creative commons

COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

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

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

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

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

심리학 석사 학위논문

구성적 문단 배열 문항의 신뢰도와 타당도

2023년 8월

서울대학교 대학원

심리학과 인지심리전공

김 휘 민

구성적 문단 배열 문항의 신뢰도와 타당도

지도교수 박 주 용

이 논문을 심리학 석사 학위논문으로 제출함
2023년 8월

서울대학교 대학원
심리학과 인지심리전공
김 휘 민

김휘민의 석사 학위논문을 인준함
2023년 7월

위 원 장 _____ 한 소 원 _____ (인)

부위원장 _____ 차 지 욱 _____ (인)

위 원 _____ 박 주 용 _____ (인)

국문초록

오늘날 대다수의 독해 검사는 글의 지엽적인 정보를 묻는 문항들이 많다. 하지만, 글의 지엽적인 정보를 묻는 것만으로는 수험자의 독해 능력을 정확히 파악할 수 없다는 문제가 있다. 이에, 본 연구에서는 지엽적인 기존의 독해 문항 대신 글의 구조적인 맥락을 파악해야 하는 문단 배열 문항을 제안하였다. 문단 배열 문항 과제의 특징은 수험자가 무선적으로 배열된 넷 또는 다섯 개의 문단을 논리적 흐름에 따라 배열해야 한다는 점이다. 본 연구에서는 단순히 답지를 고르는 선다형 대신, 배열 순서를 수험자가 직접 입력하도록 하였다. 수험자의 입력된 반응은, 정답과의 유사성을 계산하는 문자열 유사성 알고리즘(string similarity algorithm)에 따라 점수가 매겨졌다. 이 방식의 신뢰도와 타당도를 알아보기 위해 대학생을 대상으로 세 개의 연구가 수행되었다.

연구 1에서는 구성적 문단 배열 문항의 동형 검사 신뢰도를 선다형과 비교하였다. 각 문항에 대해 수험자는 먼저 직접 답을 입력하고 바로 이어서 선다형으로 반응하였다. 검사 점수를 분석한 결과, 구성적으로 반응한 두 검사 간의 상관이 선다형에서의 상관에 비해 유의미하게 높았다. 연구 2에서는 한 수험자가 4 문단짜리 16 문항으로 이루어진 검사와 5 문단짜리 8 문항으로 이루어진 검사를 모두 보았다. 검사 점수를 분석한 결과 두 검사 점수 간에 높은 상관이 확인되었다. 이 결과는 문단의 개수를 4개에서 5개로 늘려 더 어렵게 만들면, 검사 시간을 줄일 수 있음을 시사한다. 연구 3에서는 구성적 문단 배열 검사가 구조파악 능력을 측정하는 지를 알아보았다. 이를 위해 Gernsbacher와 Varner (1988)가 개발한 구조 파악 능력 측정 검사인 다중미디어 독해 배터리(Multi-Media Comprehension Battery: MMCB)가 사용되었다. 연구 참여자들로 하여금 구성적 문단 배열 검사와 MMCB의 한 부분인 그림 검사를 치르도록 한 결과, 이들 간에 높은 상관이 확인되었다. 이 결과는 구성적 문단 배열 검사가 구조 파악 능력을 측정하는 도구로 활용될 수 있음을 시사한다.

이상의 결과는 구성적 문단 배열 검사가 종래의 독해 검사들이 놓치고 있던 구조 파악 능력을 측정하는 방안으로 사용될 가능성을 제시한다. 문단 배열 문항은 오래 전부터 독해 평가를 위해 사용되어 왔지만, 선다형 방식으로 제시되어 그 장점이 충분히 발휘되지 못했다. 본 연구에서 제안한 것처럼, 구성적 방식으로 반응하게 하고 그 결과를 정답과의 유사성을 바탕으로 부분 점수를 매기면, 수험자의 독해 능력을 보다 정밀하게 측정할 수 있다. 이 방식의 또 다른 장점은, 교수자가 수업 관련 자료를 이용하여 쉽게 만들 수 있고, 문단의 개수나 글의 난도를 조절하여, 다양한 수준의 문항을 출제할 수 있다는 것이다. 요컨대 구성적 문단 배열 문항은, 그동안 간과되어 온 구조 파악 능력을 쉽게 측정할 수 있게 해, 독해 평가 그리고 나아가 독해 교육의 새로운 지평을 열 것으로 기대된다.

주요어 : 문단 배열 문항, 구성적 반응, 구조 파악 능력, 독해, 평가
학 번 : 2021-26046

목 차

제 1 장 서 론 ····· 1

제 1 절 연구의 배경 ····· 1

제 2 절 연구 문제 및 가설 ····· 4

제 2 장 이론적 고찰 ····· 6

제 1 절 독해 모형 ····· 6

제 2 절 독해 평가 ····· 9

제 3 절 구조 파악 능력과 측정 ····· 12

제 4 절 구성적 문단 배열 문항 ····· 14

제 3 장 연구 1 ····· 18

제 1 절 참여자 ····· 18

제 2 절 도구 ····· 19

제 3 절 절차 ····· 20

제 4 절 결과 ····· 21

제 5 절 논의 ····· 23

제 4 장 연구 2 ····· 26

제 1 절 참여자 ····· 26

제 2 절 도구 ····· 27

제 3 절 절차 ····· 27

제 4 절 결과 ····· 28

제 5 절 논의 ····· 29

제 5 장 연구 3 ····· 31
제 1 절 참여자 ····· 31
제 2 절 도구 ····· 31
제 3 절 절차 ····· 32
제 4 절 결과 ····· 33
제 5 절 논의 ····· 34

제 6 장 종합 논의 ····· 36

참고문헌 ····· 39
Abstract ····· 44

표 목 차

[표 1] 검사 A와 B의 평균과 표준편차 ····· 21
[표 2] 검사 A와 B의 부분점수 제거 후 평균과 표준편차 22
[표 3] 검사 C와 D의 평균과 문항 평균 ····· 28
[표 4] MMCB와 문단 배열 문항의 평균과 표준편차 ····· 33

제 1 장 서론

제 1 절 연구의 배경

독해는 가장 기본적이면서도 중요한 학습 활동이다 (Alexander and the Disciplinary reading group, 2012). 글을 읽고 이해할 수 있으면 지금까지 축적된 지식에 쉽게 접근할 수 있을 뿐만 아니라 이를 발판으로 삼아 더 발전된 지식을 만들어낼 수도 있다. 이런 이유에서, 독해 능력의 중요성은 꾸준히 강조되어 왔다. 독해 능력이 학습에 그치지 않고 직업까지 영향을 준다는 사실은, 경제협력개발기구(Organization for Economic Cooperation and Development : OECD, 2013)의 국제성인역량조사(PIAAC · 16~65세 대상) 보고서에서 확인할 수 있다. 이 보고서에 따르면 가장 높은 수준의 문해력을 갖춘 사람들이 최하위 수준에 비해 평균 시급은 60% 그리고 취업 가능성은 2배 이상 높았다.

독해의 중요성이 지속적으로 강조되고 있음에도 불구하고 사람들의 독해 수준은 기대에 못 미치고 있다. OECD (2013)에 따르면 세계적으로 성인들 중 11%만이 심층 수준의 독해 능력(deep level of reading comprehension)을 갖고 있다고 한다. 실제로 48.8%에 해당하는 성인들이 낮은 수준 이상의 추론을 요구하는 글을 이해하지 못한다 (Provasnik, 2018). 성인뿐만 아니라 초•중등 학생들의 상황도 다르지 않다. 지난 1992년부터 4, 8, 그리고 12학년을 대상으로 이루어진 미국의 국가수준 성취도평가(NAEP)에 따르면, 초•중등 학생들의 독해력도 크게 향상되지 않거나 오히려 하락하고 있다.

이런 상황을 개선하기 위한 노력이 오랫동안 이루어져왔다 (Israel & Duffy, 2017; Kamil et al., 2008; McNamara, 2007; Mohr et al., 2020). 하지만 여러 노력에도 불구하고 학생들의 독해력 향상은 물론 독해 교육 방식에서도 큰 변화를 찾아보기 어렵다. 따라서 지금까지 해왔던 것과는

다른 새로운 시도가 필요해 보인다.

본 연구는 독해 능력 향상을 위해 평가, 특히 교실 장면에서 쉽게 사용할 수 있는 문항 형식 개발에 초점을 두었다. 구성적 문단 배열 문항으로 명명된 이 방식에서, 수험자는 제시된 여러 개의 문단이 논리적으로 배열되도록 그 순서를 입력해야 한다. 구성적이라는 형용사는 선다형에서처럼 답지가 주어지지 않고, 단답형 문항에서처럼 수험자가 직접 입력해야 하는 특징을 부각시키기 위해 사용하였다 (Park, 2010).

구성적 문단 배열 문항을 제안하는 이유는 세 가지이다. 우선, 현재 독해 평가를 위해 널리 쓰이는 선다형 방식에는 많은 문제가 있기 때문이다. 선다형은 대체로 4~5개의 답지를 주고 그중에 정답을 고르게 한다. 선다형은 채점이 쉽고 경제적이라는 장점이 있다. 하지만 선다형 문항으로는 깊이 있는 독해력 측정이 어렵다는 지적이 계속되어 왔다 (Leslie & Caldwell, 2017). Park (2005, 2010)은 선다형 문항의 한계를 과정 정보의 부재로 보았다. 수험자가 정답을 골랐더라도 정말 알고 골랐는지 아니면 우연히 맞춘 것인지를 구분할 수 없다는 것이다. 실제로 제시된 글을 제대로 이해하지 못하더라도 답지를 보고 내용을 추측해 정답을 고르는 전략을 가르치기도 한다. 따라서 이런 전략을 예방할 수 있는 문항이 필요한데, 구성적 문단 배열 문항은 그 한 대안이 될 수 있다.

두 번째 이유는, 독해 평가 검사에서 전체적인 글의 구조 파악을 요하는 문항보다는 세부적인 내용 파악을 묻는 문항들이 과도하게 많기 때문이다. 예를 들어, 2016년부터 2021년의 대학수학능력시험 국어 장문독해 문제들을 분석했을 때, 내용의 사실을 확인하는 사실적 사고 문항이 66%에 해당하고 전체적인 글의 구조나 논리를 분석하여 추론하는 문항은 20%밖에 되지 않았다. 구조 파악(structure building)이란 이야기, 글, 혹은 일련의 그림에서 하나의 응집성 있는 심적 표상을 구축하는 능력을 가리킨다 (Gernsbacher, 1996). 이를 통해, 글의 여러 부분들이 대비나 비교되거나, 인과적 혹은 논리적으로 연결되면서 하나의 응집력 있는 전체로 묶이게 된다. 예를 들어, “나는 배가 맛있고 비행기는 별로야. 너는 어떤 과자가 맛있어?”에서 첫 문장을 바로 이해하기 어렵다. 배와

비행기가 서로 다른 맛을 보이는 과자의 모양이라는 것을 알게 될 때 비로소 두 문장 전체에 대한 구조가 파악되는 것이다. 이런 구조 파악 능력이 독해에서 핵심적인 역할을 하는 만큼 이를 정확히 측정할 수 있는 문항이 필요하다. 구성적 문단 배열 문항은 이런 필요를 충족시킬 수 있다.

