의 발전적인 제언을 하고자 한다. 첫째, 중학교 영어교사들보다 고등학교 영어교사들의 말하기 수행평가의 전문성에 대한 인식이 낮은 점에 대해서 영어 말하기보다는 읽기에 초점이 맞추어진 현 고등학교 교육과정의 획기적인 변화가 요구된다. 고등학교 영어교사들은 말하기 평가를 반영하지

않는 대학 입학시험에 대한 학생과 학부모의 현실적인 기대와 요구로 인하여 정규 수업에서 말하기 수업은 물론이고 진정한 말하기 수행평가를 시행하기 어려운 상황이다. 이는 바로 실용영어를 가르치는 중학교 영어교사들에 비해서 상대적으로 고등학교 영어교사들의 말하기 수행평가에 대한 전문성이 떨어지는 원인으로 볼 수 있다. 그러므로 2009 개정 영어과 교육과정에서도 명시되어 있듯이, 국제화 시대에 필요한 의사소통 능력의 함양이라는 과제를 제대로 실천하기 위해서는 먼저 국가 차원에서 현 대학 입학시험에 말하기 평가를 도입하는 등의 정책적인 결정이 선행되어야 할 것이다.

둘째, 중·고등학교 영어교사들의 말하기 수행평가의 시행에 가장 큰 장애물인 학급 규모가 크고 학생 수가 너무 많거나 또는 채점 시 객관성 확보가 어렵다는 점에 있어서 시·도 교육청이나 국가 차원에서 말하기 수행평가의 방법이나 도구 및 객관적인 기준안을 마련하는 일이 시급할 것이다. 그리고 가능하다면 정책적으로 한 학급의 학생 수를 감축하는 것도 말하기 수행평가의 원활한 시행과 신뢰성 보장에 도움이 될 것이다. 따라서 현직 영어교사들이나 영어교육 전문가들이 협의하여 바람직한 말하기 수행평가의 방법과 도구 및 기준을 구축하는 일이 필요하다. 그러면 현재 많은 학교에서 파행적으로 시행되고 있는 말하기 수행평가의 진정성에 대한 회의적인 인식이 어느 정도 줄어들 것으로 보인다.

셋째, 말하기 수행평가와 관련된 연수를 자주 이수한 중·고등학교 영어교사들이 말하기 수행평가에 대한 전문성과 효용성에 대한 인식 또한 높은 점에 있어서는 무엇보다도 말하기 수행평가와 관련된 연수를 연중 수시로 개최하여 영어교사들이 말하기 수행평가의 전문성을 신장시킬 수 있는 기회를 가급적 많이 접할 수 있도록 해야 할 것이다. 특히, 말하기 수행평가에 관한 연수의 기획 담당자나 운영자는 교사 연수 프로그램을 구성할 때 말하기 수행평가의 계획부터 준비와 시행 및 결과 활용 등 영어교사들이 갖추어야 할

모든 단계의 전문성 요소를 반드시 포함시켜야 할 것이다. 그리고 말하기 수행평가의 효용성과 전문성에 대한 인식이 서로 긍정적으로 높은 상관성을 지닌 점을 감안하여 말하기 수행평가의 교육적 효용성에 대한 연구나 실천 사례들을 연수 내용에 포함시켜서 말하기 수행평가가 교육 현장에 미치는 교육적 효과를 강조하고 학교 현장에 홍보하는 일도 중요하다.

넷째, 교직 경력에 따라 중·고등학교 영어교사들의 말하기 수행평가의 전문성에 대한 인식 차이가 나는 점에 있어서는 특히 16년 이상의 교직 경력을 지닌 영어교사들을 대상으로 말하기 수행평가의 전문성 신장에 관한 집중적인 연수가 필요할 것이다. 아무리 훌륭한 연수를 받고 학교 현장으로 돌아와더라도 시간이 흐르면서 영어교사들이 받은 연수의 교육적 효과는 점차 줄어들어 가는 경향이 있다. 그러므로 영어교사들을 위한 말하기 수행평가의 전문성에 관한 연수는 가능한 한 일정한 교직 경력에 따른 단계적 프로그램을 개발하여 보다 체계적으로 운영할 필요가 있다.

