

## 영작문평가 상황에서 FACETS와 Coh-Metrix를 활용한 에세이 분석: 학습자 단면을 중심으로

이호  
(중앙대학교)

Lee, Ho. 2019. Analyzing L2 essay with FACETS and Coh-Metrix in L2 writing assessment. *Korean Journal of English Language and Linguistics* 19-4, 796-816. The current study aims to identify Coh-Metrix indices that significantly predict writing proficiency and well discriminate misfit learners from in-fit and overfit learners. 85 undergraduate and graduate students took an on-line essay test. Two expert raters scored their essays in terms of three analytic areas such as content, organization, and language use. The researcher conducted FACETS analysis whose purpose is to form sub-groups in regard to proficiency level and fit information. Then, the researcher ran Coh-Metrix program to investigate semantic, syntactic, and linguistic features of the students' essays. The result showed that sentence length, argument overlap, and syntax similarity in adjacent sentences meaningfully discriminated proficient group from less proficient one. In addition, essays written by misfit learners showed lower value of Flesch Reading Ease than essays by counterparts. This study ended with the implication that further studies in L2 writing assessment need to explore learner dimension.

**Keywords:** many-facet Rasch model, Coh-Metrix, FACETS, Flesch Reading ease, writing assessment

### 1. 서론

최근 들어, Coh-Metrix를 활용한 연구가 국내에서 활발하게 발표되고 있다. 국내의 Coh-Metrix를 활용한 연구 주제는 교과서의 심층적 지문 분석(이승환 2013, 전문기, 임인재 2009, 홍성우, 송해성 2019), 영어 읽기평가 상황에서의 독해 지문 분석(김준구 2018, 신유선, 정영경 2018, 장지연 2018), 영작문 분야에서 다양한 담화 유형 간 혹은 원어민과 비원어민 비교(안수진 2018, 전문기, 최윤희 2019, 정미경, 김남국 2014)등으로 기술될 수 있다. 이와 같이, 다수의 국내 Coh-Metrix 연구들은 기존 지문이나 코퍼스를 심층 분석하여 교재·교육과정 영역이나 담화분석, 읽기평가의 타당화 연구 영역 등에서 의미 있는 결과들을 내놓고 있다.

이외에도, Coh-Metrix는 제2언어 쓰기평가영역에서 수험자의 쓰기 양상을 이해하고, 채

점자의 채점 행태를 분석함에 있어서 유의미한 역할을 할 수 있다. 영어쓰기평가에서의 Coh-Metrix의 활용성을 주목한 연구들은 채점자의 총괄점수(holistic score)와 Coh-Metrix의 세부 지표간의 상호작용을 면밀히 살펴보면서, 채점자의 판정에 결정적으로 작용하는 세부 요소들을 밝혀내기도 하였다(Aryadoust and Liu 2015, Xu and Liu 2011). 또한, 영어쓰기 평가와 관련 있는 다수의 해외 연구들은 Coh-Metrix 분석을 통해 학습자의 쓰기 능숙도 판별과 밀접한 관련이 있는 세부 요인들을 제시하고자 하였다(Crossley and McNamara 2012, Khushik and Huhta 2019). 이와 같이, Coh-Metrix가 영어쓰기 평가 분야에서 적절하게 활용될 수 있는 분석방식을 제공함에도 불구하고, 국내에서 영어쓰기 평가 분야와 연계한 Coh-Metrix 연구가 상대적으로 미진한 편이다.

특히, 영어쓰기 평가 혹은 영어말하기 평가 상황에서 평가의 관련 변인들을 인지하고, 그 변인들을 구체적으로 분석해주는 FACETS 프로그램과 에세이의 텍스트를 여러 가지 세부 지표로서 정밀하게 분석하는 Coh-Metrix 프로그램의 결합은 영어쓰기 평가 연구영역의 지평을 넓힐 수 있는 계기가 될 수 있다(Aryadoust and Liu 2015). 예를 들어, FACETS를 통해서 분석대상이 되는 수험자 집단이나 영역을 선정하고, Coh-Metrix를 통해서 관련 텍스트를 정밀하게 분석함으로써 해당 단면(facet)의 특징을 제대로 파악할 수 있을 것이다. 이러한 두 분석 방식의 상호 보완 관계를 활용한 국내외 연구는 아직 많이 부족한 편이며, FACETS와 Coh-Metrix의 결합한 연구방식은 그동안 베일에 쌓여있는 부적합 학습자의 특징을 살펴보는데 많은 도움이 될 수 있다.

본 연구는 다국면라쉬모형(many-facet Rasch model)에 근거한 FACETS와 텍스트의 담화적, 언어적 측면을 세밀하게 분석하는 Coh-Metrix를 활용하여 영어쓰기능력에 상대적으로 차이가 있는 상·하위 집단 간, 그리고 부적합 학습자와 적합·과적합 학생들 간의 차이를 다양한 세부지표에 걸쳐 비교·분석해 보고자 한다. 연구 질문을 정리하면 다음과 같다.

- 1) 영어쓰기능력치가 높은 학습자 집단과 상대적으로 낮은 학습자 집단 간의 차이를 변별해주는 세부 지표는 무엇이며, 그것이 시사하는 바가 무엇인가?
- 2) FACETS를 통하여 규명된 부적합집단과 적합집단, 과적합집단 간의 차이를 가장 잘 보여주는 에세이 텍스트의 세부지표는 무엇이며, 그것이 시사하는 바가 무엇인가?

## 2. 이론적 배경

### 2.1 Coh-Metrix와 영어쓰기 수행능력 비교

Coh-Metrix는 다양한 세부지표를 가지고 텍스트의 담화적 측면과 언어적 측면을 심층적으로 분석할 수 있도록 하는 프로그램이다. 기존의 언어 담화 분석 연구들이 텍스트 기반에

근거한 표층 부호 분석 단계에 머무는 경우가 많았지만(전문기, 임인재 2009), Coh-Metrix는 표층 단계를 넘어서 텍스트가 전달하고자 하는 의미와 내용에 주목해서 분석하는 심층 단계의 분석까지 가능토록 한다(Graesser, McNamara and Kulikowich 2011, McNamara, Graesser, McCarthy and Cai 2014).

Coh-Metrix의 이러한 매력적인 장점으로 인하여, Coh-Metrix는 응용언어학 분야에서 제2언어 학습자가 작성한 에세이를 분석하는데 중요하게 사용되었다(Crossley and McNamara 2012). 본 연구에서는, Coh-Metrix를 활용한 기존 연구들 중에서, 세 가지 항목(통사적 측면, 어휘적 측면, 응집성 측면)과 제2언어 학습자의 쓰기 능숙도의 관계를 조명한 연구들을 정리하여 제시하고자 한다.

통사적 측면과 관련하여, 연구자들은 통사적 복잡성이 학습자의 수준을 변별함에 있어서 유의미한 항목인가에 대하여 입장을 달리 하였다. Khushik과 Huhta(2019)는 유럽공통언어 기준(CEFR: Common European Framework of Reference)에 규정된 기초(A1), 초급(A2), 중급(B1) 단계에 속한 학생들의 에세이를 통사적 복잡성(syntactic complexity) 측면에서 비교 분석하였는데, Coh-Metrix 지표 중 문장의 길이, 종속절, 그리고 구의 밀도에서 집단 간 큰 차이를 보였다고 보고하였다. McNamara, Crossely와 McCarthy(2010)는 고수준의 에세이는 저수준의 에세이와 비교하여 상대적으로 높은 통사적 복잡성을 지니고 있다고 언급하였다. Crossely와 McNamara(2011)의 후속 연구 역시, 통사적 복잡성이 포함된 언어적 정교성과 관련된 지표들이 채점자의 에세이 평가에 실질적인 영향을 미친다고 주장하였다. 반면에, Aryadoust와 Liu(2015)는 통사적 복잡성과 통사유형별 밀도는 학습자의 수준을 변별하는데 영향을 미치지 않는다고 주장하였다.