새로운 문항을 제안하는 세 번째 이유는 최근 개발된 여러 독해 평가 도구들이, 여러 장점이 있지만, 교실 평가를 위해 사용될 가능성이 낮기 때문이다. 예를 들어, 최근 개발된 독해 평가 도구인 Dray 등 (2019)의 발달적 접근과 Sheehan & O'Reilly (2012), Sabatini 등 (2020)에 의한 시나리오 기반 평가(Scenario Based Assesment, SBA)를 살펴보자. 이론적 배경에서 더 자세하게 다룰 예정이지만, 발달적 접근이나 SBA는 탄탄한 교육 측정 이론에 근거하고 있을 뿐만 아니라 21세기에 요구되는 읽기의 특징을 반영하고 있다. 하지만 개발에 드는 시간과 비용이 크다는 문제가 있다. 여러 분야의 전문가들의 협력 작업은 물론 개발된 문항에 대한 예비 검사 작업 등에 상당한 시간과 비용이 들어 누구나 쉽게 문항을 개발하고 사용하기 어렵다. 사용되는 문서의 유형은 물론 이들을 이용하여 측정하려는 구인을 정확히 포착하는 문항은 전문가들이 시행착오를 거치며 만들어야 하기 때문이다. 따라서 교실 장면에서의 평가 도구로 쉽게 사용되기 어렵다. 반면 구성적 문단 배열 문항은 상대적으로 출제가 쉽기 때문에, 교실 평가를 위해 널리 활용될 가능성이 높다.

동시에 이 문항은 다른 단답형이나 논술 같은 구성적 문항과 마찬가지로, 채점이 쉽지 않다. 간단한 해결책인 정답과 오답으로 나누는 방법이 있기는 하다. 하지만 더 좋은 해결책은 부분 점수를 주는 것이다. 오답이라 하더라도 정답과의 유사성에서 큰 차이가 있기 때문이다. 부분 점수를 주는 한 방법은 유사성에 따라 일관된 점수를 부여할 수 있는 알고리즘을 이용하는 것이다. 본 연구에서는 컴퓨터 공학의 문자열 유사성 알고리즘(string similarity algorithm)을 이용하여 오답이라 하더라도 유사성에 따라 부분점수를 주는 방식을 택했다. 여러 알고리즘 중 예비 연구를 통해 가장 안정적인 점수 차이를 보이는 방안을 적용하였다.

정리하면 본 연구에서는 구성적 문단 배열 문항을 이용하여 기존의 독해 평가 문항에서 다루기 어려운 구조를 다룰 수 있게 하고, 직접 입력하게 하여 추측을 방지하되 반응을 정답과의 유사성을 기준으로 부분 점수를 주는 방안을 제안하였다. 제안된 방안의 신뢰도와 타당도를 알아보는 세 개의 연구가 수행되었다.

제 2 절 연구 문제 및 가설

본 연구에서는 구조 파악 능력을 측정할 수 있는 구성적 문단 배열 문항을 제안하는 동시에, 문자열 유사성을 이용한 부분점수를 부여 방안을 제안하였다. 이 문항은 제시된 몇 개의 문단을 논리적 흐름에 맞게 배열하고 그 순서를 수험자가 직접 입력하게 한다. 입력된 순서는 문자열 유사성 알고리즘에 의해 정답과의 유사성에 따라 점수가 매겨진다. 제안된 문항의 심리측정학적 특성을 알아보기 위한 연구 문제와 가설은 다음과 같다.

첫째, 제안된 방안은 같은 발문에 대해 선다형으로 반응하게 했을 때보다 더 높은 신뢰도를 보이는가? 종래의 문단 배열 문항은 발문과 답지로 이루어진 선다형으로 제시되었다. 문단 배열 문항을 선다형으로 제시하면 여러 가지 경우의 수 중에서 5개의 순서 배열만이 답지로 제시되기 때문에 추측해서 정답을 쉽게 고를 수 있다는 한계가 있었다. 이 한계를 극복하기 위해, 본 연구에서는 수험자로 하여금 배열 순서를 직접 입력하게 하는 구성적 방식을 사용하였다. 이렇게 하면 답지를 이용해 정답을 찾아낼 가능성을 줄여 더 정확한 평가가 가능할 것으로 예상하였다. 두 방식을 비교하기 위해, 같은 발문에 대해 순서를 직접 입력하는 구성적 방식으로 먼저 반응하게 한 다음 선다형 방식으로 제시된 답지 가운데 고르도록 하였다. 선다형 방식보다 구성적 방식에서 더 높은 신뢰도를 예측하면서, 두 개의 문단 배열 동형 검사를 실시하였다. 각 방식의 상관계수를 비교하여 두 방식 간의 신뢰도 차이를 확인하였다.

둘째, 문단의 개수를 늘려 난도를 높이면 더 적은 수의 문항으로 구성된 검사로 수험자의 구조파악 능력을 측정할 수 있을까? 문단 배열 문항의 난이도를 조절하는 간단한 방법은 문단의 개수다. 예를 들어, 문단의 개수를 4개에서 5개로 증가시키면 문항 난도는 높아진다. 그 대신 각 문항에 소요되는 시간이 더 길어진다. 그렇다면 문단수를 늘려 난도를 높이는 대신, 더 적은 수의 문항으로 구성된 검사를 사용하면 검사 시간을 줄이면서도 안정적으로 구조 파악 능력을 측정할 수 있을까? 이 가능성을 알아보기 위해, 수험자에게 4개의 문단을 배열하는 16개 문항으로 구성된 검사와 5개의 문단으로 된 8개의 문항으로 구성된 검사를 모두를 응시하게 한 다음, 두 검사 간의 상관을 비교하였다. 두 검사 간의 상관이 높게 나타나면, 더 어려운 대신 더 적은 문항으로 구조 파악 능력을 측정할 수 있게 된다.

마지막으로, 제안된 방식의 검사가 실제로 구조 파악 능력을 측정하는가? 본 연구에서 제안하는 문항은, 세부 사항을 묻는 기존 독해 문항과 달리, 글의 구조와 논리적 연결성을 확인하는 구조 파악 능력을 측정하고자 한다. 문단 배열 문항이 정말 구조 파악 능력을 측정하는 지를 확인하기 위해, 구조 파악 능력을 측정하는데 사용되어 온 Gernsbacher와 Varner (1988)의 다중미디어 독해배터리(Multi-Media Comprehension Battery: MMCB) 중 일부 문항과의 동시타당도를 검증하였다. MMCB는 간단한 동화를 글이나 음성, 혹은 그림으로 제시한 후 그 각각의 내용과 관련된 문제를 풀게 한다. 본 연구에서는 구성적 문단 배열 문항 검사와 함께, MMCB의 일부 검사를 차례로 실시한 다음, 이들 간의 상관을 조사하였다. 문단 배열 문항 검사가 구조 파악 능력을 측정한다면, 이 검사와 MMCB의 일부 검사 간의 상관이 높을 것으로 예상하였다.

제 2 장 이론적 고찰

이 장에서는 독해 모형에 대해 알아본 뒤 종래의 독해 평가들을 소개하였다. 기존의 독해 평가 방식의 문제점들과 한계점을 살펴보고 이를 극복하기 위해 제안된 방법들도 살펴보았다. 마지막으로 본 연구에서 초점을 맞춘 구조 파악 능력과 이를 평가하기 위한 구성적 문단 배열 문항에 대해 설명하였다.

제 1 절 독해 모형

우리가 통상 읽는 글은 담화(discourse)이다. 즉, 국지적 통일성(local coherence)을 이루는 문장들이 모여 전체로서 전달하려는 메시지(message)가 있는 일련의 문장이나 단락의 집합이다. 국지적 통일성이란 인접한 문장들이 명시적 혹은 암묵적으로 지칭하는 대상이 수렴됨을 의미한다. 문장이나 단락의 집합은 예를 들어, 주장-증거-반박-재반박이나, 인물-배경-사건 등과 같은 여러 층위로 구분되는 거시 구조를 갖는다. 메시지란 대상이나 사건에 대한 설명 혹은 글쓴이의 주장이나 글을 통해 전하고 싶은 교훈 등을 가리킨다. 독해란 결국 이 메시지를 그 구조와 함께 파악하는 활동인데, 그 과정에 개입되는 변인들이 많고 그들 간의 상호작용도 매우 복잡하다.

독해에 관련된 변인들은 물론 그 처리 과정을 명세하기 위해 다양한 독해 모형이 제시되었다. 이 중 대표적인 모형을 몇 가지 살펴보면 다음과 같다. 고전적인 단순 모형에서는 독해의 두 요소로 단어 해독(decoding)과 어휘력과 구문론과 같은 언어적 이해(linguistic comprehension)를 강조한다. 이 모형은 글을 막 배우는 사람들의 독해능

력을 비교적 잘 예측하지만, 구조를 다루지 못하고 문장 하나하나를 이해하는 데 머문다는 한계가 있다 (Woolley, 2011). 또 다른 영향력 있는 모형은 도식모형이다. 이 모형은 독해가 특정 상황에서 반복되는 경험을 통해 발달된 도식을 매개로 일어난다고 본다. 독해란 전체 담화를 특정 도식에 끼워 맞추는 과정으로 이 과정에서 담화 속의 일부 내용이 삭제되거나 각색된다는 것이다. 그렇지만 이 모형은 독해 과정의 융통성이나 맥락의 영향을 잘 설명하지 못한다. 그럼에도 독자가 갖고 있는 배경 지식이 독해에 중요함을 부각시키는데 중요한 기여를 하였다.

독해 연구에서의 획기적인 변화는 구성 통합 모형 (Kintsch, 1988)을 통해 일어났다. 이 모형에서는 독해가 구성과 통합의 두 단계에 걸쳐 일어난다고 제안한다. 구성 단계에서는 여러 수준의 관련 정보들이 활성화되면서 상향식 정보 처리가 일어난다. 이어지는 통합단계에서는 제약만족(constraint satisfaction)을 통해 여러 조건을 충족시키는 안정화된 활성화 패턴을 찾아내는 방식으로 독해가 일어난다고 본다. 간단한 예를 들자면, We waited him by the bank. He pulled the fish up onto the bank. 라는 두 문장에서 the bank는 a bank building or a river bank의 두 가지 다른 의미로 해석될 수 있다. 구성 단계에서 짧은 시간 동안, 앞 뒤 문장에 등장하는 bank의 두 의미가 모두 활성화되어 최소한 네 개 이상의 다른 명제들이 만들어지지만, 통합 단계에서는 fish가 활성화되면서 이와 연관된 a river bank는 강해지고 a bank building은 억제되어 독해가 일어난다는 것이다.

보다 최근의 van den Broek과 Kendeou (2022)의 포괄적 담화 이해 모형(Comprehensive Model of Discourse Comprehension)은 Kintsch의 모형을 더욱 정교화하였다. 이들은 독해를 텍스트 입력(text input)으로부터 추론과 타당화에 의존하는 일관성-구성 과정(coherence-building processes)을 통해 심성 모형을 구축하는 과정으로 보는데, 일관성 구성 과정은 일관성의 기준(standards of coherence)과 독자의 특성에 영향을 받는다고 가정한다. 이 기준은 개인마다 차이가 있을 뿐만 아니라 소설을 읽을 때와 교과서를 읽을 때처럼 독서의 목적에 따라 달라진다. 일반

적으로 소설의 경우 잘 이해되지 않는 부분이 있어도 읽어나갈 수 있지만, 교과서의 경우는 소설보다는 통일성 기준이 높다. 일단 그 기준에 미치지 못하면, 일관성이 없거나 연결이 되지 않는 부분을 해소하기 위한 의도적인 통제 과정이 작동한다. 따라서 독해 과정은 텍스트 정보를 바탕으로 한 자동적 처리 과정이 통일성 기준에 미치지 못할 때 작동하게 되는 의도적인 통제 과정이 작용하여 상황 모형을 만드는 과정이라 할 수 있다.

Kintsch의 모형과 van den Broek과 Kendeou의 모형이 여러 실험 연구 결과를 잘 설명하지만, 명시적으로 드러나지 않은 심층적 의미를 어떻게 포착하는지를 설명하지 못하는 한계가 있다. 토끼와 거북이 이야기를 예로 들면, 느림보 거북이가 잘 뛰는 토끼를 경주에서 이긴 것은 꾸준함 때문이라는 교훈은 텍스트 기반이나 상황 모형 이상으로 해석이라는 고차적인 처리가 필요하다.

