

문장의 다양성이 영어논술시험 평가에 미치는 영향 연구

홍효기*, 이호†
중앙대학교

Hong, Hyo-Gi & Lee, Ho. (2011). Studying the effect of syntactic diversity on rating of an L2 argumentative essay test. *Modern English Education*, 12(3), 406-427.

This study explores the effect of syntactic diversity and variety on rating of an L2 argumentative essay test. For the current study, thirty university students participated in the four series of timed essay tests. The researchers analyzed the students' essays and computed the number of sentences, the ratio of simple sentence, the ratio of complex sentence, the ratio of compound sentence, and the ratio of T-unit. At the same time, two expert raters gave holistic scores per essay and made analytic ratings in terms of nine areas including syntactic variety. In addition, two raters had semi-structured interviews. As a result, students failed to show the consistent improvement through the essays in all fluency-related features. Further, the features did not have significant correlation with the holistic essay score. Finally, the syntactic diversity as one of assessment areas was judged as an overfit, which means the syntactic diversity might be a redundant feature and did not have variance enough for an independent feature. Two expert raters reported that they consider various features in the decision of scores of syntactic variety. This study suggested that we need to redefine the construct of writing fluency in the context of writing performance tests.

[writing fluency/syntactic diversity/syntactic complexity/syntactic maturity/writing assessment/argumentative essay test/작문 유창성/문장 다양성/문장 복잡성/문장 성숙도/작문평가/논술평가]

* 제1저자

† 교신저자

I. 서론

언어학습이나 수행평가에 있어서 유창성은 정확성과 더불어 중요한 평가항목으로 인정받아 왔다. Brumfit(1984)은 언어 학습과 언어평가에 있어서 유창성과 정확성은 기본적인 토대이며 적대관계가 아닌 상보적 관계라고 제시하였으며, Brown(2003)은 정확성이 확장된 형태가 유창성이라고 제안하였다. 이러한 유창성의 중요성을 인식한 다수의 학자들은 유창성에 대하여 나름대로의 의미있는 정의를 내렸다. 예를 들어, Dechert와 Raupach(1980)은 유창성을 많은 휴지(pause)없이 원활하게 의식적 노력이 없이 잘 흘러가는 말의 상태라고 정의를 내렸으며, Lennon(1990)은 정확성과는 다소 상대적인 개념으로서 유창성을 발화의 속도와 언어의 길이, 그리고 내용의 원활함이라고 정의하였다.

그러나, 위와 같은 유창성의 정의는 말하기에 초점을 맞추어 이루어진 것으로서 상대적으로 쓰기 분야에서는 유창성에 대한 관심이 소홀하였다. Brunfit(1984)에 의하면, 실시간으로 진행되는 말하기 분야에서는 휴지의 길이나 발화의 속도 같은 유창성 관련 자질이 명확하게 표출되는 반면, 쓰기 분야에서는 빠른 속도의 의미 전달보다는 정확하게 저자의 의미를 전달하는 것이 더 중요하다는 인식으로 인하여, 유창성의 개념이 쓰기에서는 명료하게 정의내리지 못했다고 하였다.

최근 들어, 쓰기에 있어서도 유창성의 개념이 중요하다는 주장이 대두되면서(Hamp-Lyons, 1991; Wolfe-Quintero, Inagaki & Kim, 1998), 작문 유창성에 대한 정의를 내리거나 수험자들의 유창성의 향상을 측정하는 여러 세부지표들에 대한 연구와 개발에 관심을 기울이기 시작하였다. 어떤 연구자들은 상위 학습자에 있어서의 유창성 발달이 어떤 방향으로 일어나는가 관심을 기울이기도 하였고(Towell, Hawkins, & Bazergui, 1996), 어떤 연구는 하위 학습자의 유창성의 향상을 잘 판별할 수 있는 세부지표를 인식하기도 하였다(Ishikawa, 1995).

하지만, 이러한 유창성을 영어 작문 수행 평가 상황과 결부시켜 그 신뢰도 및 타당도를 조명하는 논문은 매우 드문 실정이다. 대부분의 작문 유창성과 관련된 논문들은 학습자에 초점을 맞추어, 유창성 향상을 의미있게 측정할 수 있는 지표 개발 및 연구에 관심을 가진 반면, 평가 상황에서 평가항목으로서의 유창성이 채점자에게 어떤 의미가 있고 유창성이 다른 평가 영역과 어떻게 상관관계를 이루는지에 대한 연구는 없다고 할 수 있다.

이러한 사실은 기존의 쓰기평가의 척도들을 고려해 보면 더 명확해진다. 일례로, Spalding과 Cummins(1998)에서 제시되어 있는 켄터키 주 지역의 논술 척도에서는 유창성이라는 독립적인 평가항목이 개설되지 않고 다만 문장의 다양성 측면으로 간접적으로 측정되고 있었다. 이렇게

유창성이라는 명확한 측정기준이 삽입되는 대신 문장의 다양성과 같은 관련 지표로 제시되어 있는 것은 다른 논술평가 척도표에서도 흔히 볼 수 있다(Weigle, 2002; Willard-Traub, Decker, Reed & Johnston, 1999).

위와 같은 영어수행 평가 상황에서의 유창성 연구의 부재를 인식하고, 본 논문은 유창성의 관련 지표로서의, 독립된 평가 항목으로서의 문장의 다양성을 조사해 보고자 한다. 문장의 다양성을 작문의 유창성 대신으로 활용하는 이유는 다양한 문장 구조의 사용이 유창성과 밀접하게 관련이 있다는 기존 문헌(Hunt, 1965; Sasaki, 2000)의 언급이 있었고, 기존 쓰기 평가 척도표에서 가장 빈번하게 쓰이는 척도 중의 하나가 문장의 다양성(syntactic variety)이며, 문장의 다양성이 개별 단어 단위로 유창성을 측정하는 다른 지표에 비하여 훨씬 평가하기가 용이한 실용성을 지니고 있기 때문이다. 결론적으로, 이상권, 윤인섭(2008)이 “훌륭한 텍스트는 다양성의 미학”(p. 109)이라는 주장에서 보듯, 다양한 어휘의 사용과 다양한 문장 구조의 사용이 독자의 관심을 더 끌어들일 수 있으며 이러한 요소들이 영어 텍스트의 논술 구조에서 중요하게 인식되어야 하기 때문이다.

II. 연구 배경

제2외국어로서의 작문교육 분야에서 문장의 다양성(syntactic diversity)이 영어작문 평가에 미치는 영향을 다룬 논문은 찾아보기가 어려웠다. 반면에 문장의 성숙도(syntactic maturity) 혹은 능숙도, 작문의 유창성 등 문장의 다양성을 대체하는 여러가지 용어들을 사용한 논문들이 기존 문헌에서 자주 발견되었다. 본 논문에서는 가급적 유창성이라는 개념으로 통일하여 기존문헌을 정리하고자 하는데 그 이유는 다음의 세가지가 있다. 첫째, 문장의 성숙도가 결과중심적 측정 방식에서 (특히 T-unit분석에서) 국한되어 쓰이는 경향이 있고, 둘째, 작문의 유창성이 광의의 개념으로 문장의 다양성을 포함하고 있으며(Larsen-Freeman, 2006), 마지막으로 실제 시행되는 영어작문 평가의 루브릭에서도 문장의 다양성이 유창성 항목들과 같이 중첩되는 경우가 많기 때문이다(East, 2009; Spalding & Cummins, 1998; Willard-Traub et al., 1999). 따라서, 선행연구로서 작문의 유창성과 유창성의 측정에 대해서 고찰해 보고자 한다.