Coh-Metrix의 응집성(cohesion)관련 지표가 학습자의 능숙도를 실질적으로 변별할 수 있는지에 대해서도 상반되는 의견이 제시되었다. Crossley와 McNamara(2011, 2012)는 고수준의 제2언어 학습자는 저수준의 학습자와 비교하여 응집력 있는 에세이를 쓰기보다는 언어적으로 정교한 텍스트를 양산한다고 기술하며, 응집성 관련 지표가 학습자의 능숙도를 변별할 수 있을지에 대하여 의문을 표시하였다. 이와는 달리, Aryadoust와 Liu(2015)는 McNamara와 2인의 연구에서 낮은 변별력을 가진 것으로 보고되었던 참조적 적합성 항목의 제시 방식에 문제를 제기하면서, 참조적 응집성 관련 세부 지표들이 상호 깊게 관련되어 있으며 채점자는 수험자의 에세이 수준을 결정함에 있어 응집성 항목 내 세부 요인들을 모두 연계하여 고려한다고 반론을 제시하였다. Xu와 Lui의 연구(2011)에서도, 참조적 응집성 항목에서 높은 평가를 받은 학생들이 전반적으로 질 높은 에세이를 썼다는 사실을 밝혔다.

어휘적 측면과 관련해서는, 어휘의 빈도 수준과 어휘의 다양성이 학습자의 능숙도를 변별하는데 중요한 역할을 한다고 보고하였다. McNamara, Crossely와 McCarthy(2010)는 고수준의 쓰기 학습자와 저수준의 쓰기 학습자를 변별할 수 있는 항목으로 어휘적 다양성, 그리고 단어의 빈도를 제시하였으며, Crossley와 McNamara(2012)는 어휘적 다양성, 단어의 빈도, 어휘 의미성(word meaningfulness), 단어의 친숙도, 시상의 반복(aspect repetition)이

학습자의 능숙도의 중요한 예측인자로 사용될 수 있음을 밝혔다.

## 2.2 FACETS를 활용한 학습자 적합도 연구

기존의 FACETS을 활용한 영어작문 관련 연구들은 채점자 단면을 주로 조명하였다. 기존 연구에서 학습자 단면에 대한 사안은 편향 분석(bias analysis)의 일부분으로 짧게 기술되거나, 채점자 분석의 과정에서 부분적으로 다루어지는 정도였다(이창수 2014, Knoch 2009, Schaefer 2008, Shin and Sul 2005). 따라서 학습자 단면을 초점으로 하여 부적합 혹은 과적합 판정을 받은 학생들의 에세이를 심도 깊게 살펴본 연구는 거의 없었다. 본 논문에서는 FACETS를 활용하여 부적합 채점자를 탐색한 주요 논문들을 살펴본 후, 부적합성이 학습자와 어떻게 관련이 있는지 간접적으로 조명하고자 한다. FACETS를 활용하여 채점자의 부적합 경향에 대하여 직접적으로 혹은 간접적으로 살펴본 연구들은 채점자와 평가 영역간의 편향분석(Schaefer 2008), 원어민과 비원어민간의 채점자 차이(Johnson and Lim 2009, Lee 2009), 경험이 적은 채점자와 경험이 많은 전문가 채점자 간 차이(Weigle 1998), 학습자가 채점자로 참여하였을 때 관찰되는 평가 경향성(이호, 김하나 2009)등을 주제로 삼았다.

Schaefer(2008)는 채점자의 편향분석을 통하여 부적합 채점 요인에 대하여 살펴보았다. 그는 채점자들이 내용과 구조에 대하여 혹독하게 채점하는 경향을 보이면, 상대적으로 언어 사용 측면에서는 관용적으로 채점하는 모습을 보인다고 보고하였다. 일부의 채점자들은 상대적으로 쓰기능력이 좋은 학습자들에게는 더 엄격하게 채점하는 반면, 낮은 쓰기능력을 가진 학습자들에게는 관용적으로 채점한다고 보고함으로써, 이러한 채점자의 개인적 경향이 부적합 요인을 양산할 수 있음을 제시하였다.

Lee(2009)는 원어민 채점자와 한국인 채점자의 비교를 통해서 한국인 채점자가 문법에 대하여 지나치게 낮은 점수를 주는 반면, 원어민 채점자는 내용 측면에서 특이할 정도로 혹독한 채점 경향을 보였다고 보고하였다. 한국인 채점자가 문법 영역을 평가할 때와 원어민 채점자가 내용 영역을 평가했을 때 두 가지 경우에서 모두 부적합 수치가 제시되었는데, 이는 채점자의 원어민 여부가 특정 평가 영역과 결부되었을 때 부적합한 평가 행태를 양산할 수 있음을 보여주는 결과였다. Hyland와 Anan(2006) 연구 역시 원어민 채점자가 비원어민 채점자에 비하여 문법 실수에 대하여 더 관용적임을 보여주고 있다. 이와는 반대로, Johnson과 Lim(2009)는 채점자의 언어적 배경에 따라 발생할 수 있는 편향은 매우 제한적이며, 실제 시험에 크게 영향을 미치지 않는다고 보고함으로써, 채점자의 언어적 배경이 채점자의 부적합성의 하나의 요인이 될 수 있다는 주장에 회의적이었다.

평가자 훈련을 수행하지 않는 초보 채점자의 경우에도, 부적합성이 발견되고 있다(Weigle 1998). 초보 채점자는 자신의 평가에 대하여 자신감이 부족하며 일관적이지 못한 채점 성향을 보이기에 부적합한 채점자로 분류되는 경우가 종종 있으며(Weigle 2002), 그러한 채점자의 신뢰성을 높이기 위한 방법으로 평가자 훈련이 제안되고 있다. 학습자가 평가자로

참여하는 동료평가 상황에서도 부적합성이 발견되고 있는데, 대면평가 상황에서의 채점자들은 지나치게 관용적인 평가 성향을 보인 반면, 익명평가 상황에서의 채점자들은 전문 평가자나 자기평가와는 상이한 평가 행태를 보임으로서, 부적합한 평가그룹으로 분류되었음을 보고하였다(이호, 김하나 2009).

부적합 수험자의 판정에 영향을 미칠 수 있는 요인에 대해서는 채점자와 학습자간 편향분석과 채점자의 구술자료 분석을 통해서 간접적으로 접근할 수 있다(Bond and Fox 2001). 이창수(2014)는 채점자의 구술 자료와 학생들의 에세이 자료를 동시에 분석하였는데, 그 결과 채점자가 특정 영역에 과도한 비중을 두어 평가하고, 그 영역에서 학생이 잘하거나 못할 경우 부분적인 혹은 국소적인 편향이 발생하였다고 보고하였다. Yi(2008)는 채점자의 사고구술(think aloud) 분석을 통하여, 채점자가 수험자가 누구인지를 알고 있을 경우, 그 수험자에 대한 긍정적인 혹은 부정적인 인식으로 인하여 편향이 발생할 수도 있음을 제시하기도 하였다. 그러나 이러한 편향분석은 부적합 학생들의 에세이가 가질 수 있는 특징에 대하여 간접적으로 추측할 수 있는 근거를 제시할 뿐, 심도 깊은 분석에는 한계가 있을 수 있다. 비록 FACETS 측정도구에서 채점자 변수가 수험자의 적합도 판정에 큰 영향을 미칠 수 밖에 없지만(장소영, 신동일 2009), 채점자의 점수 데이터가 아닌 부적합 학생들의 에세이로부터 공통적으로 드러나는 언어적 혹은 담화적 특성을 추출하여 분석하는 것은 매우 유의미할 것이다.

### 3. 연구 방법

#### 3.1 연구 대상

본 연구에는 서울의 대학교에서 영어를 전공하는 총 85명의 학생들이 참여하였다. 85명의 참여자 중 31명은 교육대학원 학생이며, 54명은 학부 학생이다. 참여자들은 연구자가 강의하는 두 개의 강좌(한 과목은 학부학생을 위한 과목, 다른 과목은 교육대학원생들을 위한 과목)를 수강하였고, 그 강좌에서 필수적으로 요구되는 영어 에세이 쓰기를 수행하였다. 학부 학생 중 48명은 1학년 학생이며, 나머지 6명은 고학년 학생이다.

본 연구에서 85명 학생들의 에세이를 채점할 2인은 한국인으로서, 모두 영어평가 분야의 전문가이며, 채점 당시 교육경력이 3년 이상이고 풍부한 평가 경험을 가지고 있었다. 2명의 전문 평가자는 채점자 신뢰도를 높이기 위한 일환으로, 본 연구에서 활용된 척도표를 상호 논의하였고 5명 학생들에 대한 시범 채점 과정을 통해 채점 기준을 조율하는 채점자 워크숍을 수행하였다. 본 채점이 끝난 직후 두 채점자의 채점자 신뢰도를 분석 한 결과, 전체적으로 .71의 급내상관계수(intra-class correlation) 수치를 기록하여 두 채점자간 차이는 양호한 것으로 드러났다.