끝으로, 말하기 수행평가의 전문성에 대한 인식이 높고 말하기 수행평가의 효용성에 대한 긍정적인 인식을 지닌 중·고등학교 영어교사들은 어떠한 어려움이 있더라도 말하기 수행평가를 시행하려는 의지를 비치고 있었다. 다시 말해, 말하기 수행평가의 교육적 장점이 많다고 생각하는 영어교사들은 말하기 수행평가에 대한 전문성 또한 뛰어나기 때문에, 아무리 학교 현장에서 말하기 수행평가의 시행을 가로막는 현실적인 한계점들이 있더라도 이를 극복할 수 있는 탁월한 능력과 자질을 갖추고 있다. 그리고 국제화 시대의 주역이 될 우리나라 학생들의 영어 말하기 능력의 향상에 큰 주춧돌의 역할을 하는 사람은 바로 학생들의 말하기 수행평가를 책임지고 시행하는 영어교사들이다. 따라서 영어교사들의 말하기 수행평가에 대한 전문성 신장과 더불어 말하기 수행평가의 효용성에 대한 영어교사들의 긍정적인 인식 향상을 위해서는 우선적으로 국가 차원에서 말하기 수행평가의 시행과 관련된 많은 한

계점들을 하나씩 해결하여 바람직한 평가 환경을 조성할 필요가 있다. 그렇게 된다면 학교 현장에서 영어교사들이 진정한 의사소통 중심의 말하기 수업과 수행평가를 원활하게 시행할 수 있게 되고, 나아가 영어교사들의 말하기 수행평가에 대한 긍정적인 인식 또한 점점 높아질 것으로 사료된다.

이처럼 본 연구의 결과를 통해서 연구자는 우리나라 학생들의 영어 의사소통 능력을 향상시키기 위해서는 학교 현장에서 영어교사들이 말하기 수행평가를 자주 시행하는 일이 효과적이라는 사실을 밝히고자 한다. 비록 현재 우리나라의 학교 상황으로는 말하기 수행평가를 시행하는 데 있어서 많은 어려움이 있지만, 지금까지 연구된 말하기 수행평가의 교육적 효용성을 감안했을 때 말하기 수행평가의 적극적인 시행은 궁극적인 영어교육의 정상화를 위해서도 옳은 일이다. 이를 위해서는 대학 입학시험에서 말하기 평가의 반영에 대한 정책적인 결정, 말하기 수행평가를 위한 객관적인 방법과 도구 및 채점 기준의 마련, 말하기 수행평가의 전문성과 효용성을 강화한 연수 프로그램의 개발, 일정한 교직 경력의 단계에 따른 체계적인 말하기 수행평가에 관한 연수 프로그램의 운영, 그리고 국가 차원에서의 말하기 수행평가의 시행을 위한 환경 조성 등을 제언하고자 한다.

참고문헌

- 고현우. (2014). 중등 영어교사들의 말하기 수행평가의 효용성과 시행한계성에 대한 인식 연구. *새한영어영문학*, 56(3), 43-66.
- 교육과학기술부. (2008). *중학교 교육과정 해설: 외국어과*. 서울: 교육과학기술부.
- (2011). *영어과 교육과정*. 교육과학기술부 고시 제 2011-361호. 서울: 교육과학기술부.
- 교육부. (1998). *교육비전 2002 새학교 문화 창조*. 서울: 교육부.
- 김광현. (2014.09.07). 수업 내내 문법만... 말 못하는 학교 영어교육. *SBS 뉴스*.
월드와이드웹: http://news.sbs.co.kr/news/endPage.do?news_id=N1002573544
에서 2014년 9월 14일 검색.
- 김미은. (2002). *고등학교 영어과 수행평가 실태에 관한 연구: 말하기를 중심으로*. 석사학위논문, 국민대학교.
- 김성연, 이호경. (2013). 중학교 현장의 영어 말하기·쓰기 수행평가 적용 사례 연구. *Studies in English Education*, 18(2), 159-183.
- 김수동, 이의갑, 김경희, 김선희, 박은아, 신명선, 김수진, 박가나, 서수현, 전영석. (2005). *교사의 학생평가 전문성 신장 연구(II)*. 한국교육과정평가원 연구보고 RRE 2005-3. 서울: 한국교육과정평가원.
- , 이의갑. (2005). *영어 수업에서 학생평가를 잘 하려면*. 한국교육과정평가원 연구자료 ORM 2005-51-6. 서울: 한국교육과정평가원.
- 김신영. (2002a). 현장교사의 평가 전문성 연구. *교육평가연구*, 15(1), 151-171.
- (2002b). 교실 내 평가의 특성과 교사의 평가 전문성. *한국외국어대학교 논문집*, 34, 541-558.
- 김인철. (1998). 의사소통능력향상을 위한 언어수행평가. *명지영어영문학*, 3(2), 97-112.