유창성 측정과 관련된 연구들의 체계화된 분석틀로서 Yeon Hee Choi(2010)와 Latif(2009)는 결과중심적 유창성 측정방식(product-based fluency measures)과 과정중심적 유창성 측정방식(process-oriented fluency measures)을 제시하였는데, 우선 결과중심적 유창성 측정 방식은

학생들의 쓰기 최종본을 유창성 관련 지표로서 분석하는 것이다(Latif, 2009). 결과 중심적 유창성 방식을 고수하는 논문들의 주요 관점은 학습자들이 작문에 대한 발전 과정을 종적으로 탐색하는 것이 아니라, 학생들의 결과물을 횡적으로 수집하여 여러 분석지표를 가지고 기술통계를 제시하거나 t-test나 AVNOA 방식을 수행하는 것이다. 이런 분석 방식은 정확성을 측정하는 작문에서의 오류 분석(이현구, 2001)과 그 틀을 같이 하는 것으로, 대표적으로 Hunt(1965)에 의하여 고안된 T-unit 지표가 많이 사용되었는데 여기서 T-unit이란 문법적으로 허용되는 최소한의 수준을 갖춘 문장의 단위, 즉 하나의 주어와 동사를 가진 구조라고 정의되고 있다. Ishikawa(1995)는 작문당 문장수를 포함하여 통사적 능숙도를 분석할 수 있는 세부 지표 24개를 나열하였는데, 그 중 T-unit과 관계된 것은 50%가까이 되는 11개의 지표였다.

다수의 연구에서는 T-unit이 작문의 유창성을 변별할 수 있는 유의미한 지표임을 제시하고 있다. Larsen-Freeman(1978)은 전체 문장에서의 무오류 'T-unit의 비율', '무오류 T-unit당 평균단어수의 수'가 ESL학습자의 다양한 쓰기능숙도를 변별하는데 의미있는 지표라고 제시하였다. 유재임(2005) 역시 학년의 변화에 따라 문장의 유창성이 올라간다는 결과를 제시하면서, 그 유창성의 향상 정도를 가장 분명하게 보여주는 지표가 'T-unit의 길이'라고 주장하였다.

반면, Ishikawa(1995)는 영어작문능력이 낮은 학생들에게 있어서 T-unit분석은 통계적으로 큰 의미가 없는 지표라고 보고하였다. 그는 영어작문능력이 매우 낮은 일본인 학생들을 대상으로, 주어진 그림의 내용에 대하여 직접 글로 기술하는 그룹과 그림 내용에 대한 질문들에 대답하는 그룹으로 나뉘서 두 그룹의 향상 정도를 명확하게 판별할 수 있는 측정방식을 탐색하였다. 그 결과 '무오류 절의 숫자'가 영작실력의 유의미한 향상을 가르키는 유일한 지표로 판명되었고, 반면 T-unit과 관련된 지표는 하위 단계의 학생들의 능력의 향상을 제시하는데에는 도움이 되지 않았음을 제시하였다. 박지순, 서세정(2009) 역시 T-unit관련 지표보다 '문장당 평균 단어수', '문장 당 평균 절의 수', '전체 절에서 무오류 절이 차지하는 비율' 등 세가지 지표가 문장의 유창성과 복잡성을 측정하는 적합한 지표라고 주장하였다.

과정 중심적 유창성 측정 방식은 상대적으로 최근에 각광을 받기 시작하였는데, 이 방식은 학습자가 글을 쓰고 있는 과정에 초점을 맞추었다. 과정 중심적 작문의 유창성을 조사하기 위하여 학자들은 주로 질적 연구방법론을 사용하는데 그 중 자판기 입력기법 프로그램을 이용하여 학습자로부터의 데이터를 수집한 이후(Miller, 2000; Yeon Hee Choi, 2010) 사고표출기법(think-aloud protocol analysis)으로 분석하는 연구들이 주목할 만 하다(Chenoweth & Hayes, 2001; Latif, 2009).

Latif(2009)는 글의 분량과 작문비율(분당 단어수)이 작문의 유창성을 타당하게 측정하지 못한다고 주장하면서, 기록분절단위(translating episode)를 하나의 작문유창성의 측정요소로서 제시하였다. 그는 사고표출기법을 통해 영어 작문 학습자가 계획이나 교정등의 작문활동으로 인하여 자주 글쓰기를 중단하는 모습을 관찰하면서, 한 두 단어나 구의 수준으로 중단된 그 글의 단위를 기록분절단위로 정의하고 '기록분절단위의 평균길이'가 보다 더 타당성있는 측정요소임을 제안하였다.

Yeon Hee Choi(2010)는 영어작문 능력에 따라 상하 두 집단으로 나눈 12명의 대학생들을 대상으로, 다양한 결과중심적 측정방식과 과정중심적 측정방식을 이용하여 작문의 유창성을 측정하고 상하 그룹을 변별할 수 있는 측정기준들을 선별하였다. 변별력이 있는 측정기준으로서, 결과중심적 측정방식에서는 '최종본에서의 전체 단어의 수', '오류없는 T-unit의 수', '오류없는 T-unit에서의 단어의 수'를 제시되었다. 반면, 과정중심적 측정방식에서는 '단어의 수', Latif에서 언급된 기록분절단위의 개념과 비슷한 '기록지속단위(burst)의 평균단어수', '추후 교정될 단어를 포함하고 있는 기록지속단위(R-burst)의 평균단어수', '단어간·단어내 휴지의 수', '단어내 휴지의 평균 길이' 등이 변별력있는 기준으로 인식되었다.

Chenoweth와 Hayes(2001) 역시 사고표출기법을 이용하여 제2언어 학습자들 대상으로 해당 언어에 대한 노출경험과 작문 유창성에 대한 상관관계를 연구하였다. 그들은 작문 유창성을 측정하는 지표로서 '분당 쓰여진 단어의 수', 휴지이후와 교정행동전에 기록된 평균 단어의 길이를 측정한 '기록지속단위의 평균단어수'를 포함시켰다. 연구의 결과, 해당언어에 대한 지속적 경험을 축적할수록 위에 언급된 작문 유창성의 지표가 증가하고, 교정횟수는 감소함을 발견하였다.

Miller(2000)는 제2언어학습자가 컴퓨터를 활용한 쓰기 과정을 사례연구 형식으로 관찰하였는데, 학습자의 자판기 입력 기록을 통하여 작문 텍스트의 위치에 따른 휴지의 길이에 의하여 작문의 유창성을 밝히고자 하였다. Miller의 결과에 의하면, L2학습자가 모국어 그룹에 비하여 특별히 고급어휘를 사용하는 환경에서 휴지의 길이가 길었다고 보고하였다.

