영어어문교육 *제20권* 1호 2014년 봄

다국면 Rasch 측정의 Facets 프로그램을 활용한 영어 작문 평가의 원어민 채점 검증*

이 영 식
(한남대학교)

Lee, Young Shik. (2014). An investigation into the native raters' scoring of English writing assessment using the new Facets of many-facet Rasch measurement. *English Language & Literature Teaching, 20*(1), 475-496.

The purpose of this paper is to investigate the reliability of English writing test scoring using the computer program of Facets. Based on the many-facet Rasch measurement (MFRM) model, the Facets program has recently been updated and this research uses it in an attempt to analyze the scoring reliability of native raters for an EFL writing test. For the research data the EFL compositions written by 114 English major students were collected and rated by three native raters using a three-category rating scale (Content, Rhetorical Category and Language). At first a brief theoretical basis for the Facets is introduced based on the MFRM model. Then the scores of EFL compositions are analyzed using the Facets and its analysis results are interpreted. Many-facet Rasch analysis revealed that the three native raters are at the same level in terms of rating severity. Although a few cases of unreliable scoring of native raters are found in terms of rater measurement report, it appears that the ratings of EFL compositions by the native raters are generally sound. Finally it is suggested that the Facets program of MFRM can offer many potential advantages for investigating scoring reliability and ensuring construct validation of EFL writing assessment.

[Facets/many-facet Rasch measurement/writing assessment/raters, 영어 작문 평가/채점/다국면 Rasch 측정]

I. 서론

현대 언어평가의 주된 경향은 수험자의 의사소통능력(communicative language ability)을 타당하게 측정하기 위하여 직접적으로 언어 표현력을 평가하는 것으로, 수험자의 수행에 의거한 언어평가(performance-based language testing)이다.

* 본 논문은 2013년도 한남대학교 학술연구조성비 지원에 의하여 연구되었음

교보문고 KYOBO Book Centre

언어 표현력에 의한 수행평가는 종전에 중시하던 언어적 능력(linguistic competence)보다는 의사소통 수행(communicative performance)을 다루며 간접 측정보다는 직접 측정이기 때문에 종래의 객관식 선다형 문항 위주의 언어 평가와는 여러모로 차이가 있으며 평가의 타당성을 제고할 수 있다. 특히 언어 표현력 수행평가는 시험 상황을 실제에 가깝고 의미 있게 함으로써 시험에서의 수행(test performance)을 실제 언어사용(non-test language use)과 일치시키려고 한다(Bachman, 1991).

아직도 우리는 수험자의 바람직한 의사소통능력을 제대로 평가하는 수행 위주의 언어평가에 관여하기보다는 시험의 관리 및 채점이 용이하고 객관적 채점이라는 이유로 주로 다지선다형 또는 단답식 문항위주의 영어 시험을 출제하여 왔던 것이 우리의 현실이다. 특히 외국어로서의 영어(English as a foreign language: EFL) 작문시험과 같은 직접 평가가 채점자의 주관적인 판단을 요구하므로 채점자들끼리 채점이 일치하지 않거나 채점자 개인의 주관적인 판단의 소지가 많아서 영어 작문시험이 수험자의 영어 의사소통능력을 객관적으로 공정하게 평가할 수 없는 결과를 초래할 수 있다. 특히 한국인 대학생이 쓴 영어 작문에 대해 원어민 채점자가 채점할 때에도 주어진 영어 작문을 채점할 수 있는 영어 능력이 부족하기 보다는 그들의 채점 주관성으로 인하여 영어 작문 평가에 대한 채점 결과가 일치하지 않거나 또는 주어진 채점표에 대하여 서로 다른 해석을 하는 문제가 발생한다고 볼 수 있다. 또한 영어 작문 평가에서 관찰될 수 있는 채점 비신뢰성의 문제는 모든 서술형 주관식 평가에서 공통적으로 관찰될 수 있는 채점자 엄정성의 차이나 채점의 일관성 부족 및 기타 채점 오류에서 기인되는 문제라고 할 수 있다. 이러한 여건 속에서 원어민들에 의한 영어 교실 수업에서 올바른 영어 작문 시험이 잘 정착되기 위해서는 영어 작문의 공정하고 타당한 채점이 보장되어야 한다. 이에 따라 원어민에 의한 영어 작문 채점 신뢰성도 검증할 필요가 있으며, 그러한 채점 신뢰도의 검증을 위하여 효율적인 연구 분석 방법을 모색할 필요가 있다.

II. 최근 개발된 다국면적 Rasch 측정 모형

이미 오래 전부터 다국면 Rasch측정(many-facet Rasch measurement: MFRM)의 이론적 바탕이 확립되고 그에 따른 정교한 컴퓨터 측정 프로그램이 개발되어서, 영어 작문 평가뿐만 아니라 다양한 언어수행평가에 대한 채점자들의 채점결과를 검증할 수 있는 도구로써 사용될 수 있다. 영어 작문 평가와 같은 언어 수행평가에 수반되는 채점자의 주관성, 채점자 간의 불일치 및 기타 채점 오류를 규명하기 위하여 채점자의 채점결과를 다국면 Rasch측정 모형을 이용하여 분석하여 봄으로써 언어수행평가의 채점자 오류를 줄이고 보다 타당한

교보문고 KYOBO Book Centre

채점을 할 수 있는 분석 자료를 제공할 수 있다. 특히 영어 작문 평가와 같은 표현력 시험 결과로부터 수험자의 작문 능력을 신뢰하고 타당하게 추론해낼 수 있으려면 그 무엇보다도 채점 행위의 질이 확보되어야 한다.

신뢰도가 높은 수준의 채점 행위는 언어 표현력 수행평가 방식의 타당도에 가장 중요한 요인이 될 수 있다. 이와 같은 취지로 Linacre(1989; 1994; 1996)는 덴마크 수학자 Georg Rasch(1960; 1980)의 일모수 문항반응이론(one-parameter item response theory)에 의거하여 언어표현력 평가 채점 뿐만 아니라 모든 주관적인 평가 결과 분석에 적용하는 다국면 Rasch측정 모형을 확립하였고, 그러한 평가 결과를 컴퓨터로 분석할 수 있는 Facets 컴퓨터 프로그램을 오래 전부터 개발하였다. 흔히 2모수 및 3모수의 문항반응이론에 의한 분석도구는 대부분 정답과 오답으로 구분되는 단순한 이분형(dichotomous) 자료를 분석하고 있으나, 반면에 다국면적 Rasch측정 모형에 의거한 Facets는 여러 가지 부분 점수를 주는 다분형(polytomous) 자료와 채점자의 엄정성이나 문제의 유형 등의 여러 가지 국면을 분석하는 데 적합하도록 변형하였다(Linacre, 1996; McNamara, 1996; 이영식, 1998).

다국면 Rasch측정 모형의 이론적 근거와 배경을 다음과 같이 간단히 설명할 수 있다.

$$\log(Pnmijk / Pnmijk\text{-}1) = Bn - Am - Di - Cj - Fk$$

위 모형에서 어떤 수험자 n이 문항 i에서 점수 Xnijk-1을 얻을 확률 분의 점수 Xnijk를 얻을 확률의 log 값은 그 수험자 n의 능력(Bn)에서 시험과제 m의 수준(Am)과 문항 i의 난이도(Di)를 뺀 후, 여기에 채점자의 엄정한 정도(Cj)와, 그리고 어떤 부분 점수(k-1)에서 그 다음으로 높은 부분 점수(k)로 올라갈 수 있는 데 걸리는 어려움(Fk)을 뺀 것이다. 이러한 Rasch분석 모형에서는 수험자의 능력, 문항의 난이도, 채점자의 엄정성 등 여러 가지 요인들이 국면으로 설정되어 다국면적으로 파악될 수 있으며, 이러한 요인들이 채점자의 능력이나 문항의 난이도에 미치는 영향을 분석하여 통제할 수도 있다.