## 3.2 연구 도구

### 3.2.1 에세이 문항

본 연구의 에세이 주제는 ‘대안교육으로서의 홈스쿨링’이며 구체적인 주제는 “Some people claim that completing high school at home is a good alternative to regular schooling, and others claim that home schooling is not effective and thus not recommendable. Which view do you agree with? Take a position and defend it. Be sure to follow the guidelines below.”이다. 학생들은 서론-본론-결론의 논증적 에세이의 형식을 지켜서, 20-25줄의 분량으로 4개의 문단을 작성하도록 요구되었다.

### 3.2.2 척도표

학생들의 에세이쓰기를 평가할 척도표는 크게 내용, 구조, 언어사용의 세 가지 평가항목을 가지고 있었으며, 내용과 언어사용의 경우 각 2가지 세부 영역이 포함되어 있었다. 내용 영역의 경우, ‘내용의 중복성과 전달성’(3점) 그리고 ‘근거의 효과성과 충실성’(3점) 두 가지 세부영역을 합쳐 6점 체제였다. 구조 영역의 경우는 문장 간 연결성과 문단 간 연결성, 그리고 서본결 구조 체제를 갖추었는지를 평가하는 것으로 6점의 배점이 주어졌다. 마지막 언어사용의 경우 ‘문장의 유창성(통사적 다양성 및 복잡성)’(3점)과 ‘언어사용의 오류’(3점)의 6점 체제였다. 본 연구의 척도표는 이전 수강 학생들의 점수 데이터를 기반으로 개발되고 수정된 현장친화형 척도(empirical-based rubric)로서, 채점의 편이성과 신뢰성이 양호한 것이 특징이다(Fulcher 2003). 본 연구의 FACETS분석을 위하여, 연구자는 각 영역별로 동등하게 6점 체제를 갖추도록 하였으며, 따라서 총 18점 만점이 되도록 하였다.

## 3.3 연구 절차

영어에세이 시험은 수업 이외의 시간에 온라인으로 학기 초에 시행되었다. 교수자는 학교가 자체적으로 운영하는 학습관리시스템(LMS: Learning Management System)을 활용하여, 에세이 문항을 공시한 온라인 과제방을 50분 동안 한시적으로 개설하였다. 그 시간동안 학생들은 주어진 영어에세이 주제를 보고 아래한글이나 워드로 작성하여 과제방에 제출하였다. 학생들은 시험 보는 도중 사전이나 인터넷 자료를 찾아보는 것은 금지됨을 권고받았으며, 제출 시간을 엄격하게 준수하도록 하였다. 교수자는 학생들의 부정행위를 방지하기 위하여 제출된 에세이는 카피킬러와 같은 프로그램을 활용하여 표절을 확인할 것임을 사전 공지하였다.

시험이 끝난 후, 교수자는 영어평가 분야에 전문적 식견을 갖춘 1인을 초빙하여 학생들의

에세이를 평가하였다. 채점자 신뢰도를 높이기 위한 일환으로, 연구자와 전문가는 1시간 동안의 채점 워크숍을 가졌다. 채점 워크숍 동안 채점기준표에 대하여 상세한 논의 과정을 거쳤고, 본 채점에 앞서 시범적으로 5명의 에세이에 대한 채점을 수행하여 그 결과를 상호 비교하였다.

연구자는 채점자의 점수 자료에 근거하여 FACETS 분석을 시행하였다. FACETS 분석은 참여자들을 두 가지 기준으로 분류하기 위하여 시행되었다. 하나는 능력치에 의한 분류로서, 능력치가 상대적으로 높은 상위그룹과 능력치가 낮은 하위그룹으로 분류하였다. 다른 하나는 적합도에 의한 분류로서, FACETS의 적합도 분석에 따라 부적합군, 적합군, 과적합군의 세 집단으로 분류하였다. FACETS 분석이 끝난 후, 다음 단계로 연구자는 학습자의 영어 에세이 작문 능력에 대한 세밀한 정보를 제공하는 26개 지표에서의 집단 간 차이를 조사하기 위하여, 읽기/쓰기 텍스트 정밀 분석 도구인 Coh-Metrix분석을 시행하였다.

### 3.4 연구 분석 방법

#### 3.4.1 FACETS 분석

연구자는 학생들의 능력치와 적합도를 살펴보기 위하여 FACETS라는 측정학적 도구를 사용하였다. FACETS이라는 프로그램은 다단면 라쉬 검정(Multi-facet Rasch analysis)에 의거하여 데이터를 분석하는데, 단면은 평가의 측정과정에서 개입되는 변인을 일컫는다. 본 연구에서는 3가지 단면을 포함시켰는데, 참여자 능력에 의한 차이, 채점자의 엄격도의 차이, 세 가지 평가영역(구조, 내용, 언어사용)의 난이도에 따라 학생의 점수가 결정된다는 가정을 하였다. 이 FACETS프로그램은 각 평가 단면별로 개별 항목(case)들에 대한 세밀한 분석을 수행하고, 개별 항목들의 측정치가 예측값의 범위 내에 있는지를 조사하여 그 결과를 적합/부적합의 수치로 제시해 준다. 이러한 세부항목의 측정치는 로짓(logit)을 통하여 제시되고, 양수의 로짓은 음수의 로짓에 비하여 학습자의 능력이 높거나 채점자의 채점 엄격성이 높다는 것을 의미한다(Bond and Fox 2001).

본 연구에서는 학습자 단면에 대한 적합도 분석(fit statistics)을 제시하여 그 결과에 따라 학생들을 부적합 집단, 적합집단, 과적합 집단으로 분류하였다. 여기에서, 학습자 단면에서의 적합도 수치는 특정 학생이 다른 학생들과 비교하여 특이한 양상을 보이는지 판단하는 근거로서 사용된다. 학생들이 본인의 능력 수준과는 걸맞지 않는, 예견치 못한 응답 반응을 보일 때 부적합(misfit)으로 판명되는데(장소영, 신동일 2009) 그 주요한 이유로는 두 가지를 들 수 있다. 하나는 학습자 측면의 요인으로서, 그 학습자가 특정 세부 영역에서 다른 영역과는 달리 지나치게 잘하거나 지나치게 못하는 경우가 발견되고 그러한 경향성이 그 학습자에게서만 발견되는 경우이다. 두 번째로는 채점자가 어떤 요인으로 인해서 해당 학생에 대하여 채점을 잘 못 했을 가능성이 있다. 위의 두 가지 요인으로 인한 부적합의 출현은 학습

자의 참능력(true ability)을 안정적으로 측정하지 못한다는 증거이며, 평가 점수가 외부변수의 유의미한 영향으로 인하여 학습자의 진정한 능력을 대변하지 못하는 결과를 초래한다. 이는 평가의 신뢰도를 심각하게 위협할 수 있는 요인이며 더 나아가 평가의 타당도를 훼손할 수 있다. 반면, 과적합(overfit)으로 판명된 학생들의 경우, 해당 학생의 측정치가 지나치게 측정모델에 적합할 때 발생된다. 이는 해당 수험자들이 전반적으로 유사한 내용의 에세이를 작성하며, 동일한 에세이 서술 방식을 보일 때 발견될 수 있다. 보통은 부적합이 평가 상황에서 심각한 문제를 야기하며(Bond and Fox 2001), 부적합의 원인을 찾는 것은 매우 중요한 문제로 대두되었다. 이와 같은 적합/부적합 여부를 판정하는 근거는 내적합도/외적합도 제곱평균과 z검정 값인데, 본 논문에서는 적합도 제곱평균만을 포함시켰다. 보통 적합도 제곱평균 값이 .7에서 1.3 사이일 때 적합으로 판정하며, 1.3 이상일 경우에는 부적합, .7 이하인 경우 과적합으로 판정된다(McNamara 1996). 본 연구에서는 부적합 학생들의 에세이를 적합 혹은 과적합 학생들의 에세이와 비교하고, 통계적으로 유의미한 집단 간의 차이를 보이는 분석지표를 도출해 보고자 한다.