반면, 이지연(2007b)은 교사의 인상평가 방식에 의거하여 작문 유창성이란 구인(construct)을 정리하고자 하였다. 즉, 외국어고등학교에 재학중인 195명의 학생들이 쓴 390개의 에세이를 분석하여 질적인 관점에서의 유창성과 관련있는 평가요소들을 규명하고자 하였다. 그녀는 유창성의 평가항목으로서 '적절한 분량의 글의 길이', '일관성(coherence)', '응집성(cohesion)', 그리고 '고급언어수준'을 포함시킬 수 있음을

제안하였다. 특히, 그녀의 연구에서 문장의 다양성 및 복잡성과 관련된 지표를 고급언어 수준 영역에서의 문장 구조 선택의 측면에서 간접적으로 포함하고 있었다.

지금까지, 문장의 유창성과 관련된 기존의 연구들을 제시하였다. 기존 논문들을 검토하면서 한가지 주목할 점은, 채점자가 참여하는 실제 영어작문 수행평가 상황에서 유창성이라는 항목이 평가항목으로서 적합한지를 다룬 논문이 없었다는 것이다. 다수의 논문들은 유창성과 관련있는 지표들을 가지고 학습자의 작문 데이터를 세밀하게 분석하는 방식(micro analysis)을 채택하였으나 좀 더 거시적인 안목에서 채점자 입장에서 유창성의 신뢰도를 다룬 논문들이 거의 존재하지 않았다는 점은 아쉬운 점이라고 할 수 있다.

III. 연구 방법

1. 연구 문제

연구배경에서 제기된 바와 같이 영어논술평가에서의 통사론적 다양성이 수험자 및 채점자 측면에서 어떻게 영향을 미치는지를 살펴볼 필요가 있다. 이를 위해 본 연구는 다음과 같은 세 가지 연구 질문을 중심으로 본 논문을 기술하였다.

1. 학생들의 4차에 걸친 논술 시험을 통해서 문장 다양성의 세부 지표들이 어떠한 변화 양상을 보이는가?
2. 문장다양성의 세부지표들과 영어논술 점수와의 관계는 어떠한가?
3. 채점자 측면에서 문장 다양성이라는 평가 항목이 적합한 항목인가?

위의 연구 질문들에 대한 결과를 얻기 위하여 연구자는 연구질문 1,2에 대하여 반복측정 ANOVA, Pearson 및 Spearman 상관관계 분석등의 양적연구방법을 채택하여 ‘문장의 수’, ‘단문의 비율’, ‘복문의 비율’, ‘중문의 비율’, ‘T-unit 비율’ 등을 조사하였다. 연구질문 3에 대해서는 FACETS 분석과 인터뷰 분석을 결합한 혼합연구방법을 수행하였다. 즉 연구 질문 1, 2은 학습자의 수행 측면에서 문장 다양성의 관련 지표들이 어떻게 변하는가를 살펴보기 위해 양적통계 분석을 수행하였고 연구 질문 3은 채점자 측면에서 문장 다양성의 평가항목에 대한 인식을 조사하기 위해 질적 연구 방법인 인터뷰 분석을 시행하였다.

2. 연구 참여자

본 연구를 위해 서울의 C대학교 영어과 관련 학부생 30명을 대상으로 실험을 진행하였다. 참여학생들은 “영어쓰기지도” 과목의 수강생들이며, 대부분이 영어교육 관련 활동에 관심이 많은 학생들이다. 30명의 학생들 중에서 남학생이 5명, 여학생이 25명으로서 여학생이 대다수를 차지하고 있다. 학생들의 영어논술 수준은, 학기초에 실시된 학생들의 1차 영어작문 점수가 3.43점 정도(5점 만점)인 것으로 추정된 바, 대체적으로 중급 혹은 중상급이상이라고 할 수 있다. 영어쓰기지도 과목의 교과과정은 영어 글쓰기 기술에 대한 교육이 학기초 두시간정도로만 이루어졌고, 나머지 대부분의 과정은 제2외국어로서의 쓰기지도 이론과 학생들의 영어논술시험 참여로 구성되었다. 따라서 학생들에게 문장의 다양성에 대한 명확하며 구체적인 사전지도는 이루어지지 않았다. 다만, 논술 시험 전에 채점기준표가 공개되어 학생들이 어떤 항목에 대하여 평가받는지 알 수 있도록 하였다. 학생들은 모두 50분의 시간안에 8차례의 논술 시험에 참여하였으며, 그 중 홀수 차수의 논술 시험(1차, 3차, 5차, 7차) 성적을 분석대상으로 하였다. 3차 시험의 경우, 2명의 결원인원이 생긴 이유로 28명이 참여하였으며 이로 인하여 연구 질문 3의 FACETS 분석대상에서는 3차 시험의 자료는 제외하였다.

채점자로서, 영어교육 분야에서 석사학위 이상을 받은 두 명의 전문가가 참여하였다. 두 명의 채점자는 EFL 학생들의 학술적 목적의 영어 논술 및 교양 영어 말하기 평가 상황에서 분석적 평가와 총괄적 평가를 다수 수행하여 평가자로서의 어느 정도 전문성을 지녔다고 할 수 있다. 본 논문은 이지연(2009)의 제언을 참고하여 채점의 신뢰성을 보완하기 위한 조치를 취하였다. 두 명의 채점자는 본 연구를 시행하는 연구자의 주도 아래 평가기준표의 평가항목들에 대한 상세한 논의 과정을 거쳤고, 예비채점을 통하여 채점자간 평가 차이를 조정하였다. 그 결과로서 표 1에서 제시된 바와 같이 만족할 만한 채점자간 신뢰도 수준을 도출하였다. Fleiss(1986)에 따르면, 급내상관계수에 의하여 평가자간 신뢰도를 구할 경우 평균측도값으로 그 수치가 .75이상이면 매우 좋은 수준이며 .70이상은 신뢰할 수 있는 수준이라고 언급하였다.

표 1

급내상관계수(ICC)에 의한 채점자간 신뢰도

측정단위	1차논술	3차논술	5차논술	7차논술	전체논술
단일측도	.56**	.66**	.63**	.53**	.56**
평균측도	.72**	.80**	.77**	.69**	.72**

**p<.01, Crobach alpha수치는 평균측도 값과 동일함

3. 연구 절차

참여자들은 모두 8번의 영어 논술시험을 수행하였다. 시험의 형식은 하나의 문항(혹은 주제)이 제시되면 그 주제에 대하여 주어진 시험지에 50분의 시간내에 작성하여 제출하는 것이었다. 본 논문의 분석 대상이 되는 홀수 차수의 주제들은 모두 영어교육과 관련된 사안에 대한 본인의 의견과 지식을 서술하는 것으로 지나치게 심도깊은 전공지식을 묻지 않으면서 정답의 범위가 개방적인 문항들이 출제되었다. 참여자들은 시험지에 본인의 이름 대신 학번을 적어서 제출하도록 하여, 채점자가 수험자를 알 수 없도록 조치하였다.