이런 흐름에서 최근 깊이의 차원(shallow-deep dimension)을 구분하여, 같은 글을 읽더라도 더 많이 배우고 명시적으로 드러나지 않은 추상적 의미를 찾아내도록 돕기 위한 방법을 탐색하고 있다 (Kendeou & O'Brien, 2017; Lillis et al., 2019). 또한 Common Core State Standards(CCSS)에서도 학생들이 더 깊이 글을 이해할 수 있는 방향으로 교육과정을 바꾸고 있다. 이제는 정보를 단순히 기억하는 것이 아니라 분석하고 통합하여 평가하는 과정이 중요시된다. 결과적으로 독해 과정에 대한 연구는 독해가 어휘, 텍스트의 복잡성, 독자의 사전지식 수준 그리고 읽는 목적 등에 영향을 받는 복잡한 과정으로 이해된다. 아울러 독자가 텍스트 이해 이상으로 드러나지 않은 의미를 구축하거나 자신의 목적에 맞게 활용하려면 평가, 비판, 그리고 창조와 같은 고차적 사고의 개입이 요청된다 (Cartwright, 2020). 그렇지만 이런 고차적 사고가 개입되는 포괄적인 독해 모형은 아직 제시되지 않아, 앞으로의 연구 과제로 남아 있는 상태이다.

제 2 절 독해 평가

독해에 대한 이상의 이론적 발전은 독해 과정의 복잡성을 보여준다. 하지만 이런 발전이 교실 장면에서 독해력 향상을 위한 교수법과 평가 방법에 이르지는 못하고 있다. 이상적으로는 명확한 독해력 구인을 기반으로 독해 역량을 정확히 측정하고 이를 향상시킬 수 있는 교수가 진행되어야 한다. 하지만 실제로는 교수법은 물론 평가 방법에서 큰 변화를 찾아보기 어렵다. 교수법은 많이 읽게 하는 데 초점을 두고 있고, 전통적인 독해 검사와 유사한 평가 방식이 계속 사용되고 있다.

전통적으로 사용되어 온 독해 검사는, 긴 문단(passages)과 함께 제시된 선다형 문항으로 구성된다. CBM-Maze test (Deno, 1985; Espin & Foegen, 1996)와 Gates-MacGinitie Reading Comprehension Test (MacGinitie et al., 2000)는 그 대표적인 예이다. CBM-maze 문항은 문장마다 규칙적으로 빈칸을 만들어 이를 채워 넣는 문제를 답지와 함께 제시하여 학생들의 독해능력을 측정한다. Gates-MacGinitie Reading Comprehension Test는 48개의 선다형 문항으로 구성되어 있는데, 해독기술(code-related skill)과 언어 이해 기술을 측정하는 것으로 알려져 있다 (MacGinitie, 2002). 이런 검사들이 가진 문제점은 수많은 연구자들에 의해 지적되었는데, 대표적으로 선다형 문항 형식의 과도한 사용 (Rupp et al., 2006), 발달적인 관점의 부족(lack of developmental perspective) (Dray et al., 2019) 이론적 모형과 평가 설계 간의 연결성 부족 (Sabatini et al., 2020), 지시문과의 약한 연결성 (Sabatini et al., 2020), 훈련에 대한 구체적 정보의 부족 (Dray et al., 2019) 등이 있다.

문항 제작이 쉽다는 장점이 있는 cloze 문항은 수업 시간에 다룬 텍스트나 관련 자료 중 일부를 빈칸으로 바꾸면 쉽게 만들 수 있다. 하지만, 이런 문항은 개인의 읽기 능력을 평가하는 것이 아니라 개별 텍스트의 가독성을 확인하는 정도로 얕은 수준의 읽기 능력을 측정한다는 한계가 있다 (Jensen & Elbro, 2022). 이런 한계를 극복하기 위해 cloze 문항을 고도화한 deep cloze test가 제안되었다 (Jensen & Elbro, 2022). deep

cloze 문항은 문장마다 규칙적으로 빈칸을 만드는 대신 텍스트 내에서 전략적으로 핵심이 되는 단어를 비우는 방식으로 만들어진다. 이 연구자들은 deep cloze 문항을 사용할 경우 전통적인 cloze 문항에 비해, 독해 능력을 유의미하게 더 잘 설명한다는 결과를 제시하였다.

하지만 deep cloze 문항도 다음과 같은 두 가지 한계가 있다. 하나는 문항을 만드는데 시간이 많이 들 뿐만 아니라, 사용되는 텍스트의 내용에 따라 상당한 전문성이 필요해 교사들이 쉽게 사용하기 어렵다는 것이다. 또 다른 문제는 문항이 모두 선다형으로 제시되어 빈칸에 들어갈 단어를 답지로 제시하기 때문에 답지를 보고 정답을 추측하는 역행추리가 가능하다는 점이다. 과도한 선다형 문항 사용은 컴퓨터화 평가에서 사용되는 여러 혁신 문항을 통해 어느 정도 해결이 가능하다. 예를 들어 답지를 이용한 역행 풀이를 저지하기 위해, 선다형의 발문만 제시하고 답지는 수험자가 요청하면 짧은 시간동안만 제시하고 제시된 시간 내에 반응하도록 하는 방법을 사용할 수 있다 (Park, 2005).

나머지 문제들에 대한 해결책은, Sabatini et al. (2020)과 Dray et al. (2019)에서 볼 수 있다. Sabatini 등의 시나리오 기반 평가에서는 읽기 능력(reading competency)을 읽기 기술(reading skills), 읽기 전략(reading strategies), 그리고 글(text)의 세 차원으로, 각 차원을 다시 하위 세 개의 범주로 세분하고, 각 범주는 다시 3~4개의 하위 범주로 세분하였다. 예를 들면, 읽기 전략 차원(reading strategies dimension)은 읽기 전 전략(pre-reading strategies), 모형 구성 전략(model-building strategies), 글 모형 구성을 위한 전략(strategies for model-building the text)을 포함한다. 그리고 모형 구성 전략은 질문하고 예측하고 구조를 구성하고 재구성하는 전략들을 포함한다.

이상의 핵심 역량 개념을 바탕으로 실제 과제를 설계할 때, 독해 과정이 사회적 상호작용이 내재된 목표지향적 활동으로 여러 출처의 정보를 통합해야 한다는 점에 비중을 두었다. 실제로 SBA에 기반한 대규모 독해 평가 도구인 GISA는 웹상에서, 수업 시간과 유사하게 45~50분 정도 검사를 치르게 하며, 구성적 문항의 경우 자동채점이 가능하도록 하는

제약 속에서 구현되었다. 하나의 주제로 묶일 수 있는 여러 관련 자료들을 제시한 다음, 학생들이 목적에 맞게 여러 자료를 통합하고 평가하는지와 함께, 자료 전반에 대한 심성 모형 구축, 배경지식과 사회적 상호작용 등을 측정한다. 구체적인 평가 문항은 특정 글을 요약하거나, 쟁점에 대한 찬성 반대 주장을 그래픽 오거나이저(graphic organizer)로 정리하거나 제시된 문제를 해결하는 방안을 제시하는 활동 등을 포함한다. Sabatini와 동료들은 GISA가 이처럼 자료나 문항 형식에서 큰 변화가 있음에도 불구하고, 문항들이 충분한 심리측정학적 속성을 나타내며, 전통적 검사와의 동시타당도가 높다는 연구 결과를 제시하였다 (Sabatini et al., 2020).

Dray 등 (2019)은 구성개념도 접근(construct map approach)을 통해 독해와 관련된 핵심 역량 혹은 구인에 기반하여 발달적 변화를 포착할 수 있는 문항을 개발하였다. 이들은 독해를 글의 구조 사용과 지식 생성과 관련된 방법이라는 두 구인으로 구분하고, 전자는 1. 글의 표면적 구조에 반응하기 2. 글의 특성 파악하기 3. 글의 구조 해석하기 4. 저자의 의도 판단하기 5. 의사소통 적용과 평가로, 후자는 1. 분리 2. 글의 아이디어에 참여하기 3. 중심 아이디어 식별하기 4. 글의 주요 아이디어를 상호 확인하고 조정하기 5. 지식을 종합하고 새로운 이해 창출하기의 위계적인 다섯 단계로 구분하였다. 이후 각 단계에 해당하는 활동이 일어났는지를 확인하기 위한 문항을 개발하고 개발된 문항의 심리측정학적 타당도를 확인하는 연구를 수행하였다.

SBA나 구성개념도 접근은 기존의 독해 평가가 가진 문제점들을 많이 보완한 방식임에는 틀림이 없다. 하지만 개발에 드는 시간과 비용이 많다는 문제가 있다. 그렇기 때문에 여전히 교실 현장에서 교사들이 사용하기 어렵다는 큰 단점이 존재한다. 따라서 교실 장면에서 교사가 쉽게 사용할 수 있으면서도 구조 파악을 측정할 수 있는 도구가 필요하다.

제 3 절 구조 파악 능력과 측정

앞서 살펴본 Kintsch의 구성 통합 모형에서의 통합 과정과 van den Broek과 Kendeou (2022)의 담화 이해 모형에서 등장하는 일관성-구성 과정은 둘 다 독해를 심성 모형을 구축하는 과정이 핵심이라고 파악한 점에서 일맥상통한다. 결국 글을 잘 읽기 위해서는 글의 표면만 이해하는 것이 아닌 글에서 얻은 정보들을 구조적으로 통합하여 심성 모형을 형성해야 한다는 것이다. 그렇지만 이 두 모형은 글을 이해하는 과정에 초점을 둔 모형으로 사람들의 독해 능력 간의 개인차를 설명하지 못한다는 한계가 있다. 다행히도 구조 파악 모형이 독해 과정에서의 개인차를 다룬다 (Gernsbacher, 1990).

구조 파악 모형은 정보 전달 형식, 즉 글이나 대화 혹은 그림이든 상관없이, 그 요지를 파악하는 일반적인 능력을 다룬다. Gernsbacher (1990)는 구조 파악 능력이야말로 사람들의 인지 능력의 차이를 만드는 근본적인 요인이라고 주장했다. 높은 수준의 구조 파악 능력을 갖춘 사람들은 중요하고 일관적인 심성 모형을 쉽게 구축할 수 있지만 그렇지 못한 사람들은 많은 시간을 주더라도 단편적이거나 핵심이 담기지 않은 심성 모형을 만들어낸다는 것이다.

이 모형은 건축 과정에 비유되는 세 과정으로 이루어졌다. 먼저 토대를 놓고(laying a foundation), 그 토대 위에 정보를 배치한다(mapping). 그런데 만일 새로운 정보가 기존의 토대에 배치될 수 없으면, 새로운 토대를 만드는 이동하기(shifting)가 이루어진다는 것이다. 일단 토대가 만들어진 다음 새로운 정보가 그 토대와 관련이 있으면 고양되지만 (enhanced), 관련이 없으면 억제된다(suppressed).

구조 파악 능력을 평가하기 위해 Gernsbacher와 Varner (1988)가 제안한 다중미디어 독해배터리(Multi-Media Comprehension Battery: MMCB)는 간단한 동화를 글로, 음성으로, 그림으로 제공한 후 각각에 대한 문제를 풀게 한다. 사람들은 총 6개의 이야기를 읽거나, 듣거나, 그림으로 본 다음, 각 이야기에 대한 12개의 문제를 푼다. 모든 문제는

단답형인데 이야기의 핵심과 구조를 파악해야 풀 수 있는 문제들로 구성되어 있다. MMCB는 기존의 독해 평가와는 다른 새로운 능력을 측정하는 검사로서 기존 독해평가와의 상관은 낮았다. 또한 MMCB와 SAT의 상관도 낮은데 (Gouravajhala, 2018), 그 이유는 구조 파악 능력은 단순히 글을 표면적으로 이해하는 것 외에도 다양한 인지적 과정을 포함하기 때문이다. 사건의 중요한 단서를 찾는 것뿐만 아니라 단서들을 연결시키고 중요하지 않은 정보들은 제외하는 일련의 과정이 모두 포함된다는 것이다 (McDaniel et al., 2022).