### 3.4.2 Coh-Metrix 분석과 집단 간 차이 검정

현재 웹 기반으로 무료로 이용할 수 있는 Coh-Metrix 3.0(<http://tool.cohmetrix.com>)은 기초산출치(descriptive), 텍스트 용이성 주성분 분석(text easability principal component scores), 참조적 응집성(referential cohesion), 잠재적 의미 분석(latent semantic analysis), 어휘다양성(lexical diversity), 연결사(connectives), 상황모형(situational model), 통사적 복잡성(syntactic complexity), 통사구조 밀도(syntactic pattern density), 어휘정보(word information), 이독성(readability)의 11가지 항목에서 총 106개의 세부지표에 대한 측정치를 제공한다.

이중 가장 잘 알려져 있는 기초 산출치는 단어 수, 문장 수, 문장의 평균 길이와 같은 정보를 알려주는 것으로 텍스트 길이가 길거나 어휘 수가 많다면 텍스트의 난이도가 증가하는 것으로 알려져 있다(전문기, 임인재 2009). 응집성 관련 항목으로서, 참조적 응집성은 표층적 분석 단계에서 인접 문장 혹은 한 문단 내 문장들 사이에서 내용어의 중복 여부를 알려주는 것으로 텍스트 기반 단계에서의 응집성을 측정하는데(McNamara 외 3인 2014), 이 수치가 높을수록 해당 텍스트를 이해하기가 용이해진다(김준구 2018). 연결사 항목도 텍스트 기반의 접속사 항목을 중심으로 분석한 것으로, 연결사 항목의 측정치가 높으면, 문장 간의 표층적 응집성이 좋은 것으로 판명된다. 그러나 논리적 접속사와 같은 특정 연결사가 지문에서 많이 사용될 경우, 학습자는 내용을 이해하기 위해 더 많은 작업 기억을 소요시킨다는 언급도 있다(신유선, 정영경 2018). 반면, 잠재적 의미 분석은 코퍼스 자료에 근거하여 의미론적으로 비슷한 단어들이 상호 연관되어 있는 정도를 제시한 것으로, 이 항목에서의 측정치는 0에서 1의 범위 내에서 형성되고 있으며 1의 수치에 가깝다면 의미적 응집성



이 높아서, 텍스트를 이해하기가 용이하다(McNamara 외 3인 2014).

통사론적 관련 정보를 알려주는 항목으로서, 통사적 복잡성과 통사유형별 밀도가 있다. 인접 문장 통사 동질성을 제외한 통사적 복잡성과 통사유형별 밀도 내 세부지표들은 그 수치가 높으면, 텍스트를 이해하기에 많은 작업 기억 용량을 요구하기 때문에 텍스트의 이해 난이도가 올라간다고 한다(신유선, 정영경 2018).

어휘적 측면과 관련된 항목으로서, 어휘의 다양성은 내용어의 타입과 토큰의 비율이 많이 쓰인다. 이 비율이 높다면 중복되지 않는 새로운 내용어가 출현하는 것으로 텍스트의 난이도를 높아진다고 보고 있다(김준구, 2018). 또 다른 항목인 단어정보는 품사별 출현율 정보, CELEX 코퍼스를 활용한 어휘 빈도, 심리적 평점(psychological ratings)에 근거한 어휘정보로 분류하여 볼 수 있다. 여기서 출현율(incidence score)은 전체 1,000 단어의 텍스트에서 해당 품사 영역과 관련된 단어가 차지하는 비율을 계산한 것으로서, 상대적 빈도(relative frequency)라고도 한다. 어휘 빈도는 CELEX 데이터베이스 분석에 근거하여, 분석대상 지문에 포함된 내용어들의 빈출 수준을 제시한다. 높은 빈출 수준을 지닌 어휘를 많이 포함한 텍스트는 이해하기가 용이한 반면, 낮은 빈출 수준을 가진 어휘들은 지문의 난이도를 높이는 역할을 한다. 심리적 평점에 근거한 어휘 정보는 단어의 친숙도(familiarity), 단어의 구체성(concreteness), 단어의 심상성(imagability) 등에 대하여 평정자들의 평점에 근거하여 그 측정치를 제시하는데(McNamara 외 3인 2014), 단어의 친숙도나 구체성, 심상성 등에서의 높은 수치는 텍스트를 이해하기가 용이하다는 의미를 암시하고 있다(정미경, 김남국 2014).

마지막으로, Coh-Metrix는 이독성에 대한 세 가지 세부 지표에 대한 정보를 제공하는데, 본 연구에서는 Flesch 이독지수(Flesch reading ease)와 Flesch-Kincaid 등급(Flesch-Kincaid Grade Level)을 포함시켰다. Flesch 이독지수는 단어 길이(단어 당 음절 수 평균)와 문장 길이(문장당 단어 수)를 근거로 하여, 독해의 난이도를 0에서 100점의 범위로 측정하는데, 높은 이독성 수치(대략적으로 80 이상)는 읽기 지문을 이해하기가 매우 쉽다는 것으로 해석된다(성일호 2014). 반면, Flesch-Kincaid 등급은 미국 학교의 학년 등급과 대응하도록 산출치를 만들었는데, 높은 등급은 텍스트의 난이도가 높다는 것을 의미한다(이승환 2013).

본 연구를 위하여, 연구자는 Coh-Metrix의 106개 지표 중 텍스트 용이성 주성분 분석과 상황모형을 제외한 9가지 항목에서의 26개의 세부지표분석을 수행하였다. 연구자는 86명이 작성한 에세이 텍스트를 일일이 복사하여 Coh-Metrix 분석 사이트(<http://tool.cohmetrix.com>)에 입력한 후, 표 1에 제시한 세부지표 수치를 추출하여 세부항목별로 학생들의 쓰기 양상을 살펴보았다. 다음 단계로, FACETS 분석을 통해서 확정된 집단별로 세부지표의 수치를 배치하는 과정을 거쳤다. 지표선정은 이전 Coh-Metrix 연구들(김준구 2018, 장지연 2018, 홍성우, 송해성 2019)을 참조하여, 기존 연구에서 활용이 많이 되었던 지표들 중 본 연구에 밀접하게 관련이 있을 것으로 추정되는 지표들을 추출하였다. 다음으로, 본 연구에서 쓰인 평가 척도표의 세부기준들과 추출된 지표들을 비교하여, 쓰기 평가의 세부기준들을 잘 반영하고 있다고 판단되는 26개 지표를 최종 확정하였다.

집단 간 평균 차이가 있는지를 검정하기 위하여, 연구자는 능력치와 적합도를 독립변수로, Coh-Metrix에서 추출된 지표 수치를 비독립변수로 지정하여 SPSS 분석을 수행하였다. Shapiro-Wilk의 정규성 검정을 통해 모수통계일 경우 일원분산분석(one-way ANOVA)을 시행하였고, 비모수통계일 경우 2개의 능력치 집단을 비교할 때에는 Mann-Whitney U검정을, 3개의 적합군 집단을 비교할 때에는 Kruskal-Wallis H검정을 시행하였다. 해당 지표의 데이터가 정규성 분포를 이루고 있음을 확인한 이후에는 Levene의 등분산 검정을 실시하였다. 본 연구에서 정규성이 확인된 지표에서는 모두 등분산성이 충족되었음을 확인하였다.

표 1. 본 연구에서 활용된 Coh-Metrix 분석 지표

대항목	세부지표명(지표번호)*	지표설명
기초 산출치	문장 수(2)	총 문장의 수
	단어 수(3)	총 단어의 수
	문장평균길이(6)	문장당 평균 단어의 수
참조적 응집성	인접논항중복(29)	동일한 명사 혹은 명사와 그와 대응하는 대명사가 인접 문장에서 출현하는 여부
	전체논항중복(32)	동일 논항이 전체문장에서 출현하는 여부
잠재적 의미 분석	인접문장 의미유사성(38)	의미가 비슷한 단어가 인접 문장에서 출현
	전체문장 의미유사성(42)	문단내 조합할 수 있는 문장 쌍에서 의미가 비슷한 단어가 연이어 발생되는 여부
어휘 다양성	내용어 타입-토큰(46)	내용어를 기준으로 한 Type-Token 비율
연결어	접속사 총 출현비율(50)	1000단어에서 접속사가 출현하는 비율
	인과 접속사(51)	1000단어에서 인과접속사가 출현하는 비율
	시간 접속사(54)	1000단어에서 시간접속사가 출현하는 비율
	부가 접속사(56)	1000단어에서 부가접속사가 출현하는 비율
통사적 복잡성	본동사앞 단어밀도(67)	주절의 주동사 앞에 위치한 평균 단어 수
	명사구 당 수식어(68)	명사구 수식어의 평균 단어 수
	인접문장 통사동질성(72)	인접 문장 간 문장 구조의 동질성
통사 유형별 밀도	명사구 밀도(74)	1000단어에서 명사구가 출현하는 비율
	동사구 밀도(75)	1000단어에서 동사구가 출현하는 비율
	부사구 밀도(76)	1000단어에서 부사구가 출현하는 비율
단어 정보	명사출현비율(82)	1000단어에서 명사의 출현 비율
	동사출현비율(83)	1000단어에서 동사의 출현 비율
	내용어 빈출도(92)	CELEX 데이터베이스와 분석대상 지문에 함께 포함된 내용어들의 빈도 평균
	내용어 친숙도(96)	내용어의 익숙함에 대한 여부
가독성	Flesch 이독지수 (104)	해당지문의 읽기 난이도에 대한 지수
	Flesch-Kincaid등급 (105)	해당지문의 읽기 난이도 등급

Note. (지표번호)\*는 Coh-Metrix 분석도구에서 각 지표에 명기된 번호임.