- 김정임. (2006). *말하기 수행평가가 의사소통 능력에 미치는 영향*. 석사학위논문, 경남대학교.
- 김주환, 김민규, 홍세희. (2009). *구조방정식모형으로 논문 쓰기*. 서울: 커뮤니케이션북스.
- 김진철. (1998). *초등영어 교수법*. 서울: 학문출판사.
- 문수백. (2009). *구조방정식모델링의 이해와 적용*. 서울: 학지사.
- 문지형. (2007). *중등학교 영어과 수행평가 실태 연구*. 석사학위논문, 홍익대학교.
- 박상수. (2004). *제 7차 교육과정에 의한 영어 수행평가 실태와 문제점 연구: 말하기를 중심으로*. 석사학위논문, 아주대학교.
- 박소현. (2010). *고등학교의 영어 말하기 수행평가의 문제점과 학생 인식 조사 연구*. 석사학위논문, 중앙대학교.
- 박순경, 강창동, 김경희, 이광우, 이미숙, 손민호. (2003). *제 7차 초·중등학교 교육과정 평가연구(III)*. 한국교육과정평가원 연구보고 RRC 2003-2. 서울: 한국교육과정평가원.
- 방인자, 천세영. (2011). 내신제도에 따른 고등학교 영어 교육과정 운영실태 분석. *교육연구논총*, 32(1), 1-20.
- 배병렬. (2011). *Amos 19 구조방정식 모델링: 원리와 실제*. 서울: 청람출판사.
- 배원기. (2004). *중학교 말하기 수행평가가 학습자의 정의적 요인과 학업성취도에 미치는 영향*. 석사학위논문, 울산대학교.
- 배호순. (2000). *수행평가의 이론적 기초*. 서울: 학지사.
- 성태제. (2007). *SPSS/AMOS를 이용한 알기 쉬운 통계 분석*. 서울: 학지사.
- 염지연. (2011). *중학교 영어 말하기 수행평가의 채점자 신뢰도 및 학생과 교사의 인식*. 석사학위논문, 고려대학교.
- 윤말임. (2001). *수행평가가 영어 학습태도에 미치는 영향: 말하기 과제 활동 중심으로*. 석사학위논문, 서울대학교.

- 윤미현. (2012). *중·고등학교 영어 말하기·쓰기 수업 현황과 국가영어능력평가 시험 개발에 대한 교사와 학생의 인식조사*. 석사학위논문, 건국대학교.
- 윤소영. (1999). *중학교 영어과에서 수행평가와 채점기준의 개발 및 적용에 관한 연구*. 석사학위논문, 이화여자대학교.
- 은경자. (2010). *TEE 영어회화 팀티칭에서의 말하기 활동을 통한 수행평가 연구: 고등학생 말하기 능력 향상을 중심으로*. 석사학위논문, 단국대학교.
- 이기수. (1999). 고등학교 공통영어 수행평가 계획 및 실제. *열린교육연구*, 7(1), 333-350.
- 이문복, 신동광, 조보경, 박태준, 송민영, 윤지환, 박지선. (2012). *고등학교 영어 말하기·쓰기 수업현황 분석과 정책 제언*. 한국교육과정평가원 연구자료 ORM 2012-16-2. 서울: 한국교육과정평가원.
- 이소영. (2008). 중·고등학교 영어교사의 수행평가 인식 및 운영실태 분석: 인천과 서울지역을 중심으로. *영어교육연구*, 20(4), 305-327.
- (2000). 중등영어 수행평가 실시 양상 및 효과. *Studies in English Education*, 5(1), 37-52.
- 이완기. (2012). *영어 평가 방법론* (3판). 서울: 문진미디어.
- 이은주. (2003). *중학교 영어 말하기 수업에 관한 교사 인식 및 수업 활동 조사*. 석사학위논문, 경희대학교.
- 이인제, 이범홍, 박정, 진재관, 김옥남, 서수현, 김신영. (2004). *교사의 학생 평가 전문성 신장 모형과 기준*. 한국교육과정평가원 연구보고 RRE 2004-5-2. 서울: 한국교육과정평가원.
- 이재호. (2012). *고등학교 영어과 수행평가의 실태와 개선 방안*. 석사학위논문, 영남대학교.
- 이정애. (2003). *중학교 교사의 학생평가에 대한 인식과 평가 전문성 연구*. 석사학위논문, 이화여자대학교.

- 이종윤. (2002). *수행평가에 대한 중학교 교사·학생·학부모의 인식 연구*. 석사학위논문, 연세대학교.
- 이태윤. (2012.03.20). 영어 원어민 교사 철수, 교육 불평등 논란 점화. *동아닷컴*.
월드와이드웹: <http://news.donga.com/3/all/20120319/44873910/1>
에서 2014년 5월 27일 검색.
- 이화자. (2006). 현직 교사교육이 교사들의 교수변화에 미치는 영향. *외국어교육*, 13(3), 283-320.
- 장인순. (2004). *영어과 수행평가 현황에 대한 연구 분석*. 석사학위논문, 아주대학교.
- 전병만, 박준언, 안병규, 오준일, 유제명, 이소영, 김신혜. (2005). 영어과 교사의 학생 평가 전문성 신장을 위한 연구. *영어교육*, 60(2), 3-27.
- 조영기. (2006). *초등학교 수행평가 실태분석 및 개선방안*. 서울: 한국교육정책연구소.
- 최연희. (2000). *영어과 수행평가의 이론과 실제: 대안적 평가의 제작*. 서울: 한국문화사.
- , 권오남, 성태제. (1998). *열린 교육을 위한 중학교 영어·수학 수행평가의 적용과 효과에 대한 분석 연구: 교육부 초등교육정책과 열린 교육 연구 결과 보고서*. 서울: 이화여자대학교.
- 최윤경. (2001). *제 7차 영어과 교육과정에 준한 수행평가 방안 연구*. 석사학위논문, 아주대학교.
- 최주경. (2009). *영어 말하기와 쓰기를 중심으로 한 발표 수행평가 활동에 관한 연구*. 석사학위논문, 서강대학교.
- 한국교육과정평가원. (1999). *중학교 영어과 수행평가 시행방안 및 자료개발 연구*. 서울: 한국교육과정평가원.
- (2012). *국가영어능력평가지험 대비 영어 말하기·쓰기 평가매뉴얼*