이 검사에서 높은 점수를 받은 사람과 그렇지 않은 사람은 적어도 다음과 같은 다섯 가지의 차이를 보인다. 사람들에게 같은 시간동안 글을 읽게 한 후 기억나는 부분을 쓰게 하면, 구조파악 능력이 높은 사람이 그렇지 않은 사람보다 더 많은 내용을 기억한다. 뿐만 아니라, 구조파악 능력이 높을수록 요점을 더 잘 파악하며, 이전 지식을 더 많이 활성화시키고 적용한다. 구조파악 능력이 높은 사람은 상대적으로 중요하지 않은 내용을 더 잘 억제한다 (McDaniel et al., 2022). 예를 들어, “나는 배를 타고 강을 건넜다.”라는 문장을 읽게 한 다음, 글과 그림이 포개진 자극을 제시하고 그림은 무시하고 글을 읽게 하는 과제를 사용하였다. 구조파악 능력이 낮은 사람은, 어느 정도 시간이 지난 다음, 배경으로 과일 배가 그려졌을 때 함께 제시된 단어를 읽는 시간을 느려졌다. 그 이유는 중의어인 배의 여러 의미가 짧은 시간 동안 모두 활성화되는데 어느 정도 시간이 지나면 맥락에 맞는 운송도구만 활성화되고 다른 의미는 억제되어야 하는데 억제가 충분히 이루어지지 않아 과일 배 그림으로 인해 방해를 받았기 때문이다. 마지막으로, 구조파악 능력이 높은 사람은 원격 유추 검사나 레이븐의 수서행렬 검사와의 상관이 높은데, 이 결과는 이들이 연결을 잘 지음을 시사한다.

그런데 MMCB는 6개의 이야기만으로 구조 파악 능력을 평가한다는 점에서 난도가 한정되어 있다. 그리고 30년 전에 개발된 평가 도구인 만큼 현재 학생들에게 적합한 방식인지도 불분명하다. 이런 맥락에서 본 연구에서는 다양한 난도의 문항을 쉽게 만들 수 있도록 하면서, 구

조 파악 능력을 측정할 수 있는 방안을 모색하고자 하였다. 배열형 문항은 그 자체로 새로운 문항은 아니지만 한편으로는 구성적으로 반응하게 하고 다른 한편으로는 유사성에 따라 부분 점수를 준다는 점에서 새롭다. 이와 관련된 보다 상세한 내용은 다음 절에서 살펴보도록 하겠다.

제 4 절 구성적 문단 배열 문항

본 연구는 출제와 채점이 쉬우면서도 구조 파악 능력을 측정할 수 있는 독해 평가 문항을 제안하고 그 효과를 확인하고자 하였다. 구성적 문단 배열 문항은 여러 문단을 제시한 다음 이들을 하나의 일관된 전체로 배열하도록 하되, 답지를 제시하는 대신, 수험자가 직접 순서를 입력하게 한다. 문단 배열 문항은 기존의 독해 평가에서 그리 많이 사용되지는 않았다. 하지만 이 방식은 수업에 사용된 자료는 물론 다른 관련된 자료를 이용하여 쉽게 만들 수 있어, 교실 평가에서 쉽게 활용될 수 있다. 또한 전반적인 내용은 물론 내용 간 구조적 연결성을 주목하게 함으로써 구조 파악 능력을 평가하는데 유용하다. 채점도 쉬운데, 간단한 채점표를 이용하여 손으로 채점하거나, 본 연구에서처럼 컴퓨터를 이용하여 검사를 실시할 경우 채점 알고리즘을 통해 자동채점도 가능하다. 이런 여러 장점을 고려하면, 구성적 문단 배열 문항은 독해 평가에서 더 광범위하게 활용될 필요가 있다.

배열 문항은 제시된 부분들을 적절히 배열하여 하나의 완성된 전체를 만들도록 하는 점에서 단순화된 직소퍼즐 (jigsaw puzzle) 이라 할 수 있다. 직소퍼즐 방식이 학생들의 독해 능력 향상에 도움이 된다는 연구 결과도 있다 (Neneng & Suherdi, 2017). 이 방식은 수험자가 문항의 요구사항을 쉽게 이해할 수 있을 뿐만 아니라 다양한 방식으로 활용할 수 있다. 예를 들면 다음과 같이 여러 요소를 적절하게 배열하여, 한 문장이 되도록 할 수 있다:

It
(P) reported
(Q) that do not arise readily in actual classroom
(R) spontaneous questions
(S) has long been.

A. PRQS B. QPRS C. RQSP D. SPQR.

문장 수준에서의 배열도 가능하고, 본 연구에서처럼 이를 문단 수준으로 확장할 수도 있다. 하지만 위의 예제에서처럼 선다형으로 반응하게 하면 이 장점을 살리기 어렵다. 예를 들어 4개의 문단을 제시하고 이들을 순서대로 배열하게 하도록 한다고 하자. 만일 이 문항을 우리가 뒤에서 제시하는 것처럼, 구성적 방식으로 직접 답을 쓰게 하면, 가능한 답안의 개수는 4! = 24가지이다. 하지만 선다형에서는 보통 4 ~ 5개의 답지가 제시된다. 따라서 답지를 이용하여 몇 개의 가능성은 쉽게 제거할 수 있고, 한 문단과 다른 문단간의 순서만 제대로 알아도 답지를 이용하여 정답을 골라낼 가능성을 크게 높일 수 있다. 이러한 문제점을 극복하는 방법으로 본 연구에서는 답지를 제시하지 않고, 수험자가 답을 직접 작성하는 구성적 방식으로 제안하였다. 이렇게 하면 답지를 활용한 역행풀이를 막을 수 있다.

반면 구성적으로 반응하게 했을 때의 어려움은 부분점수를 부여하는 방식이다. PQRS가 정답일 때, PQSR 혹은 PRQS 등으로 반응했을 때 과연 얼마만큼의 점수를 부여해야 할까? 본 연구에서는 이를 위해 문자열 유사성 알고리즘을 변형시킨 방법을 제안하였다. 문자열 유사성 알고리즘은 말뭉치 기반(corpus-based) 그리고 지식 기반(knowledge-based) 접근과 함께 텍스트 유사도를 측정하는 방식이다 (Gomma & Fahmy, 2013). 예를 들어 단답형 반응이나 서술형 반응을 채점할 때는, 정해진 답과 비교하면서 동시에 문장 구조를 파악해야 하는 만큼 문장을 비교하

는데 적합한 말뭉치 기반이나 지식 기반의 방식을 선택하면서 동시에 베이지안과 같은 다른 통계 처리 방법을 이용하였다 (예, Leacock & Chodorow, 2003; Mohler & Mihalcea, 2009; Mohler et al., 2011; Sukkarieh et al., 2004). 하지만, 본 연구에서처럼 정해져 있는 답지의 조합만으로 응답을 만들어낼 경우, 단어 수준의 유사성을 확인하는 것으로 충분하다. 따라서 텍스트의 순서와 철자(character)의 구성을 기반으로 유사성을 측정하는 문자열 유사성 알고리즘을 활용하였다 (Gomma & Fahmy, 2013).

대표적인 문자열 유사성 알고리즘으로는 Jaro-Winkler, Damerau-Levenshtein 등이 있는데, 이들은 단어 수준보다는 문장 수준에서 적용되기 때문에 같은 철자로 구성된 다른 문자열을 잘 구별해내지 못한다 (Hall & Dowling, 1980; Winkler, 1990). 실제로, 답지가 4개인 경우 가능한 24가지의 서로 다른 응답을 Damerau-Levenshtein을 활용하면 세 가지로 구분된다. Jaro-Winkler의 경우에는 아홉 가지이지만, 이 중 다섯 개가 .8에서 1미만으로 좁게 분포되어 변별이 잘 이루어지지 않는다. 따라서 본 연구에서는 기존의 문자열 유사성 알고리즘 대신, Lipsky 등 (2010)의 방법을 차용하였다. 이들은 두 인접한 원소를 비교하는 조작을 반복해서 전체를 일정한 규칙으로 배열하는 거품 정렬(bubble sort)을 사용하여 텍스트로 구성된 DNA의 서열 간 유사성을 측정하였다. 거품 정렬은 일반적으로 적용 범위가 넓으며 구현이 간편하다는 장점이 있지만, 시간 복잡도가 $O(n^2)$으로 느려 모든 상황의 최적화된 답이 될 수 없다는 단점이 있다. 하지만 본 연구에서처럼 답지의 개수가 많지 않을 경우, 이 단점이 큰 문제가 되지 않는다. 실제로 이 방법을 사용하면, 기존의 문자열 유사성 알고리즘에 비해 세분화된 부분점수를 제공할 수 있다.

사용된 알고리즘을 간단히 설명하면 다음과 같다. 만약 PQRS가 정답이고 문항 점수가 10점이며, 이동 횟수에 따른 감점이 2점이라 하자. 학생의 반응이 PRQS, PQSR일 경우 정답이 되기 위해서는 1회의 이동이 필요하므로 8점이 된다. 하지만 가장 정답과 다른 SRQP라는 응답은 정

답이 되기 위해 총 6회의 이동이 필요하므로 10-(2*6)으로 -2점이 된다. 이렇게 거품 정렬을 활용하여 채점하면 최대 10점에서 최소 -2점으로, 채점 결과를 세분하여 변별력을 높일 수 있다. 연구 1에서는 한 번의 이동에 2.5점씩 감점하여 최대 10점에서 최소 -5점까지 점수를 부여하였다. 새롭게 제시한 채점 알고리즘을 통해 구성적 방식도 선다형에 못지않게 빠르고 정확하게 채점할 수 있다.

지금까지 무선적으로 제시된 문단을 논리적 흐름에 따라 그 순서를 직접 입력하게 한 다음 그 반응을 정답과 비교하여 부분점수를 매기는 방식이 소개되었다. 이 방식은 Gernsbacher가 주장한 구조 파악과 일맥상통하는 면이 있다. 이 가능성은 연구 3에서 다루었다. 연구 1과 2에서는 각각 소개된 방식이 전통적 선다형에 비해 어떤 차별성을 보이는지와 문단 개수가 난도를 높이는 지를 각각 탐색하였다.

제 3 장 연구 1

연구 1은 독해 능력 중에서 구조 파악 능력에 초점을 맞추었다. 구조 파악 능력이란 글에 대한 심층적인 이해를 바탕으로 시간적 흐름, 비교나 대조 또는 인과적 연결과 같은 형식에 맞춰 정리할 수 있는 능력이다. 구조 파악 능력을 평가하기 위해 문단의 순서를 논리적 연결성에 맞춰 재배치하는 문단 배열 문항을 제안하였다.

본 연구에서 제안하는 문단 배열 문항의 특징은 두 가지이다. 하나는, 선다형이 아니라 수험자가 답을 직접 입력하는 구성적 방식이고, 다른 하나는 정답과의 유사도에 따라 부분 점수를 부여한다는 것이다. 이 방식이 선다형과 비교했을 때 어떤 차이를 보이는 지를 알아보기 위해, 4개의 문단을 논리적 순서에 맞게 배열하는 문항으로 이루어진 두 개의 동형 검사를 사용하였다. 연구 참여자는 먼저 구성적으로 순서를 직접 입력한 다음 이어 제시된 선다형 답지를 고르도록 지시받았다. 각 방식에 대한 두 동형 검사 점수 간 상관이 비교되었는데, 연구 가설은 구성적 문단 배열 문항이 선다형보다 더 높은 상관을 보일 것이라는 것이다.