### 4. 결과 및 논의

#### 4.1 FACETS 분석

그림 1은 평가의 모든 변인 내 데이터를 서로 비교할 수 있도록 제시한 척도표이다. 그림의 'Measr'칸에서 제시된 측정치의 양수는 수험자의 높은 능력치, 채점자의 엄격한 채점경향, 측정영역의 높은 난이도를 나타낸다. 절반 이상의 많은 학생들이 0 이상의 양수의 측정치를 기록한 것으로 보아, 학생들이 상대적으로 안정된 에세이 쓰기 능력을 가지고 있다고 할 수 있다. 본 논문에서는 상위 집단과 하위 집단에 소속되는 학생들의 수를 비슷하게 조절하기 위하여, 측정치 1.5를 기준으로 그 이상의 로짓을 가진 학생들을 상대적으로 높은 능력치를 가진 집단으로, 1.5 미만의 능력치를 기록한 학생들은 상대적으로 낮은 능력치를 가진 집단으로 분류하였다. 그림에서, 45명의 학생들이 1.5 미만의 능력치를 받았고 이 학생들은 상대적으로 능력치가 낮은 그룹, 1.5 이상의 능력치를 부여받은 40명의 학생들은 상대적으로 능력치가 높은 집단으로 분류되었다. 부가적으로, 교수자(그림의 rater 부분에서 'Prof')가 전문 평가자('Expert')보다 더 엄격한 채점경향을 보였고, 평가영역에서는 구조의 난이도가 제일 높았으며, 언어사용 영역은 다른 두 영역에 비하여 난이도가 낮았다.

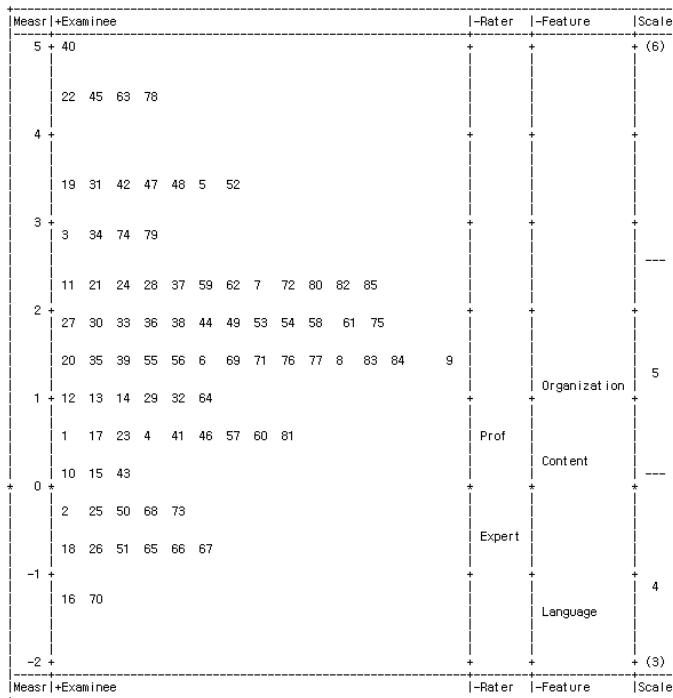


그림 1. 3단면(능력치, 채점자, 평가영역)에 대한 전체 분포도

표 2에서는 수험자 변인에 대한 FACETS의 분석결과를 제시하였다. 내적합도와 외적합도 지수 모두 1.3을 넘는 부적합군 학생들은 23명이었다. 반면, 적합 수치 범위(.7에서 1.3)에 해당되는 학생들은 모두 32명이며, 내적합도와 외적합도 지수에서 모두 .7 이하를 기록한 과적합 학생들은 30명이었다. 수험자 분리 신뢰도는 .73으로 평가영역들의 난이도에 따라 학생들의 능력분포가 어느 정도는 적절하게 이루어져 있음을 제시하고 있다. 다만, 수험자의 능력 편차가 현격하게 크다고 볼 수 없음이 1.65의 분리도 지표에서 알 수 있었다.

표 2. 적합도별 수험자 변인에 대한 정보

능력 지수	부적합 근접군			적합군				과적합 근접군			
	학생 ID	내적 합도	외적 합도	능력 지수	학생 ID	내적 합도	외적 합도	능력 지수	학생 ID	내적 합도	외적 합도
4.37	22	1.46	3.26	2.30	7	1.21	.99	1.83	36	.70	.65
-.76	65	2.88	2.71	1.83	54	1.19	1.04	2.30	62	.70	.65
-.31	68	2.33	2.11	4.37	45	1.17	.96	2.30	82	.70	.65
2.30	28	2.02	2.33	4.37	63	1.17	.96	1.39	69	.65	.63
1.39	9	2.19	2.32	2.82	74	1.17	.89	1.39	6	.59	.70
.55	48	1.90	2.08	1.39	20	1.16	1.09	1.39	56	.61	.57
.12	15	1.92	2.08	1.39	39	1.14	1.14	-.76	67	.57	.61
2.30	72	1.25	2.01	1.39	84	1.14	1.14	1.83	58	.59	.56
-.76	28	1.88	1.95	-1.23	70	1.10	1.10	1.39	76	.59	.56
-1.23	16	1.69	1.92	1.83	38	1.06	1.06	1.39	83	.59	.56
.97	14	1.66	1.66	2.30	59	1.05	1.05	.97	13	.58	.56
3.45	5	1.33	1.64	1.39	71	1.05	1.05	1.83	30	.58	.54
3.45	47	1.33	1.64	.55	17	.94	.94	1.39	8	.49	.49
-.76	66	1.56	1.58	1.39	77	.98	.98	1.83	49	.45	.46
1.39	55	1.57	1.53	-.31	2	.90	.90	3.45	19	.55	.45
.12	10	1.51	1.47	1.83	53	.94	.94	3.45	48	.55	.45
.12	43	1.51	1.47	2.82	3	.83	.83	.55	1	.44	.46
.97	64	1.46	1.39	2.82	34	.83	.83	1.39	35	.42	.42
.97	12	1.44	1.45	2.82	79	.83	.83	2.30	21	.42	.48
3.45	31	1.30	1.41	.55	81	.88	.88	2.30	85	.42	.48
1.83	27	1.36	1.17	.97	32	.87	.87	2.30	11	.42	.40
.55	4	1.30	1.30	1.83	44	.88	.88	-.31	73	.38	.35
.55	57	1.30	1.30	1.83	61	.88	.88	4.37	78	.50	.27
				2.30	37	.90	.90	2.30	24	.27	.28
				-.76	18	.75	.75	2.30	80	.27	.28
				1.83	33	.75	.75	.55	80	.24	.25
				1.83	75	.75	.75	.97	29	.22	.24
				-.31	25	.69	.69	-.76	51	.20	.21
				3.45	42	.98	.98	-.31	50	.19	.18
				3.45	52	.98	.98	5.74	40	.19	.18
				.55	41	.72	.72				
				.55	23	.66	.66				

Note. 분리도(separation) 1.65, 분리 신뢰도(separation reliability) .73

#### 4. 2 능력집단별 Coh-Matrix 분석

표 3은 정규성 검정을 통해서 비모수 통계기법에 적합하다고 판명된 6개의 지표를 대상으로, Mann-Whitney U 검정을 수행한 결과를 제시하였다. 그 결과 능력치 상위 집단과 능

력치 하위 집단 간 유의미한 차이가 발생된 지표는 단어수가 유일하였다. 이는 상위능력치에 속한 집단이 하위 집단에 비하여 현격하게 많은 단어를 활용하여 에세이를 썼다는 사실을 입증하고 있다. 다만, 잠재적 의미 분석 측면에서 두 집단 간의 차이가 통계적으로 유의미하지 않다하더라도, 상위집단이 하위집단에 비하여 의미적으로 중첩된 단어들을 문장 간 혹은 문단 내에 더 많이 활용함으로써, 의미론적인 응집성을 높인 것으로 보인다.