- (고등학교용). 연구자료 ORM 2012-67(CD). 서울: 한국교육과정평가원.
- 황정선. (2007). *중학교 영어 말하기 수업에 관한 학생과 교사의 인식 조사*. 석사학위논문, 부경대학교.
- American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). Standards for teacher competence in educational assessment of students. *Educational Measurement: Issues and Practice*, 9(4), 30-32.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- , & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191-215.
- Bentler, P. M. (1990). Comparative fit indices in structural models. *Psychological Bulletin*, 107, 238-346.
- Bol, L. (2004). Teachers' assessment practices in a high-stakes testing environment. *Teacher Education and Practice*, 17(2), 162-181.
- Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe and do. *Language Teaching*, 36(2), 81-109.
- Borko, H., Flory, M., & Cumbo, K. (1993). *Teachers' ideas and practices about assessment and instruction: A case study of the effects of alternative assessment in instruction, student learning and accountability practices*. ERIC Reproduction Service No. ED378226.
- Brookhart, S. M. (1997). A theoretical framework for the role of

- classroom assessment in motivating student effort and achievement. *Applied Measurement in Education*, 10(2), 161-180.
- Brown, D. (1994). *Principles of language learning and teaching* (3rd Ed.). Englewood Cliffs, NJ: Prentice Hall Regents.
- Brown, G., Bull, J., & Pendlebury, M. (1997). *Assessing student learning in higher education*. London: Routledge.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. *Sociological Methods Research*, 21(2), 230-258.
- Buck, G. (1988). Testing listening comprehension in Japanese university entrance examinations. *JALT Journal*, 10, 15-42.
- Busch, D. (1982). Introversion-extroversion and the ESL proficiency of Japanese students. *Language Learning*, 32, 109-132.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47.
- Chan, Y. C. (2007). *Elementary school EFL teachers' beliefs and practices of multiple assessments in the northern Taiwan*. National Science Council Research Project (NSC 92-2411-H-152-006). Taipei: National Science Council.
- Cheng, L. (1997). How does washback influence teaching?: Implications for Hong Kong. *Language and Education*, 11(1), 238-254.
- Clark, J. L. D. (1972). *Foreign language testing: Theory and practice*. Philadelphia, PE: The Center for Curriculum Development.
- Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd Ed.). Hillsdale, NJ: Lawrence Erlbaum.

- Cronbach, L. (1982). Issues in planning evaluation. In L. Cronbach (Ed.), *Designing evaluations of educational and social programs* (pp. 1-45). San Francisco, CA: Jossey-Bass.
- Day, R. (1984). Student participation in the ESL classroom or some imperfections in practice. *Language Learning*, 34, 69-102.
- Edelenbos, P., & Kubanek-German, A. (2004). Teacher assessment: The concept of “diagnostic competence.” *Language Testing*, 21(3), 259-283.
- Eisner, E. (2002). *The educational imagination: On the design & evaluation of school programs* (3rd Ed.). Upper Saddle River, NJ: Pearson Education.
- Ellis, J. (1993). Japanese students abroad: Relating language ability in class and in the community. *Thought Currents in English Literature*, 66, 45-82.
- Ely, C. (1986). An analysis of discomfort, risk-taking, sociability and motivation in the second language classroom. *Language Learning*, 36, 1-25.
- Gattullo, F. (2000). Formative assessment in ELT primary classrooms: An Italian case study. *Language Testing*, 17(2), 278-288.
- Green, K. E., & Stager, S. F. (1987). Testing: Coursework, attitudes, and practices. *Educational Research Quarterly*, 11(2), 48-55.
- Gronlund, N. E. (2006). *Assessment of student achievement* (8th Ed.). Boston, MA: Allyn & Bacon.
- Heaton, J. B. (1975). *Writing English language tests: A practical guide for teachers of English as a second or foreign language*.