특히 Facets 프로그램은 언어 표현력(말하기 및 쓰기), 포트폴리오 평가 및 기타 실기 시험 등 채점자들의 판단을 요구하는 주관적 평가의 결과를 분석하는 데 유용하다. 이미 오래전부터 많은 언어평가 전문가들이 Facets 프로그램 구 버전을 사용하여 언어수행평가 분석에 대한 연구발표를 하였다. 해외에서는 Weigle은 미국 대학생 영어작문 평가 분석(1994)과 채점자 훈련 효과 연구(1998)에 Facets를 사용하였으며, Lumley와 McNamara(1995)는 채점자의 특성과 편견(bias)을 Facets로 분석하여 채점자 훈련에 대한 시사점을 제시하였고, Brown(2003)은 말하기 시험의 면접관(interviewer)의 변인을 규명하도록 노력하였다. 특히 Lynch와 McNamara(1998)는 제2언어로서의 영어(ESL) 말하기 시험 개발에 일반화가능도 이론(G-theory)과 다국면 Rasch측정 이론을 사용하였다. 국내에서는 이영식 교수(1998)가 외국어로서의 영어 작문 평가에

교보문고 KYOBO Book Centre

대한 채점신뢰도를 분석하였다. 또한 Facets 프로그램을 이용하여 언어능력 평가에 대한 안목을 제시한 책(McNamara, 1996)도 해외에서 출판되었다.

1990년대에 개발된 Facets (Version 3.0)는 DOS에 의한 프로그램(Linacre, 1989; 1996)이어서 그 프로그램 사용자가 이용하기에는 정교하지 못하고 불편한 점이 많았다. 2000년대 중반부터 종래의 프로그램을 보완하여 최근에는 Windows용 프로그램 Facets (Version 3.67.0)를 내놓았다(Linacre, 2010; 2011). 또한 이렇게 보완된 신 버전 Facets 프로그램을 이용하여 언어수행평가를 분석한 연구 발표와 논문들도 나왔다(신동일과 설현수, 2005; Eckes, 2005; Elder 등, 2007; Davis, 2009; Kecker & Eckes, 2010; Kozaki, 2010; Barkaoui, 2011; Wu & Wu, 2012). 한편 신유선은 영어학습자(212명)와 교수자(52명)의 영어작문의 평가 기준에 대한 인식에 대해 요인 분석(factor analysis)을 이용한 연구(Shin, 2011)를 하였으나, 영어작문 채점자 특성과 편견(bias)에 대해 Facets를 이용하여 분석하였다(Shin, 2010).

최근 새로운 버전의 Facets 프로그램은 웹사이트에서 유료로 구입하여 다량의 주관적 판단 채점 데이터를 분석할 수 있다(Linacre, 2010). 한편 무료로 Facets 프로그램을 관련 웹사이트(www.winsteps.com)에서 다운 받아 사용할 수 있었는데, 최근(2012년)에 업데이트된 무료 프로그램(Minifac Version 3.70.1)으로 일반인들도 2000개 이하의 데이터를 무난히 분석할 수 있게 되었다. 또한 Facets 프로그램 사용자들이 참고할 수 있도록 매뉴얼(Linacre, 2011)을 웹사이트에서 제공하고 있다. 그러나 그러한 Facets 프로그램에 대한 매뉴얼 설명이 언어평가에만 국한되지 않고 의학 및 기타 사회과학의 데이터 분석에 적용되는 설명도 제공되므로, 실제로 언어교육에 관여한 연구자들이 그 매뉴얼을 바로 이해하기에는 어려운 점이 있으리라고 예상된다.

Facets 프로그램을 이용하여 채점 행위의 분석에 사용되는 주요 지수는 부적합(misfit)이나 과적합(overfit)과 같은 적합도 지수(misfit statistics)와 분리신뢰도 지수(reliability for separation)이다. 또한 Facets 프로그램에 의하여 실행되는 다국면적 분석의 중요성은 종합적인 요약도표(summary map) 뿐만 아니라 수험자의 능력(ability), 시험문제(task) 또는 채점항목(item)의 난이도, 채점자(rater)의 엄격하거나 관대함과 같은 각각의 국면에 대한 상세한 정보를 여러 도표를 통하여 제공하여 준다(Linacre, 1996; McNamara 1996; 이영식 1998). 물론 국내에서 장소영과 신동일(2009)이 Windows용 Facets 프로그램에 대해 소개한 책을 내놓았으나, 실제로 언어평가 연구자가 분석한 자료를 정확히 해석하기 위해서는 그 책뿐만 아니라 웹사이트에 있는 매뉴얼(Linacre, 2011)을 참조하여야 한다. 그러나 앞에서 언급한 바와 같이 그 매뉴얼을 잘 이해하여 사용하기에는 어려운 점이 있으므로, 언어평가 전문가가 실제 우리 현장의 언어표현력 채점결과를 수집하고 분석하면서 웹사이트에 있는 매뉴얼의 내용을 우리 언어평가 상황에 맞추어 해석하고 정리하여 제시할 필요가 있다.

이러한 취지에서 영어수행평가에 대한 주관적인 채점을 실시하여, 그 채점

교보문고 KYOBO Book Centre

결과를 무료로 제공되는 최신의 Facets 프로그램을 이용하여 분석 자료를 산출하고, 그 분석 결과를 웹사이트에 주어진 매뉴얼을 참고하고 정리하여 우리 영어 수행평가에 적절한 논의를 제시하는 것도 좋은 연구가 될 수 있다고 본다. 따라서 본 연구에서는 Rasch측정 모형에 대한 이론적 근거나 기술적인 문제에 대한 자세한 논의보다는, 영어 모의고사를 통하여 수집된 대학생 영어 작문에 대해 원어민 채점자가 채점한 결과를 우선 최근에 개발 보완된 Windows용 프로그램 Facets (Linacre, 2010)으로 분석 자료를 산출하고 그 매뉴얼(Linacre, 2011)을 이용하여 영어 쓰기 시험의 타당도 검증에 대해 논의하고 정리하면서 향후 영어 표현력의 타당도 검증 체제를 확립하는데 기여하고자 한다.

III. 연구방법

영어 작문 채점 신뢰도를 분석하기 위하여 한 국내 대학에서 실시된 영어 쓰기 시험에서 쓰여진 영어 작문을 수집하고 그 영어 작문에 대한 채점을 실시한 결과를 토대로 연구를 진행하였다. 즉, 외국어로서의 영어 작문 채점 신뢰도를 진단하고 분석하기 위하여 실제로 한국 대학생들에 의하여 쓰여진 영어 작문을 수집하였을 뿐만 아니라, 기존에 개발되어 해외에서 사용되는 영어 작문 평가기준(채점표)을 이용하여 학생들을 직접 가르친 원어민 3명이 채점한 결과를 분석하였다.

1. 수험자

이번 영어 작문 시험을 치른 수험자는 한 국내 대학에서 영어를 전공하는 학생들로 1학년 38명, 2학년 37명, 3학년 39명으로 총 114명이었다.

2. 시험문제

> **※ Write a composition on the following topic in 120-180 words**
> Unemployment, especially among young people, is a serious problem nowadays. What can be done about it? Do you think traditional ideas about work should be changed?

외국어로서의 영어 쓰기 능력 평가를 위하여 위와 같은 영어 논술문을 쓰게 하였다. 이 시험문제로 이미 출판된 FCE 시험의 작문 제목을 이용하였다 (UCLES, 1988).