참여자들이 제출한 에세이 자료들에 대하여 두 명의 채점자들은 미리 작성된 평가기준표에 의거하여 총괄 영역 및 세부 영역들을 채점하였다. 본 연구에서 사용된 평가기준표는 ‘영어쓰기지도’ 과목의 성격에 부합하는 것으로 판단된 Willard-Traoub 등(1999)의 진단적/포트폴리오 목적의 평가기준표를 참고하여 작성된 것으로, 지난 3년간 평가의 변별력 향상을 고려하고 학생들의 의견을 적극적으로 반영하여 수정 보완된 것이다. 여기서 총괄 점수는 학습자의 전반적인 영어 논술실력을 단일 점수로 표현한 것으로 9점 척도(1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5)체제로 평가하였다. 세부 영역점수들은 모두 9가지 평가요소에 걸쳐서 각기 5점 척도(1, 2, 3, 4, 5) 체제로 산출하였다. 9가지 평가요소로는 ‘서론·본론·결론의 완비 여부’, ‘문장의 연결성’, ‘문단간 연결성’, ‘본인 주장의 제시 여부’, ‘예증의 설득력 및 구체성’, ‘내용상 무의미한 반복의 사용 여부’, ‘정확한 문법 사용’, ‘적절한 단어 사용’, ‘문장의 다양성’으로서, 전자의 세 가지 영역은 구조와 관련된 지표이며, 후자의 세 가지 영역은 언어적 측면과 관련된 지표이다 (부록 참조).

연구자는 연구질문 1과 연구질문 2의 분석을 위하여 두 채점자들의 총괄점수를 바탕으로 각 차수별 최종 총괄점수를 결정하였다. 두 채점자간 점수의 차이가 없을 경우에는 그 점수를 그대로 확정하였고 두 채점자간 차이가 0.5점이나 1점 이하일 경우에는 두 채점자간 평균을 적용하였다. 두 채점자간 차이가 1.5이상일 경우에는 연구자가 제3채점자로 참여하여 최종점수를 결정하였다.

모든 채점과정이 완료된 이후, 연구자는 한 명의 전문 채점자와 함께 채점 과정에 대한 반구조적인 인터뷰(Semi-structured interview)를 가졌다. 인터뷰의 내용은 주로 문장 다양성의 측정 가능성과 신뢰도에 대한 질의 응답으로 이루어졌으며, 인터뷰분석이 본 연구의 주요 연구자료가 아닌 관계로 여러 번에 걸쳐서 수행되기 보다는 단 한번 이루어졌다.

4. 자료분석

1) 연구질문 1의 자료 분석법

연구자는 학생들이 4차에 걸쳐 직접 작성했던 118개의 에세이(3차에 2명의 결원 발생)들의 문장구조 다양성을 분석하기 위하여, 다섯 가지의 세부 지표를 설정하였다. 세부지표로는 문장수, 단문비율, 복문비율, 중문비율, 그리고 T-unit 비율을 설정하였다.

연구자는 우선적으로 각 에세이마다 문장의 전체 숫자를 측정하였다. 이는 문장수가 에세이의 길이와 관련있는 지표임과 동시에, 비율로서 표시되는 각 지표들의 분모로서 역할을 수행하고 있기 때문이다. 단문비율 측정시 종속절을 포함하지 않는 하나의 독립절을 단문(simple sentence)으로 보고 전체 문장 수에서 그 단문들이 얼마만큼 많은지를 계산하였다. 복문비율의 경우, 해당 문장안에 하나의 종속절을 가지고 있으면 복문(complex sentence)으로 취급하고 단문비율의 측정과 마찬가지로 계산하였다. 또한 연구자는 적어도 두 개의 독립절이 등위접속사, 세미콜론등에 의한 접속부사 등으로 연결되었을 경우 중문(compound sentence)으로 보고 전체 문장수에서 중문이 차지하는 비율을 계산하였다. T-unit비율 계산시, 종속절이 포함된 하나의 독립절을 T-unit으로 정의하고(유재임, 2005) 문장수에서 T-unit이 차지하는 비율을 측정하였다.

연구자는 위와 같은 방식으로 계산된 문장수와 네 지표의 비율을 종속변수로 하고, 문장구조의 다양성과 관계있는 세부 지표들이 각 시험차수별로 어떻게 변화되는지 조사하기 위하여 반복측정 ANOVA를 사용하였다. 참여자의 전체 응답의 구형성을 검정하기 위하여 Mauchly방식을 사용한 결과, 대부분 구형성 가정을 만족하여 다변량 검정 결과를 채택하였다. 또한 효과크기를 검정하는 부분에타제곱을 삽입하여 실제적 유의성을 살펴보았다.

2) 연구질문 2의 자료 분석법

연구자는 두 채점자간의 점수를 비교하여 작성된 총괄최종점수와 문장 다양성의 세부지표들과의 관계성을 분석하기 위하여 Pearson의 상관관계 분석과 Spearman의 등위상관계수를 동시에 제시하였다. Pearson의 상관관계 분석을 위해서는 정규분포성이 충족되어야 하는데, Shapiro-Wilkson의 검정 결과 어떤 시험 차수에서 특정 세부 지표의 분포가 정규분포를 이루고 있지 않다는 점이 발견되었다. 따라서 연구자는

비교대상이 되는 총괄최종점수와 세부지표비율을 모두 서열척도로 변환하여 등위상관계수를 계산하였다.

3) 연구질문 3의 자료 분석법

연구질문 1과 2와는 달리, 연구질문 3에서는 1차, 5차, 7차의 세차례의 논술시험에 대한 채점자료를 가지고 분석하였다. 그 이유는 3차 논술시험에 두 명의 결원이 발생하였고 학생들의 성적에 직접 반영되는 차수가 바로 1차, 5차, 7차 논술시험이기 때문에 학생들로부터 좀 더 진정성 있는 자료를 얻을 수 있기 때문이다. 30명의 수험자들이 세차례의 논술시험에서 두명의 평가자로부터 아홉개의 평가영역의 점수를 부여받았으므로 총 1620가지의 평가데이터를 FACETS 분석 프로그램에 입력하였다. 특히 본 논문에서는 평가영역으로서의 문장의 다양성에 초점을 맞추어 문장의 다양성이 전체 평가의 신뢰도에 위협이 될 수 있는 부적합한 평가기준인지 혹은 과적합한 평가기준인지를 보고자 한다.

본 연구질문 3의 탐색을 위하여 사용된 FACETS 프로그램은 각 평가변인의 항목에 대하여 자세한 정보를 제공하기 때문에 영어쓰기평가나 말하기 평가와 같은 수행평가 분석에서 자주 쓰이는 통계도구이다. FACETS는 로짓(logit)을 이용하여 채점자의 엄격성, 평가영역의 난이도, 학생들의 상대적인 능력등을 표시하는데 일반적으로 양수의 logit은 채점자가 엄격하거나 평가영역의 난이도가 높거나 학생들의 능력이 높음을 표시한다. 또한 FACETS는 특정 변인내의 특정 항목의 행태에 대한 정보를 적합성 검증 통계(fit statistics)를 통해 제시함으로써 전체 평가의 신뢰도에 위협이 될 수 있는 개별 평가 항목들을 진단하는데 있어서도 매우 중요하게 사용되고 있다.

마지막으로, 연구자는 위의 FACETS 연구외에 문장의 다양성 측정에 대한 심도깊은 이해를 위하여 본 연구에 참여한 전문채점자 두 명과 간단한 인터뷰를 시행하였고, 그 인터뷰의 내용을 모두 녹음하여 본 논문 분석에 필요한 부분을 모아 따로 전사하여 분석하였다.