표 3. 능력집단별 Mann-Whitney U검정 결과

대항목	세부지표	능력집단별 평균순위	z	유의확률
기초 산출치	문장수	상(44.83), 하(41.38)	.64	.519
	단어수	상(54.02), 하(33.20)	3.88	.000
잠재적 의미 분석	인접 문장 의미유사성	상(45.92), 하(40.40)	1.03	.303
	전체 문장 의미유사성	상(46.84), 하(39.59)	1.35	.176
통사적 복잡성 단어정보	본동사앞 단어 밀도	상(48.16), 하(38.41)	1.82	.069
	내용어 친숙도	상(45.65), 하(40.64)	.07	.351

표 4는 20개의 지표에 대하여 일원분석변량을 수행한 결과이다. 두 집단 간 차이가 통계적으로 유의미했던 지표는 문장 평균길이, 인접논항중복, 인접 문장 통사적 동질성이었다. 특히, F값(10.65)으로 보아, 두 집단 간 차이가 가장 선명하게 드러났던 지표는 문장의 평균 길이로서 상위집단이 하위집단에 비하여 문장마다 평균적으로 훨씬 더 많은 단어를 사용했음을 알 수 있었다. 인접 문장 통사적 동질성도 주목할 만한데, 하위집단이 상위집단에 비하여 동일한 통사적 구조를 인접 문장에서 연이어 구사하는 비율이 높다는 것을 알 수 있다. 다만, 상위집단은 하위집단에 비하여 동일한 명사 혹은 명사와 그 명사를 지칭하는 대명사를 인접 문장에서 더 자주 연이어 활용함으로써, 문장 간 응집성을 높이하고자 하였다. 반면, 단어정보 영역에서는 두 집단 간 유의미한 차이가 발견되지 않았는데, 이는 에세이 토픽이 상·하위 학습자들의 전공과 연계된 교육 관련된 주제이며, 본 연구의 참여자들이 평소에 교육 관련 어휘들을 잘 이해하고 사용했기 때문이라는 추정이 가능하다.

표 4. 능력집단별 ANOVA 결과

대항목	세부지표	집단별 지표 평균	F	유의확률
기초 산출치	문장평균길이	상(16.16), 하(13.89)	10.65	.00
참조적 응집성	인접논항중복	상(.62), 하(.54)	4.27	.04
	전체논항중복	상(.50), 하(.45)	3.16	.08
어휘다양성	내용어 타입-토큰	상(.62), 하(.61)	.19	.66
연결어	접속사 총 출현비율	상(106.5), 하(104.6)	.27	.61
	인과 접속사	상(36.98), 하(36.11)	.07	.78
	시간 접속사	상(17.00), 하(19.30)	1.64	.20
	부가 접속사	상(54.33), 하(53.95)	.03	.87
통사적 복잡성	명사구 당 수식어	상(.72), 하(.68)	1.00	.32
	인접문장 통사적 동질성	상(.11), 하(.13)	8.80	.00
통사 유형별 밀도	명사구 밀도	상(365.0), 하(370.7)	1.26	.26
	동사구 밀도	상(227.7), 하(226.8)	.02	.88
	부사구 밀도	상(36.41), 하(38.97)	1.08	.30
단어정보	명사출현비율	상(275.3), 하(282.6)	1.76	.19
	동사출현비율	상(115.4), 하(114.4)	.05	.82
	내용어 빈출도	상(2.48), 하(2.48)	.06	.81
	내용어 구체성	상(359.3), 하(357.1)	.61	.43
	내용어 심상성	상(399.3), 하(397.5)	.39	.53
가독성	Flesch가독지수	상(57.28), 하(57.33)	.00	.98
	Flesch-Kincaid 등급	상(9.28), 하(8.71)	2.31	.13

## 4.3 적합집단별 Coh-Matrix 분석

표 5는 8개의 세부 지표에서의 세 개의 집단(부적합집단, 적합집단, 과적합집단) 간 차이를 살펴보기 위하여, 비모수 통계방식인 Kruskal-Wallis 검정을 시행한 결과이다. 여기에서는 모든 세부 지표에서 집단 간 유의미한 차이는 발견되지 않았다. 하지만, 부적합집단의 학생들이 적합 집단이나 과적합 집단 학생들에 비하여 더 적은 문장수나 단어수를 보였고, 문장 간 혹은 문단 내에서 의미적으로 연관 있는 단어를 덜 배치함으로써 글의 의미론적인 응집성이 상대적으로 떨어짐을 볼 수 있었다.

표 5. 적합집단별 Kruskal-Wallis H검정 결과

대항목	세부지표	적합집단별 평균순위*	H	유의확률
기초 산출치	문장수	1(35.4), 2(45.0), 3(46.7.)	3.10	.212
	단어수	1(33.3), 2(46.9), 3(46.3)	4.92	.085
	문장평균길이	1(41.6), 2(46.2), 3(40.7)	.87	.648
잠재적 의미 분석	인접 문장 의미유사성	1(40.9), 2(45.2), 3(42.3)	.47	.793
	전체 문장 의미유사성	1(40.3), 2(40.7), 3(47.5)	1.53	.464
연결어 단어정보	인과접속사	1(49.1), 2(45.7), 3(35.4)	4.67	.097
	동사출현비율	1(45.8), 2(39.2), 3(44.9)	1.22	.543
	내용어 친숙도	1(36.6), 2(47.0), 3(43.7)	2.42	.299

Note. \*적합집단별 평균순위에서 1은 '부적합', 2는 '적합', 3은 '과적합' 집단을 의미

표 6과 표 7은 18개의 세부지표에서의 적합도 집단 간 차이를 제시한 것이다. 그 결과, 두 개의 세부지표(부가접속사, Flesch 가독지수)에서만 유의미한 차이가 발견되었다. 부적합 학생 집단이 상대적으로 낮은 부가적 접속사 활용도를 보이고 있는데, 이는 부적합 학생 집단이 ‘and’나 ‘moreover’ 등의 부가적 접속사를 이용하여 사실이나 근거를 보충적으로 설명하는 부분이 다른 집단에 비하여 미약했던 것으로 추정된다. 다만 적합집단 간 인과적 접속사에서 의미 있는 차이가 발견되지 않았는데, 이는 본 연구 참여자들이 본인의 근거에 대한 이유를 논리적으로 제시하는 것이 논증적 에세이에서 중요하다는 인식을 지니고 있어서, 대다수의 참여자들이 인과 접속사를 적극 활용한 측면이 있었던 것으로 추정된다. 특히, 부적합 집단이 과적합 집단에 비하여 유의미하게 낮은 Flesch 가독지수를 기록했다는 것이 발견되었는데, 이는 부적합집단의 에세이의 이해난이도가 높다는 사실과 연결된다. 이러한 경향성은 Flesch-Kincaid 등급에서도 관찰되는데, 이는 채점자가 부적합집단의 에세이 내용을 이해하고 평가함에 있어서, 상대적으로 높은 난이도를 느꼈을 가능성이 있다는 사실을 암시한다.