교보문고 KYOBO Book Centre

3. 채점자 및 채점

국내 한 대학에서 영어를 가르치고 있는 원어민 영어 교수자 3명을 동원하여 각각 총 114개의 영어 작문 채점을 실시하였다. 채점자 2명은 해당 학과에서 영어를 가르치고 있으며, 나머지 채점자 1명은 그 대학에서 영어를 가르치나 이번에 시험을 치른 학과의 학생들은 가르치지 않는다. 원어민 채점자 모두 영어교육으로 미국에서 석사 학위를 취득하였다.

4. 채점표 및 채점 방식

채점기준으로 미국 UCLA에서 외국인 유학생들이 입학 전 영어교육을 받고 영어 작문시험을 치르는데 사용된 채점표 English as a Second Language Placement Examination (ESLPE) Rating Scale을 사용하였다. 이 채점표는 작문의 내용(Content), 수사적 통제(즉, 작문의 조직: Rhetorical Control), 문법 및 어휘 등의 언어(Language)의 3영역으로 나누고 각각 10점씩 부여하고 있으나, 실제로는 언어 영역은 2배를 하여 20점으로 채점하므로 총점 40점 만점이 되게 하였다(Carr, 2011, pp. 129-131).

IV. 분석결과

1. 모든 국면의 분석 결과(All Facet Vertical "Rulers")

표 1에서 보는 바와 같이, 수집된 영어 작문을 3명의 원어민 채점자에 의하여 채점한 결과를 다국면적 Rasch 측정 모형에 의거한 컴퓨터 프로그램 Facets를 이용하여 분석한 결과를 요약도표(map)로 제공한다. 이 표에서는 8개의 칸(column)이 있는데, 첫 번째 칸에서는 측정척도(scale of measurement)를 표시하여 주고, 나머지 칸에서는 각 국면(facets)에 대한 분석결과 즉, 수험자(영어 작문), 학년, 채점자, 채점영역(Content, Rhetorical Control, Language), 채점표에 대한 분석결과를 제공하여 주고 있다.

1) 첫째 칸(Measr)에서는 단위가 'logit'로서 기준점이 "0"으로 되어있고, 위로는 '+1', 아래로는 '-1'로 진행되고 더 나아가서는 각각 양.음의 무한대 ($+\infty$,$-\infty$)까지도 진행될 수 있다.

교보문고 KYOBO Book Centre

표 1

채점표 채점기준에 의한 다국면적 분석 결과 요약

Measr |+Stu	|-Gr	|-Rtr	|-Items	| S.1| S.2 | S.3
11 +	+	+	+	+(10) +(10) +(20)
10 + *****	+	+	+	+ 9 + 9 + 18
9 + ****	+	+	+	+ + +
8 + ***	+	+	+	+ — + — + —
7 + *********	+	+	+	+ 8 + 8 + 16
6 + ****	+	+	+	+ — + — + —
5 + *****	+	+	+	+ 7 + 7 + 14
4 + ********	+	+	+	+ — + — + —
3 + *******	+	+	+	+ 6 + 6 + 12
2 + *****	+	+	+	+ — + — + 10
1 + ********	+ 1	+	+	+ 5 + 5 + —
* 0 * ********	* 2 3	* 1 2 3 *	Con LU RC	* — * — * *
-1 + ********	+	+	+	+ 4 + 4 + 8
-2 + ****	+	+	+	+ — + — + —
-3 + ***	+	+	+	+ + + 6
-4 + ****	+	+	+	+ 3 + 3 +
-5 + ****	+	+	+	+ — + — + —
-6 + *****	+	+	+	+ + +
-7 + *	+	+	+	+ 2 + + 4
-8 + **	+	+	+	+ + 2 +
-9 + ***	+	+	+	+ + +
-10 + ***	+	+	+	+ — + — + —
-11 + **	+	+	+	+ + +
-12 + *	+	+	+	+ 1 + 1 +
-13 + ****	+	+	+	+ + +
-14 +	+	+	+	+ + — + 2
-15 + *	+	+	+	+ — + +
-16 + **	+	+	+	+ + +
-17 + *	+	+	+	+ (0) + (0) + (0)
Measr | * = 1	|-Gr	|-Rtr	|-Items	| S.1 | S.2 | S.3

2) 둘째 칸(Stu)에서는 각 영어 작문에 대한 채점결과(점수)를 다루고 있다. 실제 점수 표를 보니 총 114명 중 좋은 점수를 받은 학생이 위에 있으며 낮은 점수를 받은 학생이 아래에 위치하고 있다(둘째 칸의 상단과 하단에는 Stu (학생) 앞에 '+'로 표기되어 있어서 우수한 학생이 위에 위치하고 우수하지 못한 학생은 아래에 위치한다). 표 1에서 보건대 대체로 많은 학생들이 0 logit 이상과 이하에서 고루 분포되어 있으나, 0 logit 이상에서 약간 많음을 알 수 있다

3) 셋째 칸(Gr)에서는 3개 학년의 수준이 나열되어 있는데, 2학년과 3학년 모두 0 logit에 위치하고 있으며, 1학년은 약간 위에 위치하고 있다. 2학년과

교보문고 KYOBO Book Centre

3학년은 동일 수준이나, 1학년이 2-3학년에 비해 수준이 약간 낮다는 것을 알 수 있다(셋째 칸의 상단과 하단에는 Gr(학년) 앞에 '-'로 표기되어 있어서 우수한 수준이 아래에 위치하는데, 그 수준을 반대로 위치하도록 프로그램을 짤 수도 있다).

4) 넷째 칸(Rtr)에서는 채점자에 대한 채점결과를 다루고 있는데, 3명의 원어민 채점자가 모두 같은 수준에 위치하고 있어서 채점자 간의 수준 차이는 없는 것으로 판명되었다(여기에서도 상단과 하단에는 Rtr(채점자) 앞에 '-'로 표기되어 있어서 엄격한 채점자가 위에 위치하고 관대한 채점자가가 아래에 위치한다).

5) 다섯째 칸(Items)에서는 채점영역에 대한 분석 결과를 다루고 있는데, 채점 기준표의 'Content', 'Language' 및 'Rhetorical Control'의 3개 영역 모두 동등한 난이도를 유지하였다고 볼 수 있다(상단과 하단에는 Items(영역) 앞에도 '-'로 표기되어 있어서 어려운 영역이 위에 위치하고 쉬운 영역이 아래에 위치한다)

6) 여섯째 칸(S.1)에서는 'Content' 영역에 사용된 채점 점수간의 간격을 보여주고 있다. 총 10점 만점 중 0점과 3점 이하의 간격은 넓은데, 4점 이상과 9점 이하의 간격은 좁게 나타났다. 이러한 결과는 총 114명 학생의 영어 작문의 'Content' 영역에 대해 0점과 3점의 점수 간격 보다는 4점과 9점의 간격을 적게 파악하는 채점을 하여서 채점표의 각 점수간의 간격이 잘 조정되어 있지 않거나 또는 채점자들이 채점표의 각 등급 점수의 사이를 등 간격으로 판단하지 않고 점수를 부여한 결과라고 볼 수 있다.

7) 일곱째 칸(S.2)에서는 'Rhetorical Control' 영역에 사용된 채점 점수간의 간격을 보여주고 있다. 이 영역의 채점도 0점과 3점 이하의 간격은 넓은데, 4점 이상과 9점 이하의 간격은 좁게 나타나서, 이 영역도 각 점수간의 간격이 잘 조정되어 있지 않거나 또는 채점자들이 채점표의 각 등급 점수의 사이를 등 간격으로 판단하지 않고 점수를 부여한 결과라고 볼 수 있다.