제 1 절 참여자

참여자 수는 예비 연구에서 사용된 구성적 1, 2차 검사 간의 상관이 .6임을 고려하여, G*power 3.1을 통해 계산하였다 (Faul, Erdfelder, Lang, & Buchner, 2007). 계산된 참여자 수는 최소 50명이었다. 본 연구에서는 서울 소재 종합대학교에 소속된 대학(원)생 68명이 참여하였다. 참여자들의 연령대는 만 30세 이하 대학(원)생으로, 성별은 여자 33명, 남자 35명이었다.

제 2 절 도구

본 연구에는 각각 16개의 문항으로 구성된 두 개의 검사가 사용되었다. 각 문항은 4개의 문단으로 구성되었다. 이 문항들은 대학수학능력시험과 법학적성시험에 출제되었던 글을 이용하여 만들어졌다. 기존의 글을 그대로 활용해 문단 배열 문항으로 만든 경우가 많지만 문항의 난이도를 조절하기 위해 불필요한 접속사를 제거하고 변형시킨 문항도 있다. 60명의 참여자를 대상으로 한 예비연구에서 52개의 예비문항을 사용하였고 검사의 난도를 높이기 위해 정답률이 80% 이상인 문항들을 제거하였다. 예비연구에서 답이 모호하거나 학생들이 이의를 제기한 문항들은 제외하고, 32문항을 본 연구의 검사로 구성하였다. 선다형 방식의 경우 5개의 답지가 제시되었다. 답지는, 예비 연구에서 얻어진 오답빈도가 높은 순서 답안으로 4개를 선택하여 정답과 함께 구성하였다. 본 연구에서 사용한 문단 배열 문항의 예시는 다음과 같다 :

Q. 아래의 4문단을 논리에 맞게 순서대로 배열하시오.

ㄱ. 차선의 문제는 경제학 여러 분야의 논의에서 등장한다. 관세동맹 논의는 차선의 문제에 대한 중요한 사례를 제공하고 있다. 관세 동맹이란 동맹국 사이에 모든 관세를 폐지하고 비동맹국의 상품에 대해서만 관세를 부과하기로 하는 협정이다. 자유무역을 주장하는 이들은 모든 국가에서 관세가 제거된 자유무역을 최적의 상황으로 보았고, 일부 국가들끼리 관세동맹을 맺을 경우는 관세동맹을 맺기 이전에 비해 자유무역의 상황에 근접하는 것이므로, 관세동맹은 항상 세계 경제의 효율성을 증대시킬 것이라고 주장 해왔다.

ㄴ. 가장 효율적인 자원배분 상태, 즉 '파레토 최적'상태를 달성하려면 모든 최적 조건들이 동시에 충족되어야 한다. 파레토 최적 상태를 달성하기 위해 n개의 조건이 충족되어야 하는데, 어떤 이유로 인하여

어떤 하나의 조건이 충족되지 않고 n-1개의 조건이 충족되는 상황이 발생한다면 이 상황이 n-2개의 조건이 충족되는 상황보다 낫다고 생각하기 쉽다.

ㄷ. 이 경우 경제 전체 차원에서 제기되는 문제는 현재 충족되고 있는 일부의 최적 조건들을 계속 유지하는 것이 과연 바람직한가 하는 것이다. 하나의 왜곡을 시정하는 과정에서 새로운 왜곡이 초래되는 것이 일반적 현실이기 때문에, 모든 최적 조건들을 충족시키려고 노력하는 것보다 오히려 최적 조건의 일부가 항상 충족되지 못함을 전제로 하여 그러한 상황에서 가장 바람직한 자원배분을 위한 새로운 조건을 찾아야 한다는 과제가 제시된다. 경제학에서는 이러한 문제를 '차선의 문제'라고 부른다.

ㄹ. 그러나 립시와 랭커스터는 이러한 통념이 반드시 들어맞는 것은 아님을 보였다. 즉 하나 이상의 효율성 조건이 이미 파괴되어 있는 상태에서는 충족되는 효율성 조건의 수가 많아진다고 해서 경제 전체의 효율성이 더 향상된다는 보장이 없다는 것이다. 현실에서는 최적 조건의 일부는 충족되지만 나머지는 충족되지 않고 있는 경우가 일반적이다.

제 3 절 절차

본 연구는 연구자의 감독 하에 화상회의 시스템(http://zoom.us)을 통해 비대면으로 이루어졌다. 본 연구의 실험은 두 단계로 진행하였다. 참여자는 16문항으로 구성된 두 개의 검사를 차례로 30분 동안 풀었다. 참여자는 각 문항에 대해 먼저 구성적으로 응답하고 이어 제시된 답지 가운데 하나를 선택하도록 지시를 받았다. 두 검사 간에는 3분의 휴식시간이 주어지고, 두 검사를 끝낸 다음 3분 정도의 설문이 진행되었다. 설문에는 문항 제시 방법에 따른 체감 난이도를 1 ~ 10점 척도로 응답하게 하

였다.

두 차례의 검사를 실시하는 만큼 응시 순서에 따라 체감 난이도가 달라지거나, 휴식시간 이후 응시하는 검사에만 연습 효과가 발휘될 것을 우려해 역균형화하여 참여자 중 35명은 검사 A를 먼저 응시하게 하였고, 나머지 33명은 검사 B를 먼저 응시하게 하였다. 검사가 종료된 후에 실시된 설문에서는 선다형 문항과 구성적 문항의 체감난이도를 평정하고, 검사를 받으며 느낀 두 방식의 장단점 등도 작성하게 하였다.

제 4 절 결과

분석 전 전체 문항에 대한 Cronbach's alpha를 먼저 확인하여, 음의 상관을 보이는 2개의 문항을 제거하였다. 검사 A와 B에서 각각 1문항을 제거한 후 30개의 문항에 대한 68명의 검사 점수의 평균과 표준 편차를 [표 1]에 제시하였다. 구성적 방식의 동형검사 상관은 .65로 선다형 방식보다 높았다 ($r(66) = .65$, $p < .001$). 같은 문제를 선다형 방식으로 제시한 경우에는 상관이 .43이었다 ($r(66) = .43$, $p < .001$). 두 방식 간의 상관 차이를 Fisher의 Transformation으로 확인했을 때 유의미한 차이를 확인했다 ($z = 1.75$, $p = .039$). 또한 [표 1]에서 볼 수 있듯이 두 방식 모두에서 두 검사의 평균 점수가 유사하다. 이를 통해 검사 A와 B의 난도가 유사한 것을 간접적으로 추정할 수 있다.

[표 1] 검사 A와 B의 평균과 표준편차

	검사 A	검사 B
선다형	84.85 (SD = 22.09)	85.44 (SD = 27.72)
구성형	106.98 (SD = 17.53)	109.59 (SD = 17.40)

[표 2]는 구성적 문항에 부분점수를 부여하지 않고 맞으면 10점 틀리면 0점으로 선다형의 방식대로 분석한 결과이다. 검사 B를 살펴보면 표준편차는 두 방식간의 차이가 없으나 구성적 방식의 평균이 선다형 방식의 평균보다 낮다. 같은 문항이더라도 답지가 주어진 선다형에 비해 직접 답을 적어야 하는 구성적 문항의 난도가 높아졌다. 또한 부분점수를 부여하지 않았더라도 구성적 방식의 상관은 .60으로 선다형의 상관인 .43보다 높았다. 부분 점수를 부여하는 것은 검사의 신뢰도를 높이는 하나의 요인이지만 부분 점수를 없애고 채점하여도 구성적 방식이 선다형 방식에 비해 신뢰도가 높았다.

[표 2] 검사 A와 B의 부분점수 제거 후 평균과 표준편차

	검사 A	검사 B
선다형	84.85 (*SD* = 22.09)	85.44 (*SD* = 27.72)
구성형	72.65 (*SD* = 24.95)	70.44 (*SD* = 27.61)

두 방식 간에 응답 어느 정도로 일치하는지를 알아보기 위해 답안 간의 차이를 비율로 분석하였다. 30문항 중 구성적 방식에서는 틀렸지만 선다형에서 맞춘 문제의 수는 평균적으로 3개 정도로, 비율로 표시하면 9.8%이다. 대략 10문제 중 한 문제 정도를, 선다형의 답지를 보고 추측을 통해 정답을 고른 셈이다. 또한 구성적 방식에서 틀리고 선다형에서만 맞은 문항이 5개 이상인 학생은 68명 중 12명에 해당했다. 틀렸던 문제를 답지를 보고 정답을 고르게 되는 경우가 빈번했음을 보여준다.

반대의 상황도 존재하는데, 구성적 방식에서는 정답을 적었지만 답지를 보고 다시 잘못된 답을 고르는 경우도, 비록 1문항이지만 15명이나 있었다. 많지는 않았지만, 정답을 쓰고서도 선다형의 답지에 현혹되는 경우도 쉽게 볼 수 있는데 이런 현상은 Park (2010)에서도 관찰되었다. 요컨대 수험자는 매력적인 오답지에 끌린다는 것이다.

검사 종료 후 실시한 설문에서는 문항 제시 방법에 따른 체감 난이도

를 10점 척도로 응답하게 하였다. 설문 결과, 참여자들이 평가한 선다형의 체감 난이도의 평균은 4.84점이었다. 반면 구성적 문항의 체감 난이도의 평균은 7.62로 훨씬 높았다. 두 점수 간에는 유의미한 차이가 관찰되었다 ($t(67) = 10.99$, $p < .001$, cohen's $d = 1.60$). 이상의 결과를 통해 같은 문제라도 어떤 방식으로 응답하게 하느냐에 따라 난이도에 차이가 있을 뿐만 아니라, 문항의 신뢰도에서도 차이가 나타난다는 것을 알 수 있다.

제 5 절 논의

문단 배열 문항은 뒤섞여 있는 문단을 논리적 흐름에 따라 재구성하여 전체 글이 일관성을 갖도록 배열하도록 한다. 이 문항은 기존의 독해 문항으로 접근하기 어려웠던 구조 파악 능력을 측정할 수 있게 해준다. 그럼에도 실제 독해평가에서 널리 사용되지 않고 있다. 또한 사용되더라도 선다형으로 제시되어 답지를 보고 정답을 추측해 내는 약점이 있고, 일부의 정확한 배열에 대해 부분점수를 주지 못한다는 한계가 있다. 첫 번째 연구에서는 이런 문제점을 해결하기 위해 구성적으로 반응하게 하고 그 반응을 문자열 유사성 알고리즘으로 채점하여 부분 점수를 부여할 수 있게 하였다. 이 새로운 방식이 전통적 선다형에 비해 문항 배열 능력을 더 안정적으로 측정하는지 알아보았다. 그 결과, 본 연구에서 제안한 새로운 방식이 기존의 선다형 문항에 비해 유의미하게 높은 동형검사 신뢰도를 나타냈다. 이 결과가 시사하는 바는 다음과 같다.

첫 번째는 새롭게 제안한 방식이 기존 선다형에 비해 신뢰도를 높일 수 있다는 것이다. 그 이유는 역행추리가 불가능하여 우연에 의한 편차를 줄이는 한편, 부분 점수를 통해 개인의 역량 편차를 줄였기 때문일 수 있다. 설문조사에서도, 참여자들은 선다형의 답지가 효과적인 단서로 작용했다고 응답했고, 구성적 문항의 체감 난이도가 선다형에 비해 2배 이상 높았다고 보고하였다. 부분 점수를 부여하지 않은 경우에도 선다형

에 비해 구성적 방식의 상관이 높다는 것도 확인했다. 따라서, 문단 배열 문항의 경우 선다형을 구성적 방식으로 대체할 필요가 있다.