- London: Longman.
- (1988). *Writing English Language Tests* (2nd Ed.). London: Longman.
- Heurta-Macias, A. (1995). Alternative assessment: Responses to commonly asked questions. *TESOL Journal*, 5(1), 8-11.
- Howie, S. J. (2006). *Assessment for learning strategies: Merging theory and practice within South Africa realities*. A paper presented at the GDE Assessment for learning conference, Johannesburg, South Africa.
- Hsu, Y. K. (2005). *Developing and researching multiple assessment: An action research*. National Science Council Individual Research Project (NSC 92-2411-003-031). Taipei: National Science Council.
- Jones, R. L. (1977). Testing: A vital connection. In J. K. Phillips (Ed.), *The language connection: From the classroom to the world* (pp. 237-265). Skokie, IL: National Textbook Company.
- Jonson, J. L. (1999). *Understanding barriers to teachers' use of alternative assessment*. Unpublished doctoral dissertation, The University of Nebraska-Lincoln.
- Jung, H. (2008). The washback effects of English listening test in Korean CSAT on secondary school English learning and teaching. *The Linguistic Association of Korea Journal*, 16(3), 143-162.
- Kearney, K. A., Hopkins, R. H., Mauss, A. L., & Weisheit, R., A. (1984). Self-generated identification codes for anonymous collection of longitudinal questionnaire data. *Public Opinion Quarterly*, 48, 370-378.

- Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: The Guilford Press.
- Lantolf, J., & Frawley, W. (1985). Oral proficiency testing: A critical analysis. *Modern Language Journal*, 69, 337-345.
- Lim, S. (1992). *Investigating learner participation in teacher-led classroom discussions in junior colleges in Singapore from a second language acquisition perspective*. Unpublished doctoral dissertation, National University of Singapore.
- MacIntyre, D., Baker, C., Clement, R., & Conrad, S. (2001). Willingness to communicate, social support, and language learning orientations of immersion students. *Studies in Second Language Acquisition*, 23, 369-388.
- Masole, T. M. (2011). *Enhancing the quality of performance assessment in agriculture in Botswana*. Unpublished doctoral dissertation, University of Pretoria, South Africa.
- McKay, P. (2006). *Assessing young language learners*. Cambridge: Cambridge University Press.
- McMillan, J. H. (2007). *Classroom assessment: Principles and practice for effective standardized-based instruction* (4th Ed.). Boston, MA: Allyn & Bacon.
- McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.
- Morrow, K. E. (1977). *Techniques of evaluation for a notional syllabus*. New York: Harcourt Brace and World.
- Nunan, D. (1988). *The learner-centered curriculum*. Cambridge: Cambridge University Press.

- O'Brien, J. (1997). Statewide social studies performance assessment: Threat or treat? *The Social Studies*, 88(2), 53-59.
- Oller, J. (1979). *Language tests at school: A pragmatic approach*. London: Longman.
- O'Malley, M., & Valdez Pierce, L. (1996). *Authentic assessment for English language learners*. New York: Addison-Wesley Publishing.
- Park, S. (2000). An analysis of results for practical performance assessment items in high school. *English Linguistic Science*, 5(1), 21-40.
- Pearson, E. (1988). Learner strategies and learner interviews. *ELT Journal*, 42(3), 173-178.
- Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice*, 12(4), 10-12.
- Robin, D. T. (2006). Changing practices: Influences on classroom assessment. *Assessment in Education*, 13(3), 239-264.
- Schafer, W. D. (1993). Assessment literacy for teachers. *Theory into Practice*, 32, 118-126.
- Shih, C. M., & Wang, L. Y. (2010). *Factors affecting English language teachers' classroom assessment practices: A case study at Singapore secondary schools*. Research Brief No. 13-010. Singapore: National Institute of Education.
- Shohamy, E. (1993). *The power of test: The impact of language testing on teaching and learning*. ERIC Reproduction Service No. ED362040.

- Spada, N. (1986). The interaction between type of contact and type of instruction: Some effects on the L2 proficiency of adult learners. *Studies in Second Language Acquisition*, 8(2), 181-199.
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72(7), 534-539.
- (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758-765.
- Swain, M. (1984). Large-scale communicative testing: A case study. In S. L. Savignon, & M. Berns (Eds.), *Initiatives in communicative language teaching* (pp. 185-201). Reading, MA: Addison Wesley.
- (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass, & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury House.
- Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38, 1-10.
- Vitali, G. J. (1993). *Factors influencing teachers' assessment and instructional practices in an assessment-driven educational reform*. Unpublished doctoral dissertation, University of Kentucky.
- Wall, D., & Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing*, 10(1), 41-69.
- Watanabe, Y. (1996). Does grammar translation come from the entrance examination?: Preliminary findings from classroom-based research. *Language Testing*, 13(3), 318-333.
- Weir, C. J. (1990). *Communicative language testing*. New York:

Prentice Hall.