8) 여덟째 칸(S.3)에서는 'Language' 영역에 사용된 채점 점수간의 간격을 보여주고 있다. 총 20점 만점 중 0점과 6점 이하의 간격은 넓은데, 10점 이상과 18점 이하의 간격은 좁게 나타났다. 이러한 결과는 총 114명 학생의 영어 작문 'Lanaugage' 영역에 0점과 8점의 점수 간격 보다는 10점과 16점의 간격을 적게 파악하는 채점을 하였다고 볼 수 있어서 채점표의 각 점수간의 간격이 잘 조정되어 있지 않거나 또는 채점자들이 채점표의 각 등급 점수의 사이를 등 간격으로 판단하지 않고 점수를 부여한 결과라고 볼 수 있다. 한편 각 세 영역의 점수간격 모두 거의 같은 수준을 유지하고 있는데, 다섯 째

교보문고 KYOBO Book Centre

칸(Items)에서 세 영역의 수준이 0 logit에서 동등하게 위치하는 것과 상통한다고 볼 수 있다.

2. 채점의 부적합(misfit) 및 과적합(overfit)

Facets 프로그램은 각 국면에 대한 측정치(measurement report)를 보여주고 있어서 수험자(작문), 채점자, 과제, 채점영역 국면에 대한 많은 도표를 제공하고 있는데, 여기에서는 본 연구의 초점인 채점자 국면만 살펴보기로 한다.

표 2

채점자 엄정성 및 적합도 측정치

Total Score	Total Count	Obsvd Average	Fair-M Avrage	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	N	Rtr
2121	342	6.20	5.43	.38	.09	.79	-2.6	.81	-2.2	1.09	.95	.95	1	1
2196	342	6.42	5.66	-.07	.09	.71	-3.8	.76	-3.0	.90	.96	.95	2	2
2271	342	6.64	5.79	-.31	.09	1.23	2.6	1.18	2.0	.35	.95	.95	3	3
2196.0	342.0	6.42	5.63	.00	.09	.91	-1.3	.92	-1.1		.95		Mean (Count: 3)	
61.2	.0	.18	.15	.29	.00	.23	2.9	.19	2.2		.01		S.D. (Population)	
75.0	.0	.22	.19	.35	.00	.28	3.5	.23	2.7		.01		S.D. (Sample)	

Model, Populn: RMSE .09 Adj (True) S.D. .27 Separation 3.11 Strata 4.48 Reliability .91
Model, Sample: RMSE .09 Adj (True) S.D. .34 Separation 3.88 Strata 5.50 Reliability .94
Model, Fixed (all same) chi-square: 32.0 d.f.: 2 significance (probability): .00
Model, Random (normal) chi-square: 1.9 d.f.: 1 significance (probability): .17

어떤 채점자의 '내적합 표준화값'(Infit ZStd) 및 '외적합 표준화값'(Outfit ZStd)이 '2' 이상이 되면 그 채점자에 의한 채점은 부적합(misfit)한 것으로 판명될 수 있고, '-2' 이하이면 그 채점은 과적합(overfit)하다고 볼 수 있다(또는 표준화값 대신에 '내적합 지수'(Infit MnSq) 및 '외적합 지수'(Outfit MnSq)를 참고하여 1.5이상이면 부적합, 0.5이하면 과적합한 것으로 보기도 한다). 부적합한 채점은 채점이 부정확하고 일관성이 없어서 채점의 신뢰도가 떨어지는 것을 의미한다. 반면에 과적합한 채점은 채점표의 높은 점수와 낮은 점수를 피하여 중간 점수로 몰아주는 채점이라고 할 수 있다. 이러한 과적합한 채점은 극단 점수를 피하다 보니 일반적 통계 기법에 의하여 다른 채점자와 채점의 신뢰도(inter-rater reliability) 상관관계 지수는 높게 나올 수 있으나, 이러한 과적합한 채점을 발견하여 주는 것이 다국면 Rasch측정의 장점이라 할 수 있다.

표 2에서 보는 바와 같이, 3명의 채점자들에 의한 채점결과를 분석하여 보면, 채점자 1번과 2번은 과적합한 것으로, 반면에 채점자 3번은 부적합한 것으로

교보문고 KYOBO Book Centre

나타났다. 채점자 1번과 2번은 점수를 각각 중앙으로 몰아 준 경향이 있고, 채점자 3번은 다소 일관성 없는 채점을 한 것으로 판명되어, 3명의 원어민 채점자 사이에 — 특히 채점자 1번 또는 2번과 채점자 3번 사이에 — 채점 엄격성의 차이는 많다고 볼 수 있다.

또한 Facets의 각 국면의 산출 결과 도표 하단에 분리(separation) 지수, 층(strata)의 지수, 신뢰도(reliability) 지수를 제공한다. 그 중에서 'Model, Populn' 이라고 표기된 줄에서는 현재 채점결과에 해당되는 집단(즉, 여기에서는 114명 및 3명의 채점자 등)을 전체 모집단(population)으로 보고 산출된 수치를 제공하고, 반면에 그 밑의 'Model, Sample'이라고 표기된 줄에서는 현재 채점결과의 집단을 하나의 샘플집단(sample)으로 보고 전체 모집단을 추정한 수치를 제공한다. 따라서 현재 채점결과에 해당되는 수치를 보려면 'Model, Populn' 줄을 살펴 보아야 하고, 현재 채점결과를 넘어서 전체 모집단으로 추정되는 수치를 보려면 'Model, Sample' 줄을 살펴 볼 필요가 있다. 현재 'Model, Populn'의 신뢰도 지수(reliability) 지수는 .91로 나와서 채점자들의 엄격성에는 많은 차이가 있는 채점을 하였고, 분리(separation) 지수 3.11에 의거 채점자 층(strata)의 지수가 4.48이어서 채점자 간에는 약 5개의 층으로 차이가 있는 채점결과라고 할 수 있다.

반면에 이러한 신뢰도 및 분리 지수는 수험자(학생 또는 영어작문)의 국면에서 볼 때 높은 신뢰도 지수는 변별력(즉, 능력 차이)을 나타내므로 바람직하다고 말하는데, 본 연구에서 학생(Stu) 국면의 신뢰도 지수는 .99로 나와서 이번 영작문 시험은 학생들간에 높은 변별력을 지녔다고 볼 수 있다(학생 국면에 관한 적합도 및 신뢰도 지수에 관한 도표는 지면관계상 수록하지 않았음). 그러나 채점자 국면에서는 높은 신뢰도 지수는 일단 채점 엄격성에 많은 차이가 있는 채점을 의미하므로 채점자 국면에서의 높은 신뢰도 지수는 바람직하지 않다. 분리신뢰도가 높을수록 채점자들의 엄정성이 체계적으로 차이가 있는 것을 의미하며, 반면에 (1-분리신뢰도) 값은 '채점자 간 일치도' 또는 '채점자 내 신뢰도' 지수를 나타낸다. 따라서 표에서 보는 바와 같이 이렇게 높은 신뢰도 지수는 바람직하지 않으며 채점자의 엄격성 차이의 원인을 규명하는 노력을 하여야 한다는 것을 의미한다(예를 들면, 채점표 점수의 최저 점수와 최고 점수 사이의 넓은 폭 또는 채점표에 대한 각 채점자 간의 판단 차이 등).

3. 사용된 단계의 적합성(Category Statistics)

Facets 프로그램을 이용하여 분석한 결과는 채점자에 의해 사용된 평가 척도에 대한 기술통계와 채점된 각 점수들의 변별력에 대한 전체적인 인상을 제공하여 주어서 영어 작문 평가의 채점결과를 확인할 수 있다.