표 6. 적합집단별 ANOVA 결과

대항목	세부지표	집단별 지표 평균	F	유의확률
참조적 응답성	인접논항중복	1(.56), 2(.61), 3(.56)	.88	.418
	전체논항중복	1(.47), 2(.50), 3(.44)	1.16	.319
어휘다양성	내용어 타입-토큰	1(.63), 2(.60), 3(.62)	1.64	.200
연결어	접속사 총 출현비율	1(103), 2(110), 3(102)	1.93	.151
	시간 접속사	1(16.9), 2(16.8), 3(20.8)	2.25	.112
	부가 접속사	1(49.9), 2(57.6), 3(53.7)	3.57	.033
통사적 복잡성	본동사앞 단어밀도	1(3.90), 2(3.96), 3(4.11)	.18	.840
	명사구 당 수식어	1(.69), 2(.70), 3(.71)	.20	.819
	인접문장 통사적 동질성	1(.11), 2(.12), 3(.13)	2.32	.104
통사 유형별 밀도	명사구 밀도	1(372), 2(367), 3(366)	.46	.636
	동사구 밀도	1(228), 2(226), 3(229)	.10	.907
	부사구 밀도	1(36.8), 2(39.0), 3(37.2)	.33	.724
단어정보	명사출현비율	1(270), 2(281), 3(284)	2.04	.137
	내용어 빈출도	1(2.49), 2(2.47), 3(2.49)	.40	.675
	내용어 구체성	1(356), 2(358), 3(360)	.69	.507
	내용어 심상성	1(396), 2(398), 3(401)	.80	.452
가독성	Fleschg독지수	1(53.7), 2(57.6), 3(59.8)	3.14	.048
	Flesch-Kincaid 등급	1(9.52), 2(9.02), 3(8.53)	2.17	.121

Note. \*적합집단별 지표평균에서 1은 ‘부적합’, 2는 ‘적합’, 3은 ‘과적합’ 집단을 의미

표 7. 적합집단별 다중 비교

종속변수	(I) 적합	(J) 적합	평균차이(I-J)	표준오차	유의확률
부가접속사	부적합군	적합군	-7.69	2.90	.026
		과적합군	-3.80	2.94	.403
	적합군	부적합군	7.69	2.90	.026
		과적합군	3.90	2.69	.322
	과적합군	적합군	3.80	2.94	.403
		부적합군	-3.90	2.69	.322
Flesch 이독지수	부적합군	적합군	-3.96	2.42	.236
		과적합군	-6.12	2.45	.038
	적합군	부적합군	3.96	2.42	.236
		과적합군	-2.16	2.25	.604
	과적합군	적합군	6.12	2.45	.038
		부적합군	2.16	2.25	.604

## 5. 결론

본 연구는 FACETS 프로그램을 활용하여 85명의 학습자들을 영어쓰기 능숙도와 적합도에 따라 분류하고, Coh-Metrix를 사용하여 수험자 집단별로 에세이 텍스트를 분석한 후, 집단 간 유의미한 차이를 보이는 세부지표를 살펴보았다. 연구 결과를 연구문제별로 살펴보면 다음과 같다.

첫째, 학습자의 영어 쓰기 능숙도 차이를 가장 잘 드러내는 지표로는 문장의 평균길이(한 문장 당 평균 단어 수)였으며 인접논항중복과 인접 문장 통사적 동질성 등이 유의미한 지표로 선정되었다. 상위집단의 에세이는 하위집단에 비하여 더 높은 응집성을 보이는 경향을 보였으며, 반면 하위집단은 통사적으로 비슷한 구조의 문장을 연이어 사용한 측면이 보였다. 이는 상대적으로 높은 쓰기 능숙도를 지닌 수험자가 낮은 수준의 능숙도를 지닌 수험자에 비하여 통사적으로 복잡한 문장을 쓴다는 기존의 연구(Crossely and McNamara 2011, McNamara 외 2인 2010)와 부합되며, 상위수준의 에세이에서 인접논항중복 산출치가 더 높다는 Xu와 Lui(2011)의 연구결과를 확증한다. 이는 상위수준의 학습자들이 상황에 맞는 다양한 문장 구조를 활용하면서, 동시에 내용의 원활한 흐름을 위하여 문장 간 연계에 많은 주의를 기울이고 있는 것으로 해석할 수 있다.

둘째, 적합도 집단(부적합, 적합, 과적합)간 유의미한 차이를 보여주는 지표는 부가접속사의 사용과 Flesch 이독성이었다. 세부적으로, 부적합 집단은 적합집단에 비하여 부가접속사의 사용 비율이 낮았으며, 부적합 집단은 과적합 집단에 비하여 Flesch 이독지수가 유의미하게 낮았다. 이 사실을 채점자 측면에서 조명하자면, 과적합군의 학생들이 작성한 에세이로부터는 에세이의 내용이나 구성에 있어서 일괄적으로 적용될 수 있는 경향성을 인지하며 수월하게 과적합군의 학생들의 에세이를 읽을 수 있는 반면, 부적합 집단의 학생이 작성한



에세이는 다른 학생들과는 상이한 글쓰기 양식을 이해하기가 쉽지 않았다는 점을 추정할 수 있다. 이것은 부적합 집단의 에세이가 과적합 집단의 에세이와 비교하여 유의미한 가독성의 차이를 불러일으킨 것으로 판명할 수 있다.

위에 제시된 연구 결과에 대하여, 본 연구가 가지는 다음과 같은 한계점 때문에 그 결과의 해석과 적용에 신중할 필요가 있다. 첫 번째, 학생들의 능력치를 단 한 번의 시험을 가지고 상위그룹과 하위그룹으로 분류했다는 점이다. 본 연구에서 학생들의 능력치를 절대적인 준거에 근거하여 상위와 하위로 나누기보다는, 한 번의 시험 결과로 학생들의 능숙도를 분류하였다는 점이다. 즉, 상·하위 그룹의 구분이 상대적이며 그 구분의 객관적 근거가 부족할 수 있다는 점을 지적할 수 있다. 추후 연구에서는, 경우에 따라 상위 집단과 하위 집단에 모두 분류될 수 있는 중간 정도의 학생들을 가급적 배제하기 위하여, 상위집단과 하위집단의 절대적인 기준을 명료화한 후, 학생들을 그 기준에 따라 배치할 필요가 있다. 두 번째로는, 본 연구의 참여자들이 대부분 서울지역의 대학교 혹은 교육대학원에 재학 중인 영어 관련 전공자들이므로 본 연구의 결과를 다른 지역, 다른 집단으로 일반화하기에 제약이 따른다는 점이다.

이러한 한계점에도 불구하고, 본 연구는 유의미한 시사점을 지니고 있다. 교육적 시사점으로서, 본 연구자는 채점자 교육과정에서 부적합 에세이를 이해하고 적절하게 대처할 수 있는 지침서를 만들어 활용할 것을 적극 제안한다. 지침서의 사전 작업으로서, 부적합 에세이를 수집한 데이터베이스를 구축한 후, Coh-Metrix와 같은 담화/언어 분석 프로그램을 활용하여 부적합 학습자가 쓴 텍스트를 정밀하게 분석할 필요가 있다. 이를 바탕으로, 부적합의 발생 요인이 채점자 변수와 밀접한 관련이 있는 항목(이를테면, 이독성 지표)과 부적합 발생 요인이 해당 수험자들의 특이한 쓰기 양상과 관련이 있는 항목을 분류하여, 채점자 교육 시 예비채점자들에게 각 항목과 관련된 문제점과 적절하게 대처하는 방법 등이 숙지되어야 할 것이다.

또한, 본 연구자는 연구 방법론적 측면에서 부적합 학습자와 같은 방외인(outlier)에 대한 관심과 연구가 필요함을 제안하고자 한다. 계량적 통계에서 종종 방외인은 데이터 선별 과정에서 의도적으로 제거되거나(Xu and Liu 2011), 설명되지 않는 오류인자로 취급받기도 한다. 그러나 방외인은 그가 속한 샘플군의 행동에 대해 획기적인 해석이나 설명을 내포하기도 하며, 연구자에게 새로운 연구 과제를 제시하기도 한다(Kane 2013). 본 연구 주제인 부적합 학습자들의 특성을 논의하고 그것이 평가의 타당성에 어떤 영향을 미치는지 치열하게 논의하는 과정, 즉 타당성 논증과정(validity argument)을 거치게 된다면, 부적합 학습자의 분석과 이해를 통해 평가의 타당도를 고양할 수 있는 계기가 될 수 있을 것이다.

본 연구는 영어쓰기 평가 상황에서 학습자의 능력치나 적합도와 같은 학습자 단면에 초점을 맞추어 에세이 텍스트를 분석하였다. 결론적으로, 평가 변인으로서의 학습자 단면에 대한 연구는 아직 부족한 편이며, 정량적 접근뿐 아니라 정성적 연구 방법으로 학습자 단면을 더 심도 깊게 다룰 필요가 있음을 환기하고자 한다.