- Woods, D. (1996). *Teacher cognition in language teaching: Beliefs, decision-making and classroom practice*. New York: Cambridge Applied Linguistics.
- Wudong, W. (1994). *English language development in China*. Unpublished doctoral dissertation, University of Tasmania.
- Yang, T. L. (2008). Factors affecting EFL teachers' use of multiple classroom assessment practices with young language learners. *English Teaching & Learning*, 32(4), 85-123.
- Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education*, 16(4), 323-342.
- Zhou, Y. P. (1991). The effect of explicit instruction on the acquisition of English grammatical structures by Chinese learners. In C. James, & P. Garrett (Eds.), *Language awareness in the classroom* (pp. 254-277). London: Longman.

[부록 1] 예비 연구용 설문지

중·고등학교 영어과 교사의 영어 말하기 수행평가에 대한 설문조사

안녕하십니까? 오늘도 어려운 여건 속에서 학생들을 지도하시느라 열정을 쏟고 계시는 선생님께 진심으로 존경을 표합니다.

이미 10년 전부터 학교 현장에 본격적으로 자리를 잡아 시행되던 '수행평가'는 그 중요성이 점점 높아지고 있고, 특히 EFL 환경에서 영어를 학습해야 하는 우리나라의 경우 의사소통의 표현기능인 '말하기'의 중요성은 더욱 더 중요하다고 볼 수 있습니다. 그러한 의미에서 본 설문지는 '영어 말하기 수행평가'에 대한 선생님의 생각을 알아보기 위해서 작성되었습니다. 평소 선생님께서 생각하시던 점을 솔직하게 응답해 주신다면, 본 연구에 매우 소중한 자료가 될 것 같습니다.

선생님께서 응답해 주신 내용은 비밀이 보장되며, 또한 순전히 연구의 목적으로만 사용됨을 약속드리오니 솔직한 느낌과 생각으로 응답해 주시면 대단히 감사하겠습니다.

늘 선생님의 건강과 행복을 기원합니다.

2013년 7월

부경대학교 일반대학원 영어교육전공 고 현 우

* 해당 번호에 'O' 표 해주시기 바랍니다.

1. 성별

- ① 남 () ② 여 ()

2. 근무 학교

- ① 중학교 () ② 고등학교 ()

3. 교직 경력

- ① 5년 이하 () ② 6~10년 ()
③ 10~20년 () ④ 20년 이상 ()

4. 말하기 수행평가 관련 연수 이수정도

- ① 전혀 없음 () ② 1회 () ③ 2회 () ④ 3회 이상 ()

5. 선생님의 영어 말하기 능력에 대한 자기평가

- ① 아주 우수함 () ② 우수함 () ③ 보통임 () ④ 부족함 ()

I. 다음은 선생님의 영어 말하기 수행평가의 시행도에 관한 문항들입니다. 각 진술을 잘 읽으시고, 해당하는 부분에 'O' 표 해주시기 바랍니다.

1. 선생님께서는 한 학기에 영어 말하기 수행평가를 몇 번 실시하십니까? ① 1번 () ② 2번 () ③ 3번 이상 () ④ 수시 ()
2. 선생님께서는 몇 가지 방법으로 영어 말하기 수행평가를 실시하고 계십니까? ① 1가지 () ② 2가지 () ③ 3가지 () ④ 4가지 이상 ()

II. 다음은 선생님의 영어 말하기 수행평가의 전문성에 관한 문항들입니다. 각 진술을 잘 읽으시고, 솔직한 생각이나 느낌에 'O' 표 해주시기 바랍니다.

평가 방법 선정 및 평가 도구 개발	전혀 그렇지 않다	그렇지 않다	보통이다	그렇다	매우 그렇다
1. 영어 말하기 수행평가의 목적과 내용에 적합한 수행평가 방법을 선택한다.	1	2	3	4	5
2. 사회문화적 배경, 흥미 및 적성 등 학생의 특성에 적합한 영어 말하기 수행평가 방법을 선택한다.	1	2	3	4	5
3. 학생의 성취 수준에 적합한 영어 말하기 수행평가 방법을 선택한다.	1	2	3	4	5
4. 영어 말하기 수행평가의 목적과 내용에 부합하는 평가 도구가 없을 경우 평가 도구를 직접 개발한다.	1	2	3	4	5
5. 영어 말하기 수행평가 과제 및 채점 기준의 개발 원리를 충실히 반영하여 평가 도구를 개발한다.	1	2	3	4	5

평가 실시, 채점 및 성적 부여	전혀 그렇지 않다	그렇지 않다	보통이다	그렇다	매우 그렇다
1. 학생들에게 영어 말하기 수행평가의 계획에 대해서 사전에 공지한다.	1	2	3	4	5
2. 평가 방법과 평가 도구에 대해서 정확하게 숙지한 후 영어 말하기 수행평가를 시행한다.	1	2	3	4	5
3. 영어 말하기 수행평가를 실시하기 전에 학생들에게 과제 및 채점 기준에 대해서 충분히 설명한다.	1	2	3	4	5
4. 영어 말하기 수행평가의 채점 기준을 정확하게 숙지한 후 채점한다.	1	2	3	4	5
5. 영어 말하기 수행평가의 각 영역(발음, 유창성, 과제 완성 등)에 대한 검사 점수의 의미를 이해하고 점수를 부여한다.	1	2	3	4	5