두 번째는 단순히 기존의 선다형을 구성적 방식으로 대체하는 것 이상으로 부분 점수를 합리적으로 부여하여 평가의 정확성을 향상시켰다는 점이다. 본 연구에서 활용한 채점 방식은 기존의 선다형에서처럼 맞으면 10점 틀리면 0점이 아니라 일부 맞는 부분에 대해서는 부분 점수가 부여되었다. 부분점수가 부여될 경우, 그렇지 않을 때에 비해, 중도에 포기하지 않고 할 수 있는 만큼 최선을 다해 평가에 임할 가능성이 높다. 그러나 기존의 독해 평가에서 부분점수를 주는 방식이 잘 사용되지 않았던 이유는 부분 점수를 부여할 명확한 기준을 마련하는 것이 어렵고 채점도 오래 걸리기 때문이다. 이를 보완해 본 연구에서는 거품정렬 방식을 변형하여 문단의 순서 부분마다 점수를 부여하는 방안을 새롭게 제안하였다. 또한 컴퓨터를 활용해 채점도 자동화하여 정확한 평가뿐만 아니라 빠른 채점도 가능하게 되었다는 점에서 의미가 있다.

위의 두 가지로부터 도출되는 세 번째 의의는 문단 배열 문항을 독해 평가에 더 많이 사용할 근거를 마련했다는 것이다. 기존 독해 문항에 대한 비판 중 하나는 글을 다 읽지 않고 지엽적인 정보를 찾으면 맞출 수 있는 문제가 많다는 것이었다. 이 문제를 해결하는 방법은 구조 파악 능력을 측정할 수 있는 문단 배열 문항의 수를 늘리는 것이겠다. 다행히도, 문단 배열 문항을 만드는 일은 어렵지 않다. 수업에서 사용된 내용과 관련된 자료 중 일부분을 골라 접속사나 그밖에 전후 관계를 쉽게 파악할 수 있게 하는 표현을 삭제하거나 바꾸면 만들 수 있기 때문이다. 따라서 구성적 문단 배열 문항은 교사들이 교실 평가의 질을 높이는데 사용될 수도 있겠다.

마지막으로 구성적 문단 배열 문항은 Dray 등의 구성개념도 접근이나 Sarbatini 등이 개발한 SBA에서도 사용될 수 있다. Dray 등의 접근에서는 특히 텍스트의 구조를 해석하는 단계를 평가하는데 효율적일 수 있다. SBA에서는 읽기 능력 차원의 모형 구성 능력 범주 중 조직화의 하위 범주를 측정하는데 유용해 보인다. 결론적으로 더 탄탄한 이론적 배경을

가진 독해 평가에서도 문단 배열 문항이 효과적으로 사용될 여지가 충분하다는 것이다.

제 4 장 연구 2

문단 배열 문항의 난도를 높이는 방법은 어려운 글을 사용해 문항을 만드는 방법과 배열해야 할 문단의 개수를 늘리는 방법 두 가지가 있다. 연구 2에서는 문단의 개수를 늘리는 방안을 활용하여 실제로 문단의 개수를 증가시켰을 때 검사들 간의 상관이 어떻게 달라지는지 확인하였다. 구성적 문단 배열 문항을 4문단으로 구성한 검사 C와 5문단으로 구성한 검사 D로 나누어 연달아 응시하게 한 뒤에 두 검사 간의 상관을 확인하였다.

4문단 배열 문항은 16개를 30분 동안 풀고 5문단 배열 문항은 8개를 20분 동안 푼다. 5문단 배열 문항의 경우 개수도 4문단의 절반이고 풀이 시간도 줄어든다. 그럼에도 4문단 배열 문항 검사와 높은 상관을 보인다면 5문단 배열 문항이 더 효율적으로 수험자를 평가하는 문항으로 볼 수 있다. 또한 문단의 개수를 다르게 하여 검사의 난도를 조절할 수 있다는 것을 확인할 수 있을 것이다.

제 1 절 참여자

참여자 수는 예비 연구에서 검사 간의 상관이 .66을 고려하여, G*power 3.1을 통해 계산하였다. 계산된 참여자 수는 최소 58명이었다. 본 연구에서는 서울 소재 종합대학교에 소속된 만 30세 이하 대학(원)생 71명이 참여하였다.

실험 자극과 참여자 수를 확정하기 위해 실시한 예비연구에도 서울 소재 종합대학교에 소속된 만 30세 이하 대학(원)생 37명이 참여하였다.

제 2 절 도구

본 연구에는 4문단을 배열하는 문항 16개로 구성된 검사 C와 5문단 배열 문항 8개로 구성된 검사 D가 사용되었다. 검사 C의 경우 연구 1에서 사용한 문항들 중 정답률이 80% 이하인 적절한 난도의 문항을 선별하여 재구성하였다. 검사 D의 경우 우선 연구 1에서 사용하지 않은 새로운 글을 변형시킨 20개의 5문단 배열 문항을 예비연구를 통해 난이도를 확인하였다. 이후 적절한 난도의 문항을 8개 선별하여 검사 D로 구성하였다. 연구 2에서는 참여자들에게 문단 배열 문항을 구성적 방식으로만 응답하게 하였다.

제 3 절 절차

본 연구는 연구자의 감독 하에 화상회의 시스템을 통해 비대면으로 이루어졌다. 본 연구의 실험은 두 차례의 검사로 구성되어 있다. 검사 C는 4문단 배열 문항이 16문항으로 구성되어있고 시간은 30분이 주어진다. 검사 D는 5문단 배열 문항이 8문항으로 구성되어 있으며 20분 간 진행된다. 두 검사 모두 구성적 방식으로 제시되며 참여자들은 컴퓨터로 검사에 응시한다. 두 검사 사이에는 3분의 휴식시간이 주어진다.

두 차례의 검사를 실시하는 만큼 응시 순서에 따라 체감 난이도가 달라지거나, 두 번째 검사에만 연습 효과가 발휘될 것을 우려해 역균형화하여 참여자 중 36명은 검사 C를 먼저 응시하게 하였고, 35명은 검사 D를 먼저 응시하게 하였다. 검사에 앞서 참여자들은 문단 배열 문항에 대한 간단한 설명을 듣고 모든 문항에 구성적 방식으로 응답하도록 지시받았다. 또한 문항마다 부분점수가 부여된다는 점을 참여자들에게 설명하고 모든 문항에 성실히 응답할 것을 요청하였다.

제 4 절 결과

두 검사에 대한 71명의 검사 점수의 평균과 문항 평균이 [표 3]에 제시되었다. 연구 2에서 부분 점수를 부여하는 방식은 연구 1과 같고 감점 단위는 2점으로 검사 C의 경우 한 문항 당 최대 10점에서 최소 -2점까지, 검사 D의 경우 한 문항 당 최대 10점에서 최소 -6점까지 부여된다. 검사 C의 총점은 160점이고 검사 D의 총점은 80점이다. 두 검사는 문항 수가 달라 총점의 차이가 있어 평균을 직접적으로 비교하지 않고 문항 평균을 비교하였다. 한 문항 당 총점이 10점일 때 검사 C의 문항 평균점수는 7.07이고 검사 D의 문항 평균 점수는 4.46이다. 검사 D의 문항 평균이 낮은 것을 통해 5문단 배열 문항이 수험자들에게 더 어려운 검사였다는 것을 추정할 수 있다. 또한 문항 평균의 표준편차를 보더라도 검사 D의 표준편차가 큰 것으로 보아 검사 D가 문항 수 자체는 적지만 학생들의 실력을 변별하기에 더 효과적인 검사로 보인다.

[표 3] 검사 C와 D의 평균과 문항 평균

	검사 C	검사 D
평균	113.21 (SD = 15.70)	35.69 (SD = 13.02)
문항 평균	7.07 (SD = 1.20)	4.46 (SD = 2.10)

검사 C와 D의 상관은 .70으로 연구 1에서의 구성적 방식보다 높은 것을 확인할 수 있었다 ($r(69) = .70$, $p < .001$). 두 검사는 같은 문단 배열 문항으로 구성되어 있으나 몇 개의 문단을 배열하는지에 차이가 있다. 검사 C는 4개의 문단을 배열하는 문항을 16개 푸는 것이고 검사 D는 5개의 문단을 배열하는 대신 검사 C의 반절인 8개의 문항을 푼다. 결과적으로 문항수와 검사 시간은 줄었지만 두 검사의 상관이 높게 나오는 것을 통해 검사 D가 검사 C만큼 수험자의 능력을 변별한다고 볼 수 있다.

제 5 절 논의

문단 배열 문항의 요소 중 하나인 문단의 개수를 조정하여 두 가지 검사의 상관을 확인한 연구 2는 문단 배열 문항을 좀 더 효율적으로 사용할 수 있는 방안을 제시하였다. 연구 2의 시사점은 다음과 같다.

첫 번째로 두 검사 간의 높은 상관으로 좀 더 효율적인 독해 평가 문항을 제안했다는 점이다. 평가 장면에서 정확한 평가 문항을 사용하는 것도 중요하지만 그만큼 비용을 줄이고 효율적으로 평가하는 것도 중요하다. 연구 1에서 사용한 두 개의 검사는 문항의 형식이 똑같은 동형검사로 두 검사의 신뢰도를 확인했을 때 구성적 방식에서 .65의 신뢰도를 확인할 수 있었다. 그런데 연구 2에서는 문항의 형식은 동일하지만 배열할 문단의 개수에 차이를 두어 검사를 구성했는데 두 검사 간의 더 높은 상관을 확인할 수 있었다. 4문단을 배열해야하는 검사 C의 경우 16개의 문항을 풀어야하기 때문에 30분의 시간이 소요되지만 검사 D의 경우 5문단 배열 문항을 사용한 대신 전체 문항의 수를 반으로 줄여 검사 시간도 20분으로 줄였다. 결과적으로 검사 D가 더 짧은 시간으로도 수험자를 정확하게 평가할 수 있는 효율적인 방안이라는 것이다.

두 번째로 5문단 배열 문항이 더 변별력이 높은 평가 문항이라는 것을 확인할 수 있었다. 적은 문항을 사용해 효과적으로 수험자를 평가하기 위해서는 적절한 고난도의 문항을 개발하는 것이 중요하다. 연구 2의 결과를 보면 문항 평균이나 표준편차만 보더라도 수험자들이 검사 D를 더 어려워한다는 것을 알 수 있는데, 그 이유는 경우의 수 때문이다. 4개의 문단을 배열하는 경우의 수는 4! = 24 가지이지만 5개의 문단을 배열하는 경우의 수는 5! = 120 가지로 선택지가 5배 늘어난다. 결국 수험자는 더 많은 선택지들 속에서 논리적으로 문단을 배열해야하는 것이다. 경우의 수가 많아질수록 우연에 의해 정답을 고르는 확률이 낮아지기 때문에 정확한 평가에 기여할 수 있다.

마지막으로 문단 배열 문항의 특성 중 하나인 문단 개수를 조절함으로써 문항의 난이도를 쉽게 조정할 수 있다는 점이다. 독해 평가를 다양한

연령대에 사용하기 위해서는 난이도 조절이 수월해야 한다. 그런 점에서 구성적 문단 배열 문항의 장점 중 하나는 여러 난도로 변형하여 다양한 연령대에 사용 가능한 문항이라는 점이다. MMCB도 오랫동안 사용되어 온 유용한 검사이지만 난도가 한정되어있다는 점에서 사용할 수 있는 연령대가 한정되어 있다. 그러나 문단 배열 문항의 경우 사용하는 글의 난도를 낮추거나 문단의 개수를 줄이면 초등생에게도 사용할 수 있다. 반대로 같은 글을 활용하더라도 배열해야하는 문단의 개수를 늘리면 난도가 높아지게 된다. 문단의 개수와 글의 난이도를 조절함으로써 초중고생뿐만 아니라 성인들에게도 적절한 난도의 문항을 제공할 수 있다는 점이 문단 배열 문항의 큰 강점이 될 것이다.

제 5 장 연구 3

연구 3은 문단 배열 문항이 실제로 구조 파악 능력을 측정할 수 있는지 확인하고자 수행되었다. 일련의 연구들이 독해 능력 중에서도 구조 파악 능력에 초점을 맞추었기 때문에 문단 배열 문항이 해당 능력을 측정할 수 있는 도구인지를 확인해야 한다. 그래서 구조 파악 능력을 측정하기 위해 사용되어온 검사인 MMCB의 일부 문항을 사용해 5문단 배열 문항과의 상관을 확인하고자 한다. MMCB는 30년 동안 구조 파악 능력을 측정하기 위해 사용되어온 도구로 글, 음성, 그림 문항으로 구성되어 있다. 이 중 연구 3에서는 그림 문항을 사용하여 5문단 배열 문항과의 상관을 확인하였다.