## References

- 김준구(Kim, J.). 2018. Coh-Metrix를 활용한 중학교 학업성취도 평가 영어 지문의 코퍼스 언어학적 분석(A corpus-based investigation of reading passages in the national assessment of educational achievement English test using Coh-Metrix). 《중등영어교육》(*Secondary English Education*) 11-2, 27-51.
- 성일호(Sung, I.). 2014. 이독성 공식과 Coh-Metrix를 활용한 우리나라 고등학교 영어교과서 이독성 분석(An analysis of Korean high school English textbooks through readability formulae and Coh-Metrix). 《영어영문학연구》(*Studies in English Language & Literature*) 40-4, 299-320.
- 신유선·정영경(Shin, Y. S. and Y. Jong). 2018. 코메트릭스(Coh-Metrix)를 활용한 공무원 7급과 9급 영어 시험 지문 분석: 어휘 다양성, 통사적 복잡성, 응집성과 이독성 지수를 중심으로(Analyzing reading passages in the English tests for civil servants with Coh-Metrix: Focusing on lexical diversity, syntactic complexity, cohesion, and readability). 《영미어문학》(*Studies in British and American Language and Literature*) 129, 143-161.
- 안수진(Ahn, S.). 2018. 한국인 대학생과 영어 원어민 대학생의 영어 설명문, 논설문의 코메트릭스 분석을 통한 비교(An analysis of Coh-Metrix on the differences in English expository and argumentative writing of Korean and native English university student). 《새한영어영문학》(*New Korean Journal of English Language and Literature*) 60-3, 177-205.
- 이승환(Lee, S-H.). 2013. 코메트릭스(Coh-Metrix)를 이용한 고등학교 영어1과 영어2 교과서 문어 텍스트의 코퍼스 언어학적 난이도 분석(A corpus-based analysis of the difficulty of the reading materials in high school English 1 and 2 textbooks using Coh-Metrix). 《외국어교육연구》(*Studies in Foreign Language Education*) 27-2, 131-148.
- 이창수(Lee, C. S.). 2014. 『한국 중고등학교 교사들의 영어작문 채점 훈련에서 나타나는 채점자 편향과 그 원인』 (*Rater Bias and Its Source in Secondary School Writing Assessment*). 미출간 박사학위 논문, 서울: 서울대학교.
- 이호·김하나(Lee, H. and Ha-Na Kim). 2009. 영작문에서의 대면평가와 익명평가의 비교 연구(Exploring learners' blind and face-to-face assessment in an L2 argumentative essay-writing test). 《영어학》(*Korean Journal of English Language and Linguistics*) 9-3. 403-427.
- 장소영·신동일(Jang, S-Y. and D. I. Shin). 2009. 『언어교육평가 연구를 위한 FACETS 프로그램』 (*Guidelines of FACETS Program for Language Testing Researchers*).

- 서울: 글로벌콘텐츠(Seoul: Global Contents).
- 장지연(Chang, J-Y.) 2018. Coh-Metrix를 통한 2017-2018 수능 영어 독해 지문 비교: 기초 산출치, 이독성, 용이성을 중심으로(A comparison of 2017-2018 CSAT reading passage via Coh-Metrix: Focusing on descriptive, readability, and easibility measures). 《외국어교육》(*Foreign Languages Education*) 25-4, 81-106.
- 전문기·임인재(Jeon, M-G and I-J. Lim). 2009. Coh-Metrix를 이용한 중학교 1학년 개정 영어 교과서의 코퍼스 언어학적 비교 분석(A corpus-based analysis of middle school English 1 textbooks with Coh-Metrix). 《영어교육연구》(*English Language Teaching*) 21-4, 265-292.
- 전문기·최윤희(Jeon, M-G and Y. Choe). 2019. 코메트릭스를 활용한 한국 대학생들의 영어 논설문과 설명문 요약문 비교 분석(A Coh-Metrix analysis of Korean EFL learners' summary writings in the English argumentative and expository texts). 《영어학》(*Korean Journal of English Language and Linguistics*) 19-3, 539-559
- 정미경·김남국(Jeong, M-K. and N-G. Kim). 2014. Coh-Metrix를 통한 한국인 EFL 학습자와 영어 모국어 화자의 논설문 코퍼스에 나타난 언어적 특성 분석(An analysis of the linguistic features on the corpus of Korean EFL learners and native English speakers with Coh-Metrix). 《언어학연구》(*Studies in Linguistics*) 33, 373-395.
- 홍성우·송해성(Hong, S. W. and H. S. Song). 2019. 코메트릭스(Coh-Metrix)를 활용한 2015 개정 고등학교 1학년 영어 교과서 읽기 자료의 난이도 분석(Exploring the difficulty levels of the reading materials in high school English textbooks with Coh-Metrix for the 1st year students). 《영어어문교육》(*English Language & Literature Teaching*) 25-2, 133-161.
- Aryadoust, V. and S. Liu. 2015. Predicting EFL writing ability from levels of mental representation measured by Coh-Metrix: A structural equation modeling study. *Assessing Writing* 24, 35-58.
- Bond, T. G. and C. M. Fox. 2001. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. NJ: Lawrence Erlbaum.
- Crossley, S. and D. S. McNamara, 2011. Understanding expert ratings of essay quality: Coh-Metrix analyses of first and second language writing. *International Journal of Continuing Engineering Education and Life-Long Learning* 21, 170-191.
- Crossley, S. and D. S. McNamara. 2012. Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading* 35(2), 115-135.

- Fulcher, G. 2003. *Testing Second Language Speaking*. Harlow: Pearson.
- Graesser, A. C., D. S. McNamara and J. M. Kulikowich. 2011. Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher* 40, 223-234.
- Hyland, K. and E. Anan. 2006. Teachers' perceptions of error: The effects of first language and experience. *System* 34, 509-519.
- Johnson, J. S. and G. S. Lim. 2009. The influence of rater language background on writing performance assessment. *Language Testing* 26, 485-505.
- Kane, M. T. 2013. Validating the interpretations and uses of test scores. *Journal of Educational Measurement* 50, 1-73.
- Khushik, G. A. and A. Huhta. 2019. Investigating syntactic complexity in EFL learners' writing across Common European Framework of Reference. *Applied Linguistics*, 1-28.
- Knoch, U. 2009. Diagnostic assessment of writing: A comparison of two rating scales. *Language Testing* 26, 275-304.
- Lee, H-K. 2009. Native and nonnative rater behavior in grading Korean students' English essays. *Asia Pacific Educational Review* 10, 387-397.
- McNamara, D. S., S. A. Crossely and P. M, McCarthy. 2010. Linguistic features of writing quality. *Written Communication* 27, 57-86.
- McNamara, D. S., A. C. Graesser, P. M. McCarthy and Z. Cai. 2014. *Automated Evaluation of Text Discourse with Coh-Metrix*. UK: Cambridge University Press.
- McNamara, T. 1996. *Measuring Second Language Performance*. NY: Addison Wesley.
- Schaefer, E. 2008. Rater bias patterns in an EFL writing assessment. *Language Testing* 25, 465-493.
- Shin, D. I. and H. S. Seol. 2005. Understanding rating patterns through new FACETS program. *Foreign Languages Education* 12, 191-212.
- Weigle, S. C. 1998. Using FACETS to model rater training effects. *Language Testing* 15, 263-287.
- Weigle, S. C. 2002. *Assessing Writing*. UK: Cambridge University Press.
- Xu, W. and M. Liu. 2016. Using Coh-Metrix to analyze Chinese ESL learners' writing. *International Journal of Learning, Teaching and Educational Research* 15, 16-26.
- Yi, J. Y. 2008. Necessity of developing a rating scale for writing assessment in Korean high schools. *Foreign Languages Education* 15, 261-286.

예시 언어(Examples in): 영어(English)

적용가능 언어(Applicable Languages): 한국어(Korean), 영어(English)

적용가능 수준(Applicable Level): 대학생, 성인(Tertiary)

이호 (Lee, Ho), 교수(Professor)

중앙대학교 사범대학 영어교육과

(Dept. of English Education, Chung-Ang University)

서울시 동작구 흑석로 84(84 Heukseok-ro Dongjak-gu, Seoul, Korea)

Email: holee@cau.ac.kr

논문 투고(Received): 2019년 11월 1일(Nov. 1, 2019)

논문 수정(Revised): 2019년 12월 15일(Dec. 15, 2019)

게재 확정(Accepted): 2019년 12월 20일(Dec. 20, 2019)