평가 결과 해석과 활용	전혀 그렇지 않다	그렇지 않다	보통 이다	그렇다	매우 그렇다
1. 영어 말하기 수행평가의 목적 및 내용에 기초하여 평가 결과를 해석한다.	1	2	3	4	5
2. 영어 말하기 수행평가 결과를 해석할 때 학생, 평가 도구, 평가 환경 및 평가자 등에 의한 변수를 고려한다.	1	2	3	4	5
3. 영어 말하기 수행평가의 결과에 기초하여 학생의 성취 수준, 강점 및 약점 등을 파악한다.	1	2	3	4	5
4. 영어 말하기 수행평가의 결과를 종합하여 학생뿐만 아니라 교수·학습에 대한 의사결정 자료로 활용한다.	1	2	3	4	5

III. 다음은 선생님의 영어 말하기 수행평가의 효용성에 관한 문항들입니다. 각 진술을 잘 읽으시고, 솔직한 생각이나 느낌에 'O' 표 해주시기 바랍니다.

영어 말하기 수행평가의 효용성	전혀 그렇지 않다	그렇지 않다	보통 이다	그렇다	매우 그렇다
1. 영어 말하기 수행평가를 시행함으로써 학생의 영어 말하기 능력 향상에 도움이 된다.	1	2	3	4	5
2. 영어 말하기 수행평가를 통해서 학생의 영어 말하기 능력의 발달 과정을 알아보는 데 도움이 된다.	1	2	3	4	5
3. 영어 말하기 수행평가는 학생의 영어 말하기 성취 수준을 파악하는 데 도움이 된다.	1	2	3	4	5
4. 영어 말하기 수행평가는 학생이 영어 말하기에 대한 학습 동기나 흥미를 가지는 데 도움이 된다.	1	2	3	4	5
5. 영어 말하기 수행평가는 교사의 영어 말하기 수업 방법의 개선에 도움이 된다.	1	2	3	4	5

IV. 다음은 선생님의 영어 말하기 수행평가 시행의 한계성에 관한 문항들입니다. 각 진술을 잘 읽으시고, 솔직한 생각이나 느낌에 'O' 표 해주시기 바랍니다.

영어 말하기 수행평가의 시행한계성	전혀 그렇지 않다	그렇지 않다	보통 이다	그렇다	매우 그렇다
1. 학급 규모가 크거나 학생 수가 많아 영어 말하기 수행평가를 시행하기가 어렵다.	1	2	3	4	5
2. 채점의 객관성 확보가 어려워 영어 말하기 수행평가를 시행하기가 어렵다.	1	2	3	4	5
3. 적절한 평가 도구가 없어 영어 말하기 수행평가를 시행하기가 어렵다.	1	2	3	4	5

4. 학생의 학습 부담을 가중시킴으로 영어 말하기 수행평가를 시행하기가 어렵다.	1	2	3	4	5
5. 학교 업무로 인해 시간이 부족하여 영어 말하기 수행평가를 시행하기가 어렵다.	1	2	3	4	5
6. 영어 말하기 수행평가에 도움이 되는 교사의 연수나 자료가 부족하여 영어 말하기 수행평가를 시행하기가 어렵다.	1	2	3	4	5
7. 사교육비의 증가에 대한 우려로 영어 말하기 수행평가를 시행하기가 어렵다.	1	2	3	4	5



[부록 2] 본 연구용 설문지

중·고등학교 영어과 교사의 영어 말하기 수행평가에 대한 설문지

안녕하십니까? 오늘도 어려운 여건 속에서 학생들을 지도하시느라 열정을 쏟고 계시는 선생님께 진심으로 존경을 표합니다.

이미 10년 전부터 학교 현장에 본격적으로 자리를 잡아 시행되던 '수행평가'는 그 중요성이 점점 높아지고 있고, 특히 EFL 환경에서 영어를 학습해야 하는 우리나라의 경우 의사소통의 표현기능인 '말하기'의 중요성은 더욱 더 중요하다고 볼 수 있습니다. 그러한 의미에서 본 설문지는 '영어 말하기 수행평가'에 대한 선생님의 생각을 알아보기 위해서 작성되었습니다. 평소 선생님께서 생각하시던 점을 솔직하게 응답해 주신다면, 본 연구에 매우 소중한 자료가 될 것 같습니다.

선생님께서 응답해 주신 내용은 비밀이 보장되며, 또한 순전히 연구의 목적으로만 사용됨을 약속드리오니 솔직한 느낌과 생각으로 응답해 주시면 대단히 감사하겠습니다.

늘 선생님의 건강과 행복을 기원합니다.