제 1 절 참여자

본 연구의 참여자는 서울 소재 종합대학교에 소속된 만 30세 이하 대학(원)생 42명이 참여하였다. 참여자들의 성별은 남자15명, 여자 27명이다.

제 2 절 도구

본 연구는 MMCB 중 일부 문항을 사용하였다. MMCB는 간단한 동화를 글로, 음성으로, 그림으로 제공한 후 각각에 대한 문제를 푸는 검사인데 Gernsbacher (1990)에서는 세 평가 방식 간의 높은 신뢰도를 확인하였다. 본 연구에서는 MMCB의 문항 중 그림으로 제공되는 문항만 사용하였다.

MMCB를 활용하는 연구를 진행한 Maki 등 (2002)과 McDaniel 등 (2002)에서도 MMCB 중 한 방식의 검사만을 사용해 구조 파악 능력을 측정하였다. MMCB의 그림 문제들은 두 개의 이야기로 나눠져 있는데 한 이야기 당 32개의 이미지로 구성되어 있다. 참여자들은 32개의 이미지를 순서대로 보면서 이야기의 흐름을 파악해야한다. 그리고 한 이야기마다 12개의 단답형 문항이 주어진다. 한 문항 당 3점을 부여하는데 부분점수를 부여하는 기준이 정해져있어 기준에 맞춰 채점이 이루어진다. 본 연구에서 사용한 MMCB 검사의 총점은 72점이다.

이후 5문단 배열 문항 검사로 40분 동안 16문항을 풀게 된다. 참여자들은 문단 배열 문항을 모두 구성적 방식으로 응답한다. 검사에 사용된 문항들은 연구 2 예비연구에서 적절한 난도를 확인한 문항들로 구성하여 사용하였다. 한 문항 당 최대 10점에서 최소 -6점까지 부여되며 총점은 160점이다.

제 3 절 절차

본 연구는 연구자의 감독 하에 화상회의 시스템(http://zoom.us)을 통해 비대면으로 이루어졌다. 본 연구의 실험은 두 차례의 검사로 구성되어 있다. 첫 번째 검사는 MMCB의 그림 검사로 한 이야기가 32개의 이미지로 이뤄져 있다. 참여자들에게 동시에 같은 시간 동안 이미지를 보여주기 위해 화상회의 시스템 하에서 화면공유로 이미지를 보여준다. 한 이미지 당 5초씩 노출되는 슬라이드 쇼를 사용해 자동으로 다음 이미지로 넘어가게 된다. 슬라이드 쇼 시작 전 모든 참여자들에게 공유 화면이 잘 보이는지 확인하였고 이미지들은 하나의 이야기를 이루고 있으며 딱 한 번만 보여주기 때문에 집중해서 이미지를 볼 수 있도록 안내하였다. 참여자들이 한 이야기의 이미지를 다 본 후에는 해당 이야기에 대한 문제를 풀게 된다. 문제는 12문항으로 구성되어 있고 모두 단답식 혹은 짧은 서술형 문항이다. 모든 참여자들이 첫 번째 이야기에 대한 문제를 다 풀

었을 때 다음 이야기로 넘어 간다. 다음 이야기 역시 화면공유로 이미지 슬라이드 쇼를 보여준 이후 바로 문제를 푼다.

MMCB 문제를 풀기 전 참여자들에게 몇 가지 안내가 이뤄지는데, 우선 모든 문제에 응답해야한다는 점이다. 만약 답안이 기억나지 않을 경우에는 '기억나지 않음'이라고 적고 넘어가게 하였다. 또한 모든 문제에는 부분점수가 부여되기 때문에 최대한 본인의 답안을 상세히 적을 것을 요청하였다. MMCB 검사가 종료된 후에 바로 5문단 배열 문항을 풀게 된다. 16개의 구성적 문단 배열 문항으로 구성된 이 검사는 40분 동안 진행되며 제출 후에 연구가 종료된다.

제 4 절 결과

MMCB의 경우 구체적인 채점 기준이 정해져있기 때문에 채점 기준에 맞춰 부분점수를 부여하는 방식으로 채점하였다. 구성적 문단 배열 문항의 경우 앞선 연구들에서 사용했던 채점 알고리즘을 활용하여 채점하였다. 두 검사 간의 상관계수는 .65로 높은 상관을 확인했다 ($r(40) = .65$, $p < .001$). [표 4]에서는 MMCB 검사와 배열형 문항의 참여자들 평균과 표준편차를 볼 수 있다. MMCB 그림 검사는 두 개의 이야기로 구성되어 있고 두 점수를 합하여 평균을 계산하였다.

[표 4] MMCB와 문단 배열 문항의 평균과 표준편차

	MMCB	문단 배열 문항
평균 및 표준편차	19.30 (SD = 7.01)	93.47 (SD = 21.32)

제 5 절 논의

앞선 연구 1,2에서 문단 배열 문항의 신뢰도를 확인한 후 실제로 문단 배열 문항이 구조 파악 능력을 측정할 수 있는지를 연구 3에서 확인하였다. 연구 3의 시사점은 아래와 같다.

첫 번째로 본 연구에서 제안한 문단 배열 문항이 구조 파악 능력을 측정할 수 있는 문항이라는 것을 확인했다. 30여 년 동안 구조 파악 능력 측정을 위해 사용된 MMCB의 일부 문항과 문단 배열 문항 간의 상관을 확인해 본 결과 높은 상관계수를 확인했다. Gouravajhala (2018)의 연구에서는 MMCB와 SAT의 낮은 상관을 밝혔다. 또한 Maki 등 (1994)의 연구는 Science news에서 발췌한 글로 평가 문항을 만든 검사와 MMCB 일부 문항들 사이에서 .53의 상관을, Nelson-Denny Reading Test (Brown, Nelson, & Denny, 1973) 와 MMCB는 .45의 상관을 확인했다. Maki의 연구에서는 관련 연구들 중 가장 높은 상관계수를 확인했지만 계속해서 기존 독해 평가 문항으로는 구조 파악 능력을 측정하기 어렵다는 주장이 나왔다 (Gouravajhala, 2018). 문단 배열 문항의 경우에는 앞선 두 검사보다도 MMCB와의 상관이 높은 것으로 보아 기존의 독해 검사들보다도 구조 파악 능력을 측정하기에 적합한 문항이라고 볼 수 있다

두 번째로 문단 배열 문항은 구조 파악 능력을 측정할 수 있을 뿐 아니라 능력을 향상시키기 위한 훈련으로도 적합한 문항이라는 점이다. 앞서 서론에서 밝혔듯이 평가가 중요한 이유는 학생들이 평가 방식에 맞춰 학습하기 때문이다. 기존의 독해 문항들은 지엽적인 정보에 초점을 맞춰 구조 파악 능력을 측정하기 어려웠다. 그래서 학생들이 독해 학습을 하더라도 구조 파악 능력에 대해 알지 못하거나 훈련할 기회가 없었다. 그러나 문단 배열 문항으로 평가할 경우 학생들은 구조 파악 능력의 중요성을 알게 되어 새로운 학습의 방향을 세울 수 있다. 또한 문단 배열 문항을 꾸준히 푸는 것만으로도 구조 파악 능력을 기르는 훈련이 될 수 있다.

마지막으로 구조 파악 능력을 측정하는 방식에 대해 형식적, 이론적

측면으로 기여했다는 점이다. MMCB의 경우 인물, 사건, 배경을 파악하는 이야기 형식의 글로 구조 파악 능력을 측정한 검사이지만 문단 배열 문항의 경우 정보를 전달하는 글을 변형하여 만든 문항이다. 구조 파악 능력을 측정하는 방식에 있어 이야기 형식의 한 장르로만 평가하는 것보다 정보성 글을 활용한 평가 문항이 추가된다면 더 정확하게 평가될 수 있을 것으로 기대된다. 결과적으로 문단 배열 문항은 MMCB와는 다른 구조 파악 능력 측정 방안이 될 뿐만 아니라 독해 능력을 평가할 수 있는 효율적인 문항이 될 것으로 기대된다.

제 6 장 종합 논의

정보에의 접근이 그 어느 때보다 쉬워지면서 독해의 중요성은 그만큼 커지고 있다. 그럼에도 불구하고, 스마트 폰에의 과몰입으로 인해, 독서 시간은 줄고 결과적으로 독해 능력이 쇠퇴하고 있다. 이런 배경에서 본 연구는 새로운 독해 평가 방식을 제안하고 그 신뢰도와 타당도를 탐색하였다.

본 연구에서는 독해의 핵심을 구조 파악 능력으로 보았다. 구조란 글을 이루는 여러 부분들이 응집력 있게 묶인 표상을 가리킨다. 이 구조는 예를 들어 시간적 흐름, 비교나 대조, 혹은 인과적 논리적 연결 혹은 문제 제기와 해결과 같은 형식을 취할 수 있다. 구조 파악 능력을 측정하는 방안으로 원 글의 문단을 임의로 배열하여 제시하고, 올바른 순서로 배열하게 하였다. 이런 문단 배열 문항은 이전에도 선다형으로 사용되었다. 하지만 본 연구에서는 두 가지 차이가 있었다. 하나는 가능한 순서를 답지로 제시하는 대신, 직접 입력하도록 한 것이고, 다른 하나는 입력된 순서를 정답과 비교하여 부분점수를 부여하였다. 직접 입력하도록 한 이유는 답지를 보고 풀거나 아니면 무선 추측으로 맞추지 못하게 하기 위해서였다. 부분점수는 여러 문자열 유사성 알고리즘 중 차이를 가장 크게 만드는 하는 방법으로 선정하였다.

이 새로운 독해 평가 문항의 심리측정학적 특징을 알아보기 위해, 세 개의 연구가 수행되었다. 연구 1에서는 연구 참여자를 대상으로 4개의 문단으로 이루어진 문항으로 구성된 두 개의 동형 검사를 실시하였다. 각 문항은 구성적 방식과 선다형 방식으로 제시하였다. 즉, 같은 발문에 대해 먼저 구성적으로 그 다음에는 선다형으로 반응하게 하였다. 그 결과 구성적 방식의 두 검사 점수 간 상관이 두 선다형 점수 간의 상관보다 유의미하게 높음을 확인하였다. 이 결과는 새로운 문단 배열 방식이 전통적 선다형에 비해 수험자의 구조 파악 능력을 안정적으로 측정함을 시사한다.

연구 2에서는 수험자에게 4문단으로 구성된 검사와 5문단으로 구성된 더 적은 문항의 검사를 동시에 실시하였다. 난도를 높이고 문항수를 줄여도 두 검사 간 상관이 높게 나타나는지를 탐색하였다. 그 결과 두 검사 간 상관이 높게 나타났는데, 이 결과는 문단 개수로 난도를 높이고 문항수를 줄여 전체적으로 구조 파악 능력을 효율적으로 측정할 수 있음을 시사한다. 요컨대 사용되는 문단의 개수로 구조 파악 능력 평가의 난도를 쉽게 조절할 수 있다는 것이다.

연구 3에서는 문단 배열 문항과 종래의 구조 파악 능력 측정 도구인 MMCB와의 동시 타당도를 확인하였다. MMCB의 세 양태 중 그림을 이용한 자극으로부터의 점수와 5개의 문단 배열 문항으로 이루어진 검사 간에 높은 상관이 발견되었다. 이 결과는 문단 배열 문항 검사와 MMCB 검사가 같은 능력을 측정할 가능성을 시사한다. 이 결과는 또한 MMCB를 다양한 언어권이나 연령대에 사용하는 데 있어서의 어려움을 해결하는데 도움을 줄 가능성을 시사한다.