IV. 결과

1. 문장 다양성의 세부지표들의 변화 정도

표 2는 4차례의 논술시험을 통해서 문장 다양성을 측정하는 분석지표들이 어떻게 변화했는지를 기술통계로 제시하고 있다.

총괄점수 측면에서는, 학생들의 영어 논술 성적은 계속 향상되고 있는 반면, 다른 지표에서는 뚜렷한 경향성을 보이고 있지 않다. 예를 들어, 단문비율의 경우 1차 논술시험에 비하여 3차 논술시험에서 그 수치가 높아진 반면 5차와 7차 논술시험에서는 단문비율이 다시 줄어드는 경향을 보이고 있지만, 단문비율의 변화가 통계적으로 유의미하지는 않다.

표 2

각 논술시험과 문장 다양성 세부지표별 기술통계

분석단위	1차논술 평균(SD) N=30	3차논술 평균(SD) N=28	5차논술 평균(SD) N=30	7차논술 평균(SD) N=30
총괄점수	3.49(.55)	3.64(.56)	3.70(.52)	3.88(.39)
문장수	19.33(5.38)	22.70(6.37)	20.50(6.77)	20.50(4.35)
T-Unit비율	118.10(12.82)	119.55(10.16)	113.26(9.76)	116.29(11.11)
단문비율	46.93(14.66)	47.64(11.21)	45.37(14.02)	43.33(12.30)
복문비율	35.83(16.03)	32.50(9.26)	41.20(13.75)	40.70(13.77)
중문비율	10.04(7.75)	10.04(6.57)	5.64(6.23)	7.75(7.26)

혼합문비율의 평균치로 각각 7(1차), 10(3차), 8(5차), 8(7차)을 기록

표 3과 표 4는 반복측정 ANOVA를 실행한 이후의 결과를 정리하여 제시한 것이다. 표 3의 개체내 효과 검증에 의하면, 총괄점수와 문장수에서 F값이 각각 10.58과 3.99로 유의확률 0.01이하의, 복문비율과 중문비율에서 F값이 각각 3.29와 3.38로 유의확률 0.05이하의 유의미한 차이가 발견되었다. 그 유의미한 차이가 어디서 발생되는지를 자세하게 살펴보기 위하여 대응별 비교 검정을 수행하였는데, 총괄점수의 경우는 3차와 5차간의 시험성적 차이를 제외하고는 모든 차수간 시험성적에서 유의미한 향상을 보였다. 복문비율은 3차-5차, 3-7차 시험사이에 비율이 급격하게 높아졌음을 보이고 있고 중문비율은 각기 1차-5차, 3차-5차 사이에 그 비율이 크게 줄고 있음이 발견되었다. 문장수의 경우 1차와 3차시험사이에 큰 폭으로 증가하였다가 3차-5차, 3차-7차간에는 통계적으로 유의미하게 문장수가 감소하였다.

정리하자면, 학생들이 다수의 영어 논술시험을 치를 수록, 총괄점수는 계속하여 유의미한 향상을 지속하는 경향을 보이고 있는 반면에 단문비율은 통계적으로 유의미하지 않지만 다소 하락하는 모습을 보이고 있었다. 복문비율과 문장수, T-unit 비율의 뚜렷한 경향성을 도출하기가 어려웠다.

표 3
각 문장 다양성 세부지표별 개체-내 효과 검증

분석단위	제3유형 제곱합	자유도	평균제곱	F	유의확률	부분에타 제곱
총괄점수	2.600	3	.867	10.581	.000	.282
문장수	178.025	3	59.342	3.992	.010	.121
T-Unit비율	488.438	3	162.813	1.879	.140	.065
단문비율	356.714	3	118.905	.860	.465	.031
복문비율	1561.241	3	520.414	3.293	.025	.109
중문비율	374.384	3	124.795	3.382	.022	.111

표 4
각 문장구조 세부지표에 대한 시험 차수 대응별 비교

차수(I)	차수(J)	총괄점수		문장수		복문비율		중문비율	
		평균차 (I-J)	유의 확률	평균차 (I-J)	유의 확률	평균차 (I-J)	유의 확률	평균차 (I-J)	유의 확률
1차	3차	-1.15	.021	-3.37	.003	3.70	.194	.00	1.00
	5차	-2.21	.020	-1.17	.367	-5.86	.161	4.39	.012
	7차	-3.38	.000	-1.17	.202	-4.29	.253	2.29	.171
3차	5차	-0.05	.463	2.20	.025	-9.54	.003	4.39	.003
	7차	-2.23	.006	2.20	.013	-7.96	.007	2.29	.145
5차	7차	-1.18	.030	.00	1.00	1.57	.679	-2.11	.241

2. 문장 다양성의 세부지표들과 총괄점수와의 관계

표 5는 문장 분석의 세부지표들과 각 시험 총괄점수와의 상관관계를 제시하였는데, 전반적으로 모든 지표에 걸쳐서 상관계수가 낮아서 유의미한 상관관계를 발견할 수 없었다. 즉, 문장수, T-unit 비율, 복문의 비율의 증가가 좋은 총괄점수의 획득과 관련이 있다고 보기 어렵다는 것이다. 마찬가지로 단문비율 및 중문비율의 감소가 높은 총괄점수와 연관이 없는 것으로 보여진다. 결과적으로, 세부평가지표 가운데 어느 것도 전체 작문점수와 관련성이 없다고 할 수 있다.

표 5

차수별 총괄점수와 문장구조 세부지표들간의 상관관계

분석단위	측정단위	1차총괄	3차총괄	5차총괄	7차총괄	전체총괄
문장수	Pearson	.346	.261	-.045	-.018	.118
	rho	.365*	.278	.278	.070	.245
T-Unit비율	Pearson	.120	.185	.041	-.040	.248
	rho	.132	.219	.196	.030	.277
단문비율	Pearson	-.032	-.236	-.043	-.134	-.281
	rho	-.044	-.179	.072	-.161	-.232
복문비율	Pearson	-.039	.083	.006	.035	.057
	rho	-.033	.158	-.075	.075	.060
중문비율	Pearson	-.079	.162	-.122	-.193	-.034
	rho	-.142	.109	.051	-.026	.061

Pearson: Pearson correlation, rho: Spearman rank-order correlation

3. 평가영역으로서의 문장 다양성 측정의 적합성

그림 1은 전체 변인에 대한 정보를 비교할 수 있도록 정리한 척도표이다. 여러가지 평가변인들 중에서 특히 평가영역 변이(그림에서는 Item으로 표시)에 주목하여 보자면, 예증의 설득력 및 구체성(그림에서는 예증)이 다른 평가 항목에 비하여 가장 점수가 낮았던 분야이며, 서론·본론·결론의 완비 여부(그림에서는 서본)가 제일 쉬운 난이도의 평가영역이다. 특별히, 문장의 다양성¹은 다소 어려운 평가항목으로서 제시되었는데, 언어적 측면을 측정하는 평가 영역들이 대체적으로 난이도가 높았음이 발견되었다. 채점자변인의 경우 채점자2가 채점자1에 비하여 더 엄격한 채점경향을 보이고 있으며, 주제변인의 경우 논문 주제1(조기영어교육 문항)에서 학생들이 제일 낮은 점수를 기록하였다.