교보문고 KYOBO Book Centre

표 3

평정척도의 기술 통계 (Category Statistics)

DATA				QUALITY CONTROL		
Category	Counts	%	Cum %	Avge Meas	Exp Meas	Outfit MnSq
0	11	3	3	-16.36	-15.81	.3
1	33	10	13	-11.34	-11.56	.8
2	36	11	23	-7.39	-7.17	.9
3	33	10	33	-3.37	-3.47	1.0
4	52	15	48	-1.00	-.90	.8
5	39	11	60	1.06	1.14	.7
6	32	10	70	3.27	3.18	.6
7	43	13	82	5.18	5.07	.9
8	30	9	91	7.12	7.15	.9
9	24	7	98	8.89	8.85	.8
10	6	2	100	10.02	9.82	.8

표 3에서 보는 바와 같이 'Content' 영역에 대한 평정 척도의 기술통계를 제공하여 주는데, 각 점수 0점에서 10점까지의 빈도(Counts), 비율(%)과 누적 비율(%), 평균 측정값(Average Measure), 기대측정값(Expected Measure) 등을 알려준다. 여기에서 가장 많이 받은 점수는 4점으로 52개(15%)임을 알 수 있으며, 마지막 칸에서 0점 단계를 제외한 모든 점수의 외적합 지수가 .5에서 1.5사이에 있어서 부적합이나 과적합한 척도는 발견되지 않았으나, 0점 단계의 외적합 지수(.3)는 .5보다 적어서 과적합 채점 결과라 할 수 있다. 또한 첫째 칸(Category)의 점수가 0점에서 10점으로 올라 갈수록 다섯째 칸(Average Measure)의 수치도 같이 증가하기 때문에 'Content' 영역의 양호한 채점 결과라 할 수 있다. 만약 증가하는 어떤 특정 영역 등급 점수의 수치가 바로 위의 수치보다 적어서 별표(*)가 표시되면 그 등급 점수에 대한 문제 있는 채점 결과라고 볼 수 있는데, 실제로 신뢰롭지 못한 채점 점수 분석에서는 간혹 별표(*)가 나타나기도 한다.

Facets 프로그램을 이용하여 분석한 결과는 채점 영역에서 부여된 각 점수들의 변별력에 대한 두 개의 그림을 제공하여 주어 해당 채점자들이나 평가 집단에게 전체적인 인상을 파악할 수 있게 한다. 아래 두 개의 그림에서 보는 바와 같이 그림 1은 채점기준표의 각 점수간의 사용된 단계에 대한 확률곡선(Probability Curve)과, 그림 2에서 보는 것처럼 기대점수 오자이브 곡선(Expected Score Ogive Curve)을 그려준다. 종래의 DOS용 Facets 프로그램(Linacre, 1989; 1996)에서는 그림 1과 같은 확률곡선만을 보여 주었는데, 최근 개발된 Windows용 Facets 프로그램(Linacre, 2010)과 매뉴얼(Linacre, 2011)에서는 그림 1뿐만 아니라 그림 2와 같은 오자이브 곡선도 보여주고 있다.

교보문고 KYOBO Book Centre

그림 1

'Content' 채점 영역의 확률 곡선(Probability Curves)

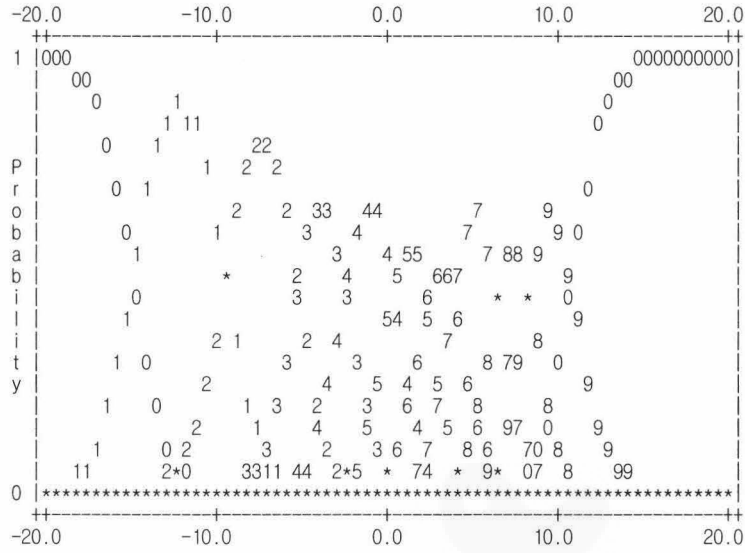
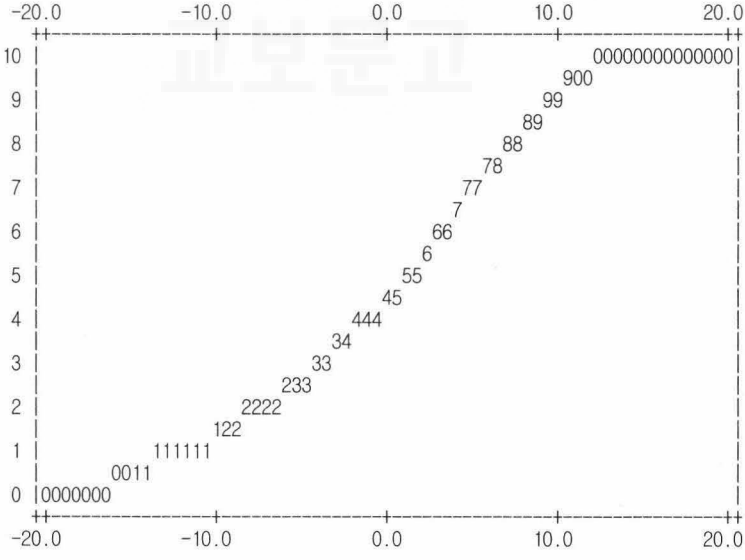


그림 2

'Content' 채점 영역의 오자이브 곡선(Expected Score Ogive: Model ICC)



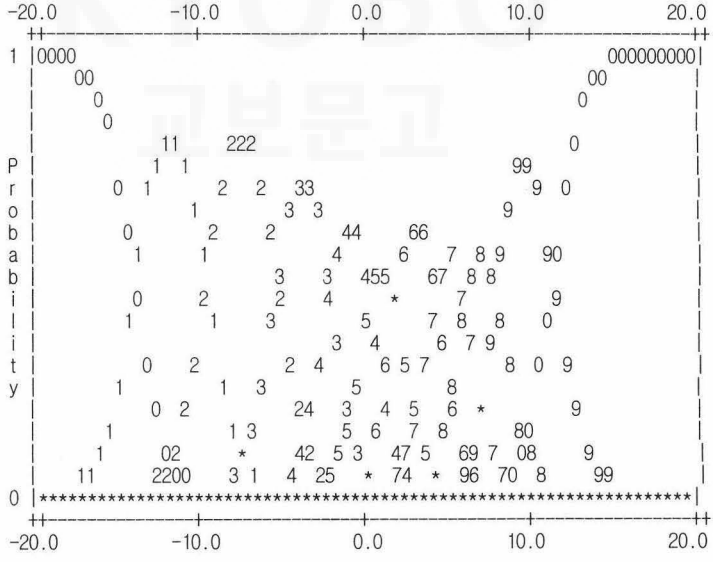
교보문고 KYOBO Book Centre

그림 1에서는 'Content' 영역 척도의 점수에서 특히 5점, 6점, 8점은 낮게 나타나서 다른 점수와 비교하여 상대적으로 변별력이 약한 것으로 나타났다. 따라서 채점표의 'Content' 영역의 등급에서 5점, 6점, 8점의 변별력을 높이도록 채점표를 보완하든지 또는 채점자 훈련을 실시하여야 한다는 것을 시사하고 있다. 그림 2에서는 기대점수 오자이브 곡선을 그려주어 각 점수의 누적 단계를 파악할 수 있게 해준다. 표 3에서 척도 점수가 증가 할수록 평균측정값 역시 증가하는 것처럼, 그림 2에서도 측정치가 오른쪽으로 증가할수록 척도등급이 증가하고 있는 것을 쉽게 파악할 수 있다. 이렇게 아래로 내려가는 굴곡이 없는 척도등급의 증가는 비교적 무난한 채점의 결과로 볼 수 있다.