2014년 4월

부경대학교 일반대학원 영어교육전공 고 현 우

* 해당 번호에 'O' 표 해주시기 바랍니다.

1. 성별

- ① 남 () ② 여 ()

2. 근무 학교

- ① 중학교 () ② 고등학교 ()

3. 교직 경력

- ① 5년 이하 () ② 6~10년 () ③ 11~15년 ()
④ 16~20년 () ⑤ 21년 이상 ()

4. 말하기 수행평가 관련 연수 이수정도

- ① 전혀 없음 () ② 1회 () ③ 2회 이상 ()

5. 선생님의 영어 말하기 능력에 대한 자기평가

- ① 부족함 () ② 보통임 () ③ 우수함 ()

I. 다음은 선생님의 영어 말하기 수행평가의 시행도에 관한 문항들입니다. 각 진술을 잘 읽으시고, 해당하는 부분에 'O' 표 해주시기 바랍니다.

1. 선생님께서는 한 학기에 영어 말하기 수행평가를 몇 번 실시하십니까? ① 1번 () ② 2번 () ③ 3번 이상 ()
2. 선생님께서는 몇 가지 방법으로 영어 말하기 수행평가를 실시하고 계십니까? ① 1가지 () ② 2가지 () ③ 3가지 이상 ()

II. 다음은 선생님의 영어 말하기 수행평가의 전문성에 관한 문항들입니다. 각 진술을 잘 읽으시고, 솔직한 생각이나 느낌에 'O' 표 해주시기 바랍니다.

평가 방법과 도구의 선정 및 개발	전혀 그렇지 않다	그렇지 않다	보통이다	그렇다	매우 그렇다
1. 영어 말하기 수행평가의 목적과 내용에 적합한 수행평가 방법과 도구를 선택한다.	1	2	3	4	5
2. 학생의 성취 수준에 적합한 영어 말하기 수행평가 방법과 도구를 선택한다.	1	2	3	4	5
3. 목적과 내용에 부합하는 평가 방법과 도구가 없을 경우 영어 말하기 수행평가 방법과 도구를 직접 개발한다.	1	2	3	4	5
4. 과제 및 채점 기준의 개발 원리를 충실하게 반영하여 영어 말하기 수행평가 방법과 도구를 직접 개발한다.	1	2	3	4	5

평가의 실시 및 채점	전혀 그렇지 않다	그렇지 않다	보통이다	그렇다	매우 그렇다
1. 영어 말하기 수행평가를 실시하기 전에 학생들에게 평가 계획에 대해서 사전에 공지한다.	1	2	3	4	5
2. 영어 말하기 수행평가를 실시하기 전에 평가 방법과 도구에 대해서 정확하게 숙지한다.	1	2	3	4	5
3. 영어 말하기 수행평가를 실시하기 전에 학생들에게 과제 및 채점 기준에 대해서 충분히 설명한다.	1	2	3	4	5
4. 영어 말하기 수행평가의 채점 기준을 정확하게 숙지한 후 채점을 한다.	1	2	3	4	5

평가 결과의 해석 및 활용	전혀 그렇지 않다	그렇지 않다	보통 이다	그렇다	매우 그렇다
1. 영어 말하기 수행평가의 목적과 내용에 기초하여 평가 결과를 해석한다.	1	2	3	4	5
2. 영어 말하기 수행평가의 결과를 해석할 때 학생, 평가 도구, 평가 환경 및 평가자 등의 변수를 고려한다.	1	2	3	4	5
3. 영어 말하기 수행평가의 결과를 바탕으로 학생의 성취 수준, 강점 및 약점 등을 파악한다.	1	2	3	4	5
4. 영어 말하기 수행평가의 결과를 종합하여 차후 교수·학습에 대한 의사결정의 자료로 활용한다.	1	2	3	4	5

III. 다음은 선생님의 영어 말하기 수행평가의 효용성에 관한 문항들입니다. 각 진술을 잘 읽으시고, 솔직한 생각이나 느낌에 'O' 표 해주시기 바랍니다.

영어 말하기 수행평가의 효용성	전혀 그렇지 않다	그렇지 않다	보통 이다	그렇다	매우 그렇다
1. 영어 말하기 수행평가 시행은 학생의 영어 말하기 능력 향상에 도움이 된다.	1	2	3	4	5
2. 영어 말하기 수행평가를 통해서 학생의 영어 말하기 능력의 발달 과정을 알아보는 데 도움이 된다.	1	2	3	4	5
3. 영어 말하기 수행평가는 학생의 영어 말하기 성취 수준을 파악하는 데 도움이 된다.	1	2	3	4	5
4. 영어 말하기 수행평가는 학생이 영어 말하기에 대한 학습 동기나 흥미를 가지는 데 도움이 된다.	1	2	3	4	5
5. 영어 말하기 수행평가는 교사의 영어 말하기 수업 방법의 개선에 도움이 된다.	1	2	3	4	5