표 6은 평가영역에 대한 자세한 정보를 제공하고 있다. 각 평가영역의 적합성을 측정하기 위해서는 특별히 내외적합 표준화 값(infit z과 outfit z)에 주목해야 하는데, 그 수치가 -2.0에서 2.0사이에 존재해야 적합한 평가 영역으로 인정받을 수가 있다. 문장의 다양성 영역은 -2.5로서 과적합(overfit)으로 판명되는데, 이것은 문장의 다양성 영역이 타 영역의 점수분포와 지나치게 비슷하여 독립된 평가 영역으로서 인정받기 어렵다는 것이다. 즉, 문장의 다양성 영역이 다른 평가기준들의 점수 분포와는 차별화될 수 있는 정도의 편차가 필요한데, 그 편차를 갖추지 못했다는 것으로 밝혀졌다. 반대로 서론·본론·결론의 완비 여부는 그

¹ 문장의 다양성은 그림1에서 간단하게 ‘문장’으로 표기되며 채점 척도표상에서는 문장 구조의 다양성을 측정하는 지표로 제시되고 있다.

수치가 6.5로서 다른 평가 영역과는 확연히 다른 채점 경향을 보여주고 있어서 부적합한 평가 기준으로 판명되었다.

Measr	+Examinee	-Rater	-Topic	-Item	Scale
4	+	+	+	+	(5)
	*				

3	+	+	+	+	+
	*				
	*				
2	+	+	+	+	4
	*				

	*				
	*				
	**				
1	+	+	+	+	+
	**				
	**				
	*			예증	---

			T1		
		Rater 2		문법 문장 반복	
*	0	*	T5	*문단연결	*
	**				
	*	Rater 1		주장	3
			T7		
	*			문단연결	
	*			서론	
-1	+	+	+	+	(1)
Measr	* = 1	-Rater	-Topic	-Item	Scale

그림 1. 전체변인 척도표.

표 6
각 평가영역별 분산 분석표

평가영역	Logit	SE	Fair-M	Infit MS	Infit z	Outfit MS	Outfit z
서본결구성	-.72	.12	3.92	1.85	6.5	1.76	5.9
문단간연결	-.58	.12	3.87	.97	-.2	1.00	.0
주장제시	-.22	.11	3.72	.75	-2.5	.81	-1.9
문장간연결	-.01	.11	3.63	.91	-.7	.92	-.7
단어적절성	.14	.11	3.56	.67	-3.5	.68	-3.4
문장다양성	.20	.11	3.53	.75	-2.5	.75	-2.5
문법정확성	.21	.11	3.53	.86	-1.3	.87	-1.3
반복효율성	.23	.11	3.52	1.16	1.4	1.16	1.4
예증효과성	.76	.11	3.27	1.12	1.2	1.12	1.1

Separation=3.87, Reliability =.94

지금까지 제시된 수치적 측면에서의 문장 다양성의 신뢰성 조사를 보완하고자, 연구자는 채점자 두 명과 문장의 다양성에 관련하여 짧은 인터뷰를 수행하였다. 문장의 다양성이라는 세부평가항목에 의거하여 학습자들에게 점수를 부여하는 것이 가능한가에 대한 질문에 대하여, 채점자 1(그림 1에서 rater 1)은 한 문장내의 다양한 품사들의 활용 여부와 접속사의 사용여부를 중요한 기준으로 삼은 반면, 채점자 2(그림 1에서 rater 2)는 단문이나 복문의 대체적인 비율을 염두해 두고서 채점한다고 보고하였다. 채점자들의 답변은 다음과 같다.

문장구조에 있어서 뭐 단순하게 주어 동사 그리고 그것을 기본으로 해서 다양한 품사를 넣은 문장을 구사했는지 아니면 다양한 접속사를 활용을 해서 문장의 길이를 늘렸는지 그런 면들을 보았죠. (채점자 1)

의식을 안하고 보면 잘 모르겠지만...문장의 복잡성을 생각하면서 보면 일단..느껴지긴 느껴지는 것 같아요. 너무 단순하게 썼던가 아님 복문이 어느정도 들어가 있던가... 그런데 이사람이 얼마만큼의 비율로 복잡하게 썼고.. 복문의 비율이 어땠고 정확한.. 그런 것은 확 모르겠고 다만 종합적으로.. 한번 보고 이건 어느 정도 될까다... (채점자 2)

채점자들은 문장의 다양성을 평가함에 있어서 한가지 심각한 문제를 제시하였는데, 그것은 문장의 다양성이라는 평가 항목이 다른 항목과

비교하여 독립적인 평가영역을 확보하고 있지 못한 것에 대한 느낌의 토로였다. 채점자 1은 내용적 측면을, 채점자 2는 구성적 측면을 같이 고려하여 문장의 다양성을 평가하였음을 보고하였다.

단순히 문장 길이만 늘린 것이 아니고 정말로 다양한 예시를 들면서 읽는 사람으로 하여금 이해를 도울 수 있는 그런 문장들을 보았을 때 문장 구조가 다양하다고 본 거죠... (채점자 1)

솔직히 말하자면 채점할 때 문장의 다양성이라던가 복잡성에 대하여 깊이 생각하지 않았어요. 문장이..전체 문장의 논리의 흐름이 좋다던가..구성이 좋다던가 그러면 점수를 알아서 잘 준 것 같아요.. 그냥 그 점수를 따라간거죠. (채점자 2)

FACETS의 결과와 인터뷰 결과를 정리하자면, 문장의 다양성이라는 평가 항목은 타 평가영역과 비교하여 과적합 영역으로 판명될 수 있으며, 이것은 채점자가 타 영역과는 차별화되지 않는 비슷한 기준을 적용하여 문장의 다양성을 평가했다는 의미로 해석할 수 있다. 채점자들의 인터뷰를 통해서 드러난 사실은, 채점자들이 내용이나 구성과 같은 다른 평가 요소를 같이 고려하여 문장의 다양성을 평가하였다는 것을 알 수 있었다.

V. 논의 및 결론

본 연구는 작문의 유창성과 관련된 지표인 문장 다양성의 세부 지표들의 향상 정도, 세부지표와 총괄점수와의 상관관계, 그리고 문장 다양성의 평가항목으로서의 신뢰도등을 조사하였다. 그 결과를 요약하면 다음과 같다.

먼저, 연구질문 1인 문장 다양성의 세부 지표의 향상 정도를 보면 문장 다양성과 관련된 세부 지표들은 학생들의 향상도에 대하여 유의미한 정보를 제공하지 못하였다. 비록 단문비율은 어느 정도 감소하는 추세를 보였으나, 학생들이 네차례의 시험을 치루면서 그들의 영어논술 실력이 유의미하게 향상되고 있음을 증명하는 총괄점수의 명확한 추이와 비교한다면 문장 다양성 관련 세부지표들이 유의미하게 변화하고 있다고 단언하기 어렵다. 따라서 자연스럽게 연구질문 2에서 제시된 세부지표들과 총괄점수와의 상관관계 역시 대부분 통계적으로 무의미하다는 사실이 발견되었다.