아래 두 개의 그림 중 그림 3에서 'Rhetorical Control' 영역의 4점에서 8점까지의 점수가 변별력이 낮은 것으로 나타났다. 따라서 채점표의 'Rhetorical Control'영역의 등급에서도 앞에서 말한 바와 같이 해당 점수의 변별력을 높이도록 채점표를 보완하든지 또는 채점자 훈련을 실시하여야 한다는 것을 의미한다고 볼 수 있다. 또한 그림 4와 같이 채점표의 'Rhetorical Control' 영역에 대한 기대점수 오자이브 곡선(Expected Score Ogive Curve)에서도 측정치가 오른쪽으로 증가할수록 척도등급이 증가하고 있는 것을 알 수 있다.

그림 3

'Rhetorical Control' 채점 영역의 확률 곡선(Probability Curves)



교보문고 KYOBO Book Centre

그림 4

'Rhetorical Control' 채점 영역의 오자이브 곡선(Expected Score Ogive: Model ICC)

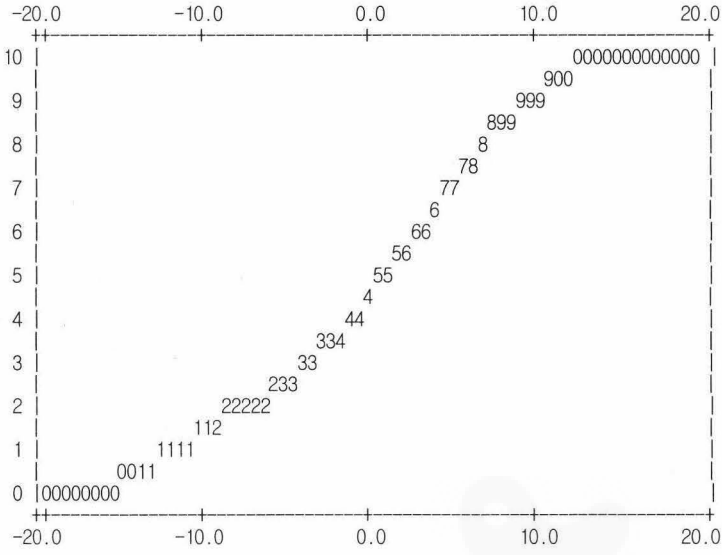
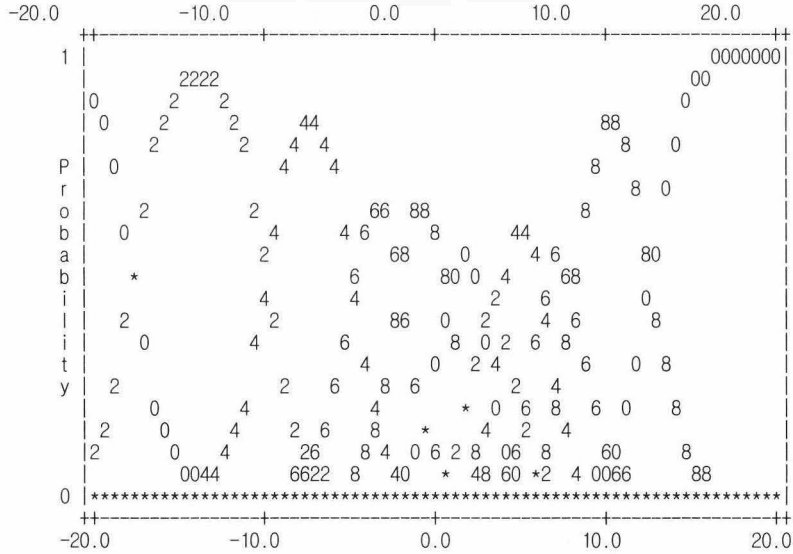


그림 5의 'Language' 영역은 본래 10점의 등급에서 두 배를 하여 20점 만점인데, 6점에서 16점까지의 점수가 변별력이 낮은 것으로 나타났고, 그림 6에서 보는 바와 같이 측정치가 증가할수록 척도등급이 증가하고 있다.

그림 5

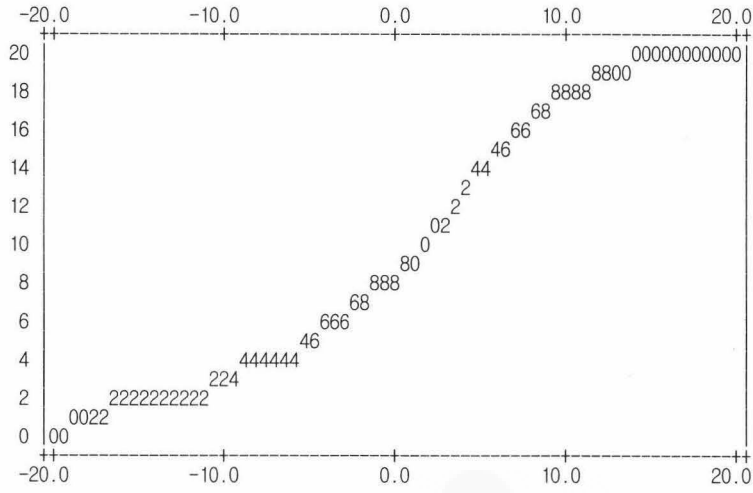
'Language' 채점 영역의 확률 곡선(Probability Curves)



교보문고 KYOBO Book Centre

그림 6

'Language' 채점 영역의 오자이브 곡선(Expected Score Ogive: Model ICC)



4. 채점결과 비신뢰성의 검증(Unexpected Responses)

수집된 채점결과를 입력하여 Facets 프로그램을 이용하면 표 4와 같은 분석모형에서 벗어나는 결과가 나오게 되는데, 이것은 특정 채점자의 비정상적이고 오류가 있는 실제 사례로 파악할 수 있다. 이전의 DOS에 의한 Facets 프로그램에서는 이러한 결과표를 앞에 제시하였는데, 최근 Windows에 의한 Facets 프로그램에서는 이 표를 가장 나중에 제시하고 있으며, 전반적 모형 적합도(global model fit) 통계도 제공한다. 표 4는 표준 잔차(standardized residual) 수치를 보여주고 있는데, 그 절대값이 3이상이면 문제 있는 채점이라고 할 수 있다.

표 4

비 신뢰적인 채점

Score	Exp.	Resd	StRes	Stu	Gr	Rtr	Item
0	1.0	-1.0	-4.8	328	3	1	LU
0	1.1	-1.1	-4.3	328	3	2	LU

* Score 점수 Exp.기대점수 Resd 잔차 StRes 표준잔차 Stu 학생 작문
Gr 학년 Rtr 채점자 Item 채점영역

예를 들어 첫 번째 줄에서 보듯이, 학생 328번의 작문(3학년)의 'Language' 영역에 대해 채점자 1번과 2번이 0점을 주었는데 기대점수는 각각 1.0점과

교보문고 KYOBO Book Centre

1.1점으로 그 차이도 각각 -1.0과 -1.1이다. 표준잔차는 각각 -4.8과 -4.3으로 나타나 절대값이 3이상이어서 문제 있는 채점이라고 밝혀졌다. 이에 따라 학생 328번의 점수를 살펴보면 다음과 같다.