이상의 결과는 새로운 문단 배열 문항이 구조 파악 능력을 비교적 잘 측정함을 보여준다. 문단 배열 문항은 언어에 상관없이 다양한 나라에서 활용될 수 있으며 학교 현장에서 교사도 쉽게 출제할 수 있다는 강점이 있다. 또한 이야기 형식의 글만을 가지고 평가한 MMCB와는 다르게 정보를 전달하는 글의 형식으로 구조 파악 능력을 측정하는 문단 배열 문항은 기존의 평가를 보완하며 정확하게 구조 파악 능력을 측정할 수 있게 해줄 것으로 기대된다.

이런 여러 장점에도 불구하고, 본 연구는 다음과 같은 한계점이 있다. 우선, 한 대학의 대학생을 대상으로만 연구를 진행했다는 점이다. 후속 연구에서는 더욱 다양한 글로부터 추출된 문단을 사용하면서, 다양한 연령대를 대상으로 그 사용가능성이 검증될 필요가 있다. 또 다른 한계는 문단 배열 문항만을 이용하여 연구가 진행되었다는 점이다. 다른 문항들과 섞여 있는 상황에서도 본 연구에서와 같은 결과가 관찰되는지 확인할 필요가 있다.

이런 한계를 극복하는 연구와 함께 새로운 문단 배열 문항을 풀 때 일

어나는 인지 과정을 탐색할 필요가 있다. 이미 구조 파악 능력을 측정할 수 있다는 것을 확인했지만 이외에도 문단 배열 문항이 어떤 능력들을 평가할 수 있는지를 알아볼 수 있다. 이를 위해 독해 연구에서 많이 사용되는 프로토콜 분석을 이용하여, 문단 배열 문항을 풀 때 어떤 과정이 일어나며 언제 어떤 어려움을 겪는지를 알아본다면 좀 더 고도화된 문항을 개발하는데 도움이 될 것이다. 또한 수험자들이 독해 과정에서 겪는 어려움을 확인하여 극복 방안을 모색할 수 있겠다.

이상의 여러 후속 연구를 통해 이 방식을 더욱 효율적으로 사용할 수 있는 방법을 찾을 수 있을 것으로 예상한다. 우리의 궁극적인 바람은 구성적 문단 배열 문항이 여러 교육 현장에서 정확한 독해 평가의 한 도구로 손쉽게 활용되는 것이다.

참 고 문 헌

민경석, & 박주용. (2009). 구성적 선다형 검사에서 선다형과 단답형의 문항 특성 비교. **교육평가연구, 22(2)**, 451-469.

박주용. (2010). 구성적 선다형 검사 방식에서 배점 조작의 효과. **한국심리학회지: 인지 및 생물, 22(4)**, 573-588.

Arnold, K. M., Eliseev, E. D., Stone, A. R., McDaniel, M. A., & Marsh, E. J. (2021). Two routes to the same place: learning from quick closed-book essays versus open-book essays. *Journal of Cognitive Psychology*, *33*(3), 229-246.

Bogaerds-Hazenberg, S. T., Evers-Vermeul, J., & van den Bergh, H. (2021). A meta-analysis on the effects of text structure instruction on reading comprehension in the upper elementary grades. *Reading Research Quarterly, 56(3)*, 435-462.

Brown, J. I. (1973). The Nelson-Denny reading test–form C and D. *Boston: HoughtonMifflinCo.*

Butler, A. C., & Roediger III, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, *19*(4-5), 514-527.

Callender, A. A., & McDaniel, M. A. (2009). The limited benefits of rereading educational texts. *Contemporary Educational Psychology*, *34*(1), 30-41.

Cartwright, K. B., Lee, S. A., Taboada Barber, A., DeWyngaert, L. U., Lane, A. B., & Singleton, T. (2020). Contributions of executive function and cognitive intrinsic motivation to university students' reading comprehension. *Reading Research Quarterly*, *55*(3), 345-369.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, *52*(3), 219-232.

Dray, A. J., Brown, N. J., Diakow, R., Lee, Y., & Wilson, M. R. (2019).

A construct modeling approach to the assessment of reading comprehension for adolescent readers. *Reading Psychology*, *40*(2), 191-241.

Espin, C. A., & Foegen, A. (1996). Validity of general outcome measures for predicting secondary students' performance on content-area tasks. *Exceptional Children*, *62*(6), 497-514.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.

Gernsbacher, M. A., & Varner, K. R. (1988). *The multi-media comprehension battery* (No. 88-07). Tech. Rep.

Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 430.

Gernsbacher, M. A. (1997). Two decades of structure building. *Discourse processes*, *23*(3), 265-304.

Gomaa, W. H., & Fahmy, A. A. (2017, December). SimAll: A flexible tool for text similarity. In *The Seventeenth Conference on Language Engineering ESOLEC* (Vol. 17, pp. 122-127).

Hall, P. A., & Dowling, G. R. (1980). Approximate string matching. *ACM Computing Surveys (CSUR)*, 12(4), 381-402.

Jensen, K. L., & Elbro, C. (2022). Clozing in on reading comprehension: a deep cloze test of global inference making. *Reading and Writing*, 1-17.

Kendeou, P., & O'Brien, E. J. (2017). Reading comprehension theories:

A view from the top down. In *The Routledge handbook of discourse processes* (pp. 7-21). Routledge.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological Review*, *95*(2), 163.

Leslie, L., & Caldwell, J. S. (2017). Assessments of reading comprehension: Challenges and directions. *Handbook of research on reading comprehension*, 219-271.

Lillis, T. (2019). 'Academic literacies': sustaining a critical space on writing in academia. *Journal of Learning Development in Higher Education*, (15).

Lin, C., McDaniel, M. A., & Miyatsu, T. (2018). Effects of flashcards on learning authentic materials: The role of detailed versus conceptual flashcards and individual differences in structure-building ability. *Journal of Applied Research in Memory and Cognition*, *7*(4), 529-539.

Lipsky, O., Porat, B., Porat, E., Shalom, B. R., & Tzur, A. (2010). String matching with up to k swaps and mismatches. *Information and Computation*, *208*(9), 1020-1030.

MacGinitie, W. H., MacGinitie, R., Mria, K., & Dreyer, L. G. (2000). *Gates-MacGinitie reading tests: Directions for administration*. Riverside Publishing.

MacGinitie, W. (2002). Gates-MacGinitie Reading Comprehension Test.

Maki, W. S., & Maki, R. H. (2002). Multimedia comprehension skill predicts differential outcomes of web-based and lecture courses. *Journal of Experimental Psychology: Applied*, *8*(2), 85.

McDaniel, M., Marsh, E., & Gouravajhala, R. (2022). Individual Differences in Structure Building: Impacts on Comprehension and Learning, Theoretical Underpinnings, and Support for Less Able Structure Builders. *Perspectives on Psychological Science*, 17(2),

385-406.

McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychology of Learning and Motivation*, *51*, 297-384.

Namaziandost, E., Gilakjani, A., & Hidayatullah. (2020). Enhancing pre-intermediate EFL learners' reading comprehension through the use of Jigsaw technique. *Cogent Arts & Humanities*, 7(1), 1738833.

Neneng, H., & Suherdi, D. (2017). The Effectiveness of Jigsaw in Improving Students'Reading Comprehension. *Journal of English and Education*, *5*(1), 1-12.

Nurwanti, N., Asrifan, A., & Haedar, H. (2019). THE APPLICATION OF COOPERATIVE LEARNING: JIGSAW II TECHNIQUE IN IMPROVING STUDENTS'READING COMPREHENSION OF EXPOSITORY TEXT. *Journal of Advanced English Studies*, *2*(1), 31-40.

Park, J. (2005). Learning in a New Computerized Testing System. *Journal of Educational Psychology*, *97*(3), 436.

Park, J. (2010). Constructive multiple-choice testing system. *British Journal of Educational Technology*, *41*(6), 1054-1064.

Rivers, M. L., Dunlosky, J., & Persky, A. M. (2020). Measuring metacognitive knowledge, monitoring, and control in the pharmacy classroom and experiential settings. *American Journal of Pharmaceutical Education*, *84*(5).

Roehling, J. V., Hebert, M., Nelson, J. R., & Bohaty, J. J. (2017). Text structure strategies for improving expository reading comprehension. *The Reading Teacher*, *71*(1), 71-82.

Rupp, D. E., Ganapathi, J., Aguilera, R. V., & Williams, C. A. (2006). Employee reactions to corporate social responsibility: An organizational justice framework. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and*

Organizational Psychology and Behavior, *27*(4), 537-543.

Sabatini, J., O'Reilly, T., Weeks, J., & Wang, Z. (2020). Engineering a twenty-first century reading comprehension assessment system utilizing scenario-based assessment techniques. *International Journal of Testing*, *20*(1), 1-23.

Sheehan, K., & O'Reilly, T. (2012). The case for scenario-based assessments of reading competency. *Reaching an Understanding: Innovations in How We View Reading Assessment*, 19-33.

Traxler, M. J., & Gernsbacher, M. A. (1995). Improving coherence in written communication. *Coherence in Spontaneous Text*, 215-238.

Winkler, W. E. (1990). String comparator metrics and enhanced decision rules in the Fellegi-Sunter model of record linkage.

Woolley, G. (n.d.). Reading Comprehension Assisting Children with Learning Difficulties / (1st ed. 2011. ed.). Dordrecht :: Springer Netherlands :.

Van Boekel, M., Lassonde, K. A., O'Brien, E. J., & Kendeou, P. (2017). Source credibility and the processing of refutation texts. *Memory & Cognition, 45(1)*, 168-181.

Van den Broek, P., & Kendeou, P. (2022). Reading Comprehension I: Discourse. *The Science of Reading: A Handbook*, 239-260.

Abstract

Reliability and Validity of Constructive Paragraph Rearrangement Items

Hwimin Kim
Department of Psychology
The Graduate School of Social Science
Seoul National University

One of the shortcomings of most reading comprehension tests is the abundance of items that ask for trivial details. In order to provide a substitute for these items, this study focused on the global structure of the text and introduced modified paragraph rearrangement items. In paragraph rearrangement items, the student's task is to rearrange four or five randomly arranged paragraphs in a logical flow. I revised traditional paragraph rearrangement items as follows, calling them constructive paragraph rearrangement items: i) Instead of choosing an option in the multiple choice format, students have to write down the correct order; ii) their responses are scored using a string similarity algorithm. Three studies were conducted on undergraduate and graduate students to confirm the reliability and validity of the constructive paragraph rearrangement items.

In Study 1, the equivalent-form reliability of the constructive paragraph rearrangement items was compared with the multiple-choice format items. The result showed that the correlation between the two tests in the constructive paragraph rearrangement format was significantly higher than that in the multiple-choice format. This result suggests that the constructive paragraph arrangement test measures test takers' structure comprehension ability more reliably compared to the multiple choice format test.

In Study 2, a high correlation was found between a 16 item test that asks to rearrange 4 paragraphs and a 8 item test composed of 5 paragraphs. The times allowed for the two tests were 30 minutes and 20 minutes, respectively. This result suggests that by increasing the number of paragraphs, which raises the difficulty of the test, we can create a more efficient test. In Study 3, we examined whether constructive paragraph rearrangement items actually measure structure building ability. To test this hypothesis, Multi-Media Comprehension Battery(MMCB), a structure building ability test developed by Gernsbacher and Varner (1988), was used. Participants were instructed to take both the constructive paragraph rearrangement test and the MMCB picture subtest. The result showed a high correlation between the two tests. This suggests that structure building ability can be assessed through constructive paragraph rearrangement items.

The above results show that constructive paragraph rearrangement items, which are easily created using relevant text, can be an effective tool that brings new vitality into the reading comprehension assessment field.

keywords : constructive format, paragraph rearrangement item, structure building, reading comprehension, assessment

Student Number : 2021-26046