문장 다양성의 신뢰도를 정량적·정성적으로 살펴본 연구질문 3에서는, 문장 다양성의 평가 항목이 과적합으로 판정되어, 타영역과는 차별화되지 못하고 있음이 드러났다. 채점자와의 인터뷰를 통해서, 문장 다양성이 다른 영역의 점수를 그대로 따라가는 한편, 채점자의 지나친 안전지향적 채점전략이 작용했음을 발견하였다. 결과적으로, 문장 다양성은 평가항목으로서 꼭 필요한 만큼의 고유 변수가 없는 지나친 신뢰도를 가지고 있어서 적합한 평가 요소라고 주장하기에 다소 문제가 있다.

본 연구는 몇 가지 교육적, 이론적 시사점을 제시하고 있다. 우선은, 문장의 다양성이 작문의 유창성에 포함되는 개념으로 본다면(Hunt, 1965), 본 논문은 유창성과 관련해서도 여러 가지 시사점들을 제시하고 있다고 볼 수 있다. 특히 유창성이라는 개념이 외국어로서의 작문 분야의 학자들에게 많이 논의되는 주제인 만큼 유창성에 대한 논의는 매우 유의미하다고 판단된다.

영작문 평가 상황의 실용성과 결부하여 작문의 유창성 구인을 평가상황에 맞게 재정립할 필요가 있다. 기존의 문헌에서 제시된 단어 혹은 그 이상의 세밀한 수준의 분석 단위로 유창성을 정의내린다는 것은, 실행 불가능한 평가기준을 만드는 것에 다름아니다. 예를 들어, 작문의 유창성을 평균 기록지속단위의 관점(Chenoweth & Hayes, 2001)이나 작문시 적절한 어휘를 선택할 수 있는 속도의 관점(Snelling, van Gelderen & de Glopper, 2004)으로 유창성을 평가한다는 것이 평가 현장의 속성을 고려할 때 매우 어렵다는 것이다. 혹은 추상적인 관점, 이를테면 유창성을 학습자의 기존 언어지식 체제에 접근하여 해당 아이디어와 텍스트를 신속하게 꺼낼 수 있는 능력(van Gelderen & Oostdam, 2002)으로만 정의내린다면 유창성에 대하여 무엇을 평가해야 하는지 혼란이 가중될 수 있다.

평가 상황에 적합한 유창성의 평가 구인을 정의하기가 어렵다면, 관련된 지표들로서 유창성을 간접 평가하는 방안도 고려해 볼 수 있을 것이다. 이런 간접지표들로서는 본 논문에서 제시한 바와 같이 문장의 다양성도 고려될 수 있겠지만 이와 같은 통사적 관점뿐만 아니라 상황에 맞는 단어의 적절한 사용이나 논리의 원활한 흐름 같은 내용적 측면도 고려하여 유창성을 종합적으로 평가할 수 있다. 특히 본 연구를 통해서 문장의 다양성이 학습자 수행측면이나 평가측면에서 의미있는 지표가 되기 어렵다는 결과를 고려한다면, 다른 평가 영역을 유창성 측정과 관련지어 고려할 필요가 있다. 다시 말해, 유창성을 하나의 명목화된 지표로서만 고집할 것이 아니고, 다수의 세부평가영역들에 포괄적으로 걸쳐있는 경계가 느슨한 평가구인으로서 진지하게 숙고해야 한다.

채점자 측면에서는 평가의 신뢰도를 높이기 위하여 문장의 유창성에 대한 구인을 정확하게 이해하는 것이 필요하다. 특별히 이지연(2007a)의 연구에서 제시된 논문수집자료 방식을 주목할 만 한데, 채점자가 직접 작문샘플들을 시범 채점하면서 유창성에 대한 요인들을 파악하고 분류하며 구인을 정의하는 실증적 방식은 평가의 신뢰도를 높이는 데 매우 효과적이다. 즉, 채점자를 평가설계 과정에 직접 참여시켜서 실증적 데이터에 의거하여 척도표를 제작 방식(empirical-driven construct)은 채점 과정의 투명성과 신뢰도를 높이는 작용을 할 수 있다(Fulcher, 2003; Sung Sam Hwang, 2010).

본 연구는 몇 가지 한계점을 지니고 있다. 첫번째로, 기존의 연구에서 자주 수행되는 단어 수준의 분석이 없다는 점이다. 따라서, 무오류 T-unit의 평균 단어수와 같은 단어를 활용한 세부지표 분석이 이루어지지 않았다. 이런 점으로 인하여, 본 논문에서 문장 다양성에 해당되는 관련 지표들을 완전히 활용했다고 볼 수가 없다. 만약 Yeon Hee Choi(2010)나 Larsen-Freeman(1978)의 연구와 같이 단어수를 기본 단위로 해서 무오류 절의 평균 단어수 같은 세밀한 분석을 시행하였다면 결과는 달리 나올 수 있었을 것이다. 그러나, 짧은 시간에 세부영역에 대한 점수를 매기고 거기에 따라 피드백을 제공하는 영어수행평가 상황에서 단어수준의 정밀 분석이 그 실용성에 문제가 있다는 것을 감안한다면, 어느 정도는 채점자가 인상적으로 파악이 가능한 전체문장에서의 단문, 복문, 중문의 비율을 고려해 보는 것은 가치가 있을 것이다.

두번째로는 깊이 있는 질적연구가 포함되지 못했다는 점이다. 본 연구가 연구질문 3을 위하여 전문 채점자와의 인터뷰 내용을 부분적으로 발췌하기는 했지만, 일회성 인터뷰에 그쳤고 그 인터뷰 내용의 신뢰성을 뒷받침할 수 있는 다른 정성적 기법이 부재했다는 점이다. 부가적으로 채점 과정을 좀 더 자세히 관찰해 볼 수 있는 사고표출기법등이 연구방법으로 수행되었다면 본 논문의 연구방법에 대한 신뢰도가 고양될 수 있었을 것이다.

마지막으로, 본 연구의 실험참여자 및 채점자 수가 충분하지 못하다는 점이다. 적은 실험 참여자의 수를 보완하기 위하여 각 수험자마다 4차례 이상의 논술시험을 수행하도록 했을 지라도, 본 논문의 결과를 다른 상황에 적용시켜 일반화하기에는 다소 무리가 있다. 본 논문에서 나온 결과에 대한 신중한 해석과 적용이 필요하다고 할 수 있다.

위와 같은 한계점을 인식하여 앞으로의 추후 연구에서는 문장의 유창성 및 다양성을 측정할 수 있는 관련 평가영역들을 탐색하여 발굴하고 다양한 세부분석지표들을 이용하여 작문의 유창성 및 다양성과 수험자의 전반적인 영어작문 수행실적과의 관계를 밝혀야 할 것이다. 본 논문에서는 연구자가 애초에 예상했던 의미있는 결과를 도출하지 못한

측면이 있는데, 이는 부분적으로 본 논문에서 제시된 세부지표들에 대한 치밀한 사전 분석이 부족했기 때문이라고 본다. 따라서 추후 연구에서는 관련 전문가들을 참여시켜 문장의 유창성 및 다양성을 나타내는 지표들을 공동 협의하에 추출하고 그런 지표들이 과연 작문평가과 관련된 다양한 평가요소들과 어떤 관계가 있는지를 살펴볼 필요가 있다. 마지막으로, 정성적·정량적 연구방법을 적절하게 사용하여 채점자가 갖는 유창성에 대한 인식의 연구, 유창성 영역과 관련된 세부분야의 점수 결정과정 혹은 다양한 채점 전략을 종합적으로 연구해 볼 필요가 있을 것이다.