표 5
학생 328번의 영어 작문 점수

채점자	Content	Rhetorical Control	Language	총점
1	1	1	0	2
2	1	1	0	3
3	2	1	2	5

위 표에서 보듯이, 채점자 3번은 학생 작문의 'Language' 영역(20점 만점)의 낮은 점수 2점을 주었으나, 채점자 1번과 2번은 각각 0점을 주어 결국 문제가 있는 채점 결과가 되었다고 볼 수 있다. 표 4에서 학생 328번의 작문의 'Language'에 대해 채점자 1번 및 2번의 표준 잔차 수치가 모두 음수로 -4 이하이어서 과적합(overfit)하게 채점하였다고 밝혀졌다. 만약 표준 잔차의 수치가 양수로 3이상이면 부적합(misfit)으로 판명되는데, 다행히 본 연구에서는 부적합한 채점 사례는 나타나지 않았다. 한편 학생 114명의 작문에 대해 채점자 3명 및 채점 3영역을 고려한다면 총 1,026개의 사례 중 문제 있는 채점은 위 2개만 발견되어 결국 전반적으로는 양호한 채점을 이루었다고 볼 수 있다.

V. 결론

위와 같이 국내 한 대학의 영어 전공 114명을 대상으로 실시한 영어 작문 시험에 대한 원어민 교수 채점자들의 채점 결과를 다국면 Rasch측정 모형에 의한 측정분석 프로그램 Facets를 이용하여 분석하여 보았다. 분석결과로 산출된 통계 및 도표는 많지만 본 논문의 지면관계상 모두 나열하여 해석하지 못한 것이 본 연구의 제한점이라 하겠다. 우선 발견된 문제점은 비록 3인의 원어민의 채점을 실시하였지만, 채점자 국면에 대한 신뢰도 지수가 높아서 채점자마다 채점 엄격성이 서로 차이가 있는 것으로 드러난 결과, 그러한 채점자의 엄격성 차이에 대한 심층적 연구를 후속적으로 하여야 하겠다. 한편 본 연구의 영어 작문시험 채점에 참여한 원어민 채점자의 상호 일치성에 문제가 없는 것은 아니나, 과적합으로 판명된 채점결과가 많지 않아서 영어 작문 평가 채점에 대해 비관적인 견해를 가질 필요는 없다고 본다. 이번에 선발된 채점자들은 영어 교육 경험자임을 고려한다면 향후 우리의 교실 수업을 통한 영어 작문평가를 활성화시키고 채점자 2인 이상의 복수 채점으로

교보문고 KYOBO Book Centre

채점신뢰도가 보장되는 채점 결과를 산출할 수 있는 영어 평가 환경을 진작시킬 수 있다. ''

Davies (1990)에 의하면, 언어평가는 '요구되는 명확성과 내재되어 있는 불확실성의 창조적인 타협(a creative compromise between the required explicitness and the underlying uncertainty)'라고 말하였다. 그러나 우리 영어교육자들은 '요구되는 명확성'은 기존의 표준화된 시험의 평가에만 맡겨서 다지선다형의 이해력 위주의 영어평가에 치우치고, 반면에 영어 말하기 및 쓰기의 표현력과 같은 중요한 의사소통능력의 측정이 불확실하여 심지어 교실 수업을 통한 평가에서도 기피되었던 것도 사실이다. 물론 표현력 평가로서의 영어 작문 시험은 여러 가지 불확실성을 안고 있지만, 우선 교실 수업을 통하여 영어 작문시험 과제를 잘 개발하고 채점기준표를 합리적으로 제작한 후 자기 또는 상호 채점자 훈련 및 조정을 통하여 채점을 실시한다면 상당히 타당한 채점 결과를 산출하여 학생들에게도 영어작문에 대한 많은 동기를 부여 할 수 있다고 기대할 수 있다. 특히 본 연구의 영어 작문 채점 분석을 통하여 영어 작문평가와 같은 표현력 시험의 답안에 대해 원어민 채점자들이 채점한 결과를 다국면적 Rasch 측정 모형에 의거 분석함으로써 여러 채점의 양상을 파악하고 또한 채점 오류를 발견하여 영어 쓰기 시험의 채점표와 채점과정을 보완할 수 있다는 것을 알 수 있다. 이러한 Rasch 모형을 이용하여 영어작문 채점결과를 검증하는 것은 영어 표현력 평가의 불확실성을 극복하고 종래에는 신비스럽고 막연하게 느껴졌던 주관적 판단을 보다 명확히 밝히는 노력이라고 볼 수 있다.

그러나 지난 10년간 영어 교육 및 평가에 관한 학술 논문 중에서 측정에 관한 논문이 간혹 나와 있으나, 전문적인 연구가 많지 않은 것을 보면 영어 표현력 시험이 타당하고 채점의 신뢰성을 보장하는 평가로 인식되기에는 아직 요원할 수 있다. 향후 영어 표현력 시험이 영어 학습자에게 교실 수업을 통하여 타당하고 훌륭한 언어 시험으로 인식이 되기 위해서는 영어 표현력 측정 및 분석에 관한 측면도 적극적으로 관여하여 심도 높은 채점 분석에 대한 연구를 실시할 필요가 있다. 이러한 연구 결과는 일시에 성취되는 것이 아니라, 수 차례에 걸친 꾸준한 연구를 통하여 타당한 채점모형을 도출하고 자기 또는 상호 채점자 훈련을 통하여 영어 능력 시험 점수의 신뢰성을 확보한 후에 이뤄질 수 있다고 본다. 이에 따라 본 연구에서 영어 작문 채점 점수의 분석에 활용하였던 다국면 Rasch 측정 모형과 같은 심리측정 모형(psychometric model)이 많은 도움이 될 것으로 기대한다. 비록 본 연구의 채점 분석을 통하여 영어 작문 시험에 대한 약간의 문제는 발견되었지만, 영어 표현력을 지도하는 영어 교육자들도 영어 평가의 측정 문제에도 많이 관여하여 연구결과를 산출한다면 앞으로의 우리 영어 작문 평가는 전문성을 확보할 수 있다고 기대할 수 있다. 이제는 표현력 언어표현력 수행평가로서 영어 작문 시험을 개발·실시하고 주관적인 채점에 대한 신뢰성을 분석·파악하여 언어평가에 요구되는 명확성에 부응하는 자세가 필요하다고 본다.

교보문고 KYOBO Book Centre

참고문헌

신동일, 설현수. (2005). NEW FACETS을 활용한 채점자료 분석 방법. *외국어교육, 12*(2), 191-211.

이영식. (1998). 영어작문평가의 채점신뢰도에 대한 분석. *영어교육, 53*(1), 179-200.

장소영, 신동일. (2009). *언어교육평가 연구를 위한 Facets 프로그램*. 서울: 글로벌콘텐츠.

Bachman, L. F. (1991). What does language testing have to offer? *TESOL Quarterly, 25*(4), 671-704.

Barkaoui, K. (2011). Effects of marking method and rater experience on ESL essay scores and rater performance. *Assessment in Education: Principles, Policy & Practice, 18*(3), 279–293

Brown A. (2003) Interviewer variation and the co-construction of speaking proficiency. *Language Testing, 20*(1), 1-25.

Carr, N. T. (2011*). Designing and analyzing language tests*. Oxford: Oxford University Press.

Davies, A. (1990). *Principles of language testing*. Oxford: Basil Blackwell.

Davis, L. (2009). The influence of interlocutor proficiency in a paired oral assessment. *Language Testing, 26*(4), 367-396.

Eckes, T. (2005). Examining rater effects in TestDaf writing and speaking performance assessment: A many-facet Rasch analysis. *Language Assessment Quarterly, 2*(3), 197-221.