참고문헌

- 박지순, 서세정. (2009). 쓰기 텍스트 분석을 통한 한국어 학습자의 통사적 숙달도 측정 연구. *언어와 문화*, 5(2), 151-173.
- 유재임. (2005). T-Unit 분석방법과 문장합성연습. *신영어영문학*, 31, 259-287.
- 이상권, 윤인섭. (2008). 영어 텍스트의 논술 형식에 대한 연구. *영어교육연구*, 36, 101-116.
- 이지연. (2007a). 영어 교사들의 주관적 평가 방법에 따른 영작문 평가시 평가항목의 타당도. *응용언어학*, 23(1), 147-172.
- 이지연. (2007b). 고등학교 영작문에 있어서의 유창성의 개념에 대한 고찰: 실제 작문 샘플 분석을 바탕으로. *Foreign Languages Education*, 14(3), 339-364.
- 이지연. (2009). 영작문 평가를 위한 채점자 훈련의 방안. *현대영어교육*, 10(3), 217-241.
- 이현구. (2001). 대학영어작문에서 나타나는 전치사 오류 분석. *현대영어교육*, 2(2), 239-253.
- Brumfit, C. (1984). *Communicative methodology in language teaching: The roles of fluency and accuracy*. Cambridge, UK: Cambridge University Press.
- Brown, J. D. (2003, May). *Promoting fluency in EFL classrooms*. Paper presented at the 2nd Annual JALT Pan-SIG Conference, Kyoto, Japan. Retrieved July 1, 2011, from <http://www.jalt.org/pansig/2003/HTML/Brown.htm>
- Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in writing: Generating text in L1 and L2. *Written Communication*, 18(1), 80-98.
- Choi, Yeon Hee. (2010). Fluency measures in L2 Writing. *Korean Journal of English Language and Linguistics*, 10(4), 913-937.

- Dechert, H. W., & Raupach, M. (1980). *Temporal variables in speech: Studies in honor of Frieda Goldman-Eisler*. The Hague, The Netherlands: Mouton.
- East, M. (2009). Evaluating the reliability of a detailed analytic scoring rubric for foreign language writing. *Assessing Writing, 14*, 88-115.
- Fleiss, J. L. (1986). *The design and analysis of clinical experiments*. New York: Wiley.
- Fulcher, G. (2003). *Testing second language speaking*. Harlow, UK: Pearson Longman.
- Hamp-Lyons, L. (1991). Pre-text: Task-related influences on the writer. In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts* (pp. 87-107). Norwood, NJ: Ablex.
- Hunt, K. W. (1965). *Grammatical structures written at three grade levels*. Champaign, IL: NCTE.
- Hwang, Sung Sam. (2010). *Discourse analysis of EFL interview test at secondary school: A focus on oral fluency*. Unpublished Doctoral Dissertation, Chung-Ang University, Seoul.
- Ishikawa, S. (1995). Objective measurement of low-proficiency EFL narrative writing. *Journal of Second Language Writing, 4*(1), 51-69.
- Larsen-Freeman, D. (1978). An ESL index of development. *TESOL Quarterly, 12*, 439-448.
- Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics, 27*(4), 590-619.
- Latif, A. (2009). Toward a new process-based indicator for measuring writing fluency: Evidence from L2 writers' think-aloud protocols. *The Canadian Modern Language Review, 65*(4), 531-558.
- Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40*(3), 387-417.
- Miller, K. S. (2000). Academic writers on-line: investigating pausing in the production of text. *Language Teaching Research, 4*(2), 123-148.
- Sasaki, M. (2000). Toward an empirical model of EFL writing processes: An exploratory study. *Journal of Second Language Writing, 9*(3), 259-291.
- Snellings, P., van Gelderen, A., & de Glopper, K. (2004). Validating a test of second language written lexical retrieval: A new measure of fluency in written language production. *Language Testing, 21*(2), 174-201.
- Spalding, E., & Cummins, G. (1998). It was the best of times. It was a waste of Time: University of Kentucky students' views of writing under KERA. *Assessing Writing, 5*(2), 167-199.

- Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, 17, 84-119.
- van Gelderen, A., & Oostdam, R. (2002). Improving linguistic fluency for writing: Effects of explicitness and focus of instruction. *Educational Studies in Language and Literature*, 2, 239-270.
- Weigle, S. C. (2002). *Assessing writing*. London, UK: Cambridge University Press.
- Willard-Traub, M., Decker, E., Reed, R., & Johnston, J. (1999). The development of large-scale portfolio placement assessment at the University of Michigan: 1992-1998. *Assessing Writing*, 6(1), 41-84.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy & complexity* (Technical Report #17). Honolulu: University of Hawaii, Second Language Teaching and Curriculum Center.

부록

채점기준표 (Scoring Guide)

채점영역	세부영역	해당 사항
구 조	서론본론결론	서론, 본론, 결론 구조가 없거나 서론, 본론, 결론 중에 하나 이상이 빠져있다.
	문장간 연결성	문장과 문장을 연결하는 접속사가 거의 없다. 문단내 글의 흐름이 끊겨 메시지를 알 수 없다. 주제문(thesis statement)이 전혀 없다.
	문단간 연결성	의미가 서로 연결되지 못할 정도이고, 문단과 문단을 연결하는 연결사가 없다.
내 용	토픽 준수	제시된 토픽과는 거리가 멀다.
	구체적 예증제시	구체적인 증거나 예시가 전혀 제시되어 있지 않다.
	자신의 주장표현	자신의 주장이 전혀 보이지 않고, 기존의 설만 반복하고 있다.
	무의미한 반복	문장 안에서 불필요한 단어 사용이나, 문단 내 특정 단어의 남용이 심각하다. 문장 형태도 똑같은, 같은 의미를 가진 문장들이 자주 출몰한다.
어휘사용	문법의 정확성	거의 모든 분야에서 문법상 오류를 저지르고, 의미전달을 막는 오류 투성이다.
	적절한 단어사용	문장의 흐름상 전혀 의미가 맞지 않는 단어, 숙어가 각 문단마다 네 다섯개이상 쓰이고, 전체 에세이에서 철자상의 오류가 15개 이상 자주 발생하여 글의 의미를 이해하기 힘들다.
	다양한 문장구조	제대로 끝나지 못한 문장들이 대다수이고 간단한 단문만 나타나고 있다.

홍효기

중앙대학교 영어교육과

156-756 서울시 동작구 흑석동 221

Tel: 02-880-4538

Email: dkny2351@snu.ac.kr

이호

중앙대학교 영어교육과

156-756 서울시 동작구 흑석동 221

Tel: 02-820-5393

Email: holee@cau.ac.kr

Received 15 July 2011

Revised 25 August 2011

Accepted 3 September 2011