Elder, C., Barkhuizen, G., Knoch, U., & von Randow, J. (2007). Evaluating rater responses to an online training program for writing assessment. *Language Testing, 24*(1), 1-28.

Kecker, G., & Eckes, T. (2010). Putting the manual to the test: the TestDaF—CEFR linking project. In W. Martyniuk. (Ed.), *Aligning tests with the CEFR* (pp. 50-79). Cambridge: Cambridge University Press.

Kozaki, Y. (2010). An alternative decision-making procedure for performance assessments: Using the multifaceted Rasch model to generate cut estimates. *Language Assessment Quarterly, 7*(1), 75-95.

Linacre, J. M. (1989). *Many-facet Rasch measurement*. Chicago, IL: MESA Press.

Linacre, J. M. (1994, March). *An introduction to many-facet Rasch measurement*. The pre-colloquium workshop on FACETS, The 16th Language Testing Research Colloquium. Washington DC.

Linacre, J. M. (1996). *A user's guide to Facets: Rasch measurement computer program* (Version 3.0). Chicago, IL: MESA Press.

Linacre, J. M. (2010). *Facets: Rasch measurement computer program* (Version 3.67.0)

교보문고 KYOBO Book Centre

[Computer software]. Chicago, IL: Winsteps.com.

Linacre, J. M. (2011). *A user's guide to FACETS: Rasch-model computer programs* (Program Manual 3.68.1). Chicago, IL: Winsteps.com.

Lumley, T., & McNamara, T. F. (1995). Rater characteristics and rater bias: Implications for training. *Language Testing, 12*(1), 54-71.

Lynch, B. K., & McNamara, T. F. (1998). Using G-theory and many-facet Rasch measurement in the development of performance assessments of the ESL speaking skills of immigrants. *Language Testing, 15*(2), 158-180.

McNamara, T. F. (1996). *Measuring second language performance*. New York: Addison Wesley.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago, IL: University of Chicago Press. (Original work published in 1960)

Shin, Y. (2010). A FACETS analysis of rater characteristics and rater bias in measuring L2 writing performance. *English Language & Literature Teaching, 16*(1), 123-142.

Shin, Y. (2011). Comparing perceptions of evaluative criteria in EFL writing between learner and instructor group. *English Language & Literature Teaching, 17*(1), 191-208.

University of Cambridge Local Examinations Syndicate. (1988). *Cambridge first certificate examination practice 2*. Cambridge: Cambridge University Press.

Weigle, S. C. (1994). Effects of training on raters of ESL compositions. *Language Testing, 11*(2), 197-223.

Weigle, S. C. (1998). Using FACETS to model rater training effects. *Language Testing, 15*(2), 263-287.

Wu, J. R. W., & Wu, R. Y. F. (2012, October). Parallel-forms reliability of the GEPT speaking test. Paper presented at the 14th Academic Forum on English Language Testing in Asia, Xi'an, China.

부 록

UCLA English as a Second Language Placement Exam (ESLPE)
Composition Rating Scale (Carr, 2011, pp. 129-131)

Content

9-10: The essay fulfills the assignment well and treats the topic with sophistication. The main idea is clear and well-developed. Support is relevant, thorough and credible.

7-8: The essay addresses the assignment appropriately* and is developed competently. The main idea is clear and competently developed, but with less sophisticatin and depth than

교보문고 KYOBO Book Centre

the 9-10 paper. Arguments/ideas are competently supported.

5-6: The essay addresses the writing task adequately, but may not be well-developed. OR The essay only addresses part of the topic, but develops that part sufficiently. The main idea is clear but may not be fully developed. Ideas/arguments may be unsupported or unrelated to main idea.

3-4: The essay only partially fulfills the writing task OR the main idea is somewhat clear, but requires the reader to work to find it. The essay contains unsupported or irrelevant statements.

1-2: The essay fails to fulfill the writing task, and lacks a clear main idea and development. Most ideas/arguments are unsupported, and ideas are not developed. OR Not enough material to evaluate.

*Note: *Appropriate* is defined as addressing all aspects of a topic, for example, all characteristics in questions involving choices. Furthermore, all parts of the writing task should be touched on in the writer's response.

Rhetorical Control

9-10: Introduction and conclusion effectively fulfill their separate purposes: the introduction effectively orients the reader to the topic and the conclusion not only reinforces the thesis but effectively closes off the essay. Paragraphs are separate, yet cohesive, logical units which are well-connected to each other and to the essay's main idea. Sentences form a well-connected series of ideas.

7-8: The introduction presents the controlling idea, gives the reader the necessary background information, and orients the reader, although there may be some lack of originality in the presentation. The conclusion restates the controlling idea and provides a valid interpretation but not as effectively as 9-10 paper. Paragraphs are usually cohesive and connected to the essay's main idea. Sentences are usually well-connected.

5-6: Introduction presents the controlling ideas but may do so mechanically or may not orient the reader to the topic effectively. The conclusion does not give the reader new insights or may contain some extraneous information. Paragraphs may exhibit a lack of cohesion or connection to the essay's main idea. Sentences may not be well-connected.

교보문고 KYOBO Book Centre

3-4: Introduction and conclusion do not restate the controlling idea. Introduction fails to orient the reader adequately, and the conclusion may be minimal or may not be tied to the rest of the essay. Paragraphs often lack cohesion and are not appropriately connected to each other or to the essay's main idea. Sentences are not well-connected.

1-2: Introduction and conclusion are missing or unrelated to rest of the essay. There is no attempt to divide the essay into conceptual paragraphs, or the paragraphs are unrelated and the progression of ideas is very difficult to follow. OR Not enough material to evaluate.

Language (grammar, vocabulary, register, mechanics)
9-10: Except for rare minor errors (esp. articles), the grammar is native-like. There is an effective balance of simple and complex sentence patterns with coordination and subordination. Excellent, near-native academic vocabulary and register. Few problems with word choice.

7-8: Minor errors in articles, verb agreement, word form, verb form (tense, aspect) and no incomplete sentences. Meaning is never obscured and there is a clear grasp of English sentence structure. There is usually a good balance of simple and complex sentences both appropriately constructed. Generally, there is appropriate use of academic vocabulary and register with some errors in word choice OR writing is fluent and native-like but lacks appropriate academic register and sophisticated vocabulary.

5-6: Errors in article use and verb agreement and several errors in verb form and/or word form. May be some incomplete sentences. Errors almost never obscure meaning. Either too many simple sentences or complex ones that are too long to process. May be frequent problems with word choice; vocabulary is inaccurate or imprecise. Register lacks proper levels of sophistication.
3-4: Several errors in all areas of grammar which often interfere with communication, although there is knowledge of basic sentence structure. No variation in sentence structure. Many unsuccessful subordinated or coordinated structures. Frequent errors in word choice (i.e. wrong word, not simply vague or informal). Register is inappropriate for academic writing.

1-2: There are problems not only with verb formation, articles, and incomplete sentences, but sentence construction is so poor that sentences are often incomprehensible. Sentences that are comprehensible are extremely simple constructions. Vocabulary too simple to express meaning and/or severe errors in word choice. OR Not enough material to evaluate.

교보문고 KYOBO Book Centre

예시언어(Examples in): English
적용가능 언어(Applicable Languages): English
적용가능 수준(Applicable Levels): Tertiary

이영식
한남대학교 사범대학 영어교육과
[306-791] 대전광역시 대덕구 한남로 70
Tel: 042)629-7415 / Fax: 042)629-8395
E-mail: youngslee@hnu.kr

Received in January 15th, 2014
Reviewed in February 26th, 2014
Revised version received in March 5th, 2014

KYOBO 교보문고
교보문고 KYOBO Book Centre