영어어문교육 제 24 권 4 호 2018 년 겨울

Cloze 테스트의 구인과 안면 타당도

한 혜 령
(서원대학교)

Hye-ryeong Hahn (2018). The constructs of cloze test and face validity. *English Language & Literature Teaching, 24*(4), 221-243.

The present study addresses the usefulness of cloze test in terms of its construct and face validity. While cloze has been claimed to be a useful and practical tool that measures language learners' proficiency, discussions on its validity have been mostly limited to its correlations with other tests. Aiming to explore the constructs underlying cloze test, this study conducted a factor analysis based on Korean EFL leaners' test scores on a cloze test. In addition, the study analyzed participants' responses to a questionnaire on the learners' percieved validity of the test. Factor analysis of the paticipants' cloze scores identified genral English ability, as well as phrase-, sentence-, and discourse-level language ability as underlyng constructs. The analysis of the participants' responses to the questionnaire revealed that the cloze had only moderate degree of overall face validity, although the learners tended to perceive the tests' relevance to word-, sentence-, and discourse level language abilities. Correlation analyses suggested that poor topic familiarity is a factor that can lower learners' perceived validity of the test.

[cloze test/construct validity/face validity/EFL, 규칙 빈칸 채우기/구인 타당도/안면 타당도/외국어로서의 영어]

I. 서론

한 개인의 영어 점수는 대학입학, 취업과 같이 인생 경로를 좌우하는 사건들로부터 수준별 분반 작업과 같은 일상의 학습 현장에 이르기까지 적지 않은 영향을 미친다. 영어 학습자들은 자신의 영어 능력을 증명하기 위하여 TOEIC, TOEFL, OPIC 등 공인 영어능력 검증시험을 치른다. 이와 같은 점수를 얻기 위해서는 많은 시간과 비용이 들뿐더러, 교육 현장에서 교사-학생에게 편리한 시공간에서 시행하기도 어렵다.

그 실질적 대안 중 하나로 떠오르는 것이 바로 cloze 테스트이다. 일찍이 읽기교육을 위한 글의 가독성(readability) 측정도구로 제안되었던 cloze 테스트는 글 안에 일정 간격마다 빈칸으로 만든 후, 응답자에게 각 빈칸을 채워 넣도록 하는 시험으로서, 교육자에게나 연구자에게나 여러 면에서 매력적인 도구일 수 있다. 우선 기존의 글에 빈 칸을 만드는 것만으로 시험문제를 만들 수 있기 때문에 학생들의 수준에 맞는 글을 선정하여 cloze를 손쉽게 제작할 수 있고, 시행에 있어서도 짧은 시간 안에 여러 명의 능숙도를 평가할 수 있어 편의성과 실용성을 갖춘 평가도구라 할 수 있다.

이 뿐 아니라 cloze는 그 자체로서 학습자의 언어능력을 잘 대변해 줄 수 있는 타당한 측정도구라는 것을 수많은 연구들이 주장해왔다(Alderson, 1979; Cohen, 1980; Hale et al., 1989; Hinofotis, 1987; Jonz, 1976; Oller, 1979; Oller & Conrad, 1971; Stubbs & Tucker, 1974). cloze 테스트에서 주기적으로 글 안에 배치된 빈 칸에 알맞은 단어를 적어넣기 위해서는 단어지식과 통사적 지식 같은 문장 내적 능력 뿐아니라, 문장 간의 관계와 일관성 등 상위적 지식을 발휘하게 된다는 것이다. Cloze가 이처럼 내용적 측면에서의 타당성까지 갖추었다면 그 잠재적 효용은 매우 크다 할 것이다.

그러나 cloze의 타당성 검증은 준거타당도(criterion-related validity) 측면에 치중되어 왔다. 준거타당도는 타 평가도구와의 상관관계의 크기를 보여줄 뿐, 그 상관관계 이면에 있는 cloze의 구인적 측면, 즉 “cloze 가 대변하는 언어능력이 실제로 무엇인가”라는 측면에서의 타당성을 입증해 주지는 못한다 (Bachman, 1982; Sadeghi, 2013).

또다른 문제는 cloze가 매우 생소한 형태의 시험이라는 사실과 관련된다. 수능이나 TOEIC처럼 짧고 많은 지문과 다양한 주제 및 상황을 중심으로 답을 선택하도록 하는 기존 시험에 익숙한 한국인 영어 학습자들이 단 한 개의 지문, 단 한 개의 주제에 수십 개의 답이 걸려 있는 cloze를 거부감 없이 받아들일 수 있는지, 자신의 영어 능력이 제대로 측정된다고 인식할지 등, 학습자 관점에서의 타당성 문제, 즉 안면 타당도는 cloze 연구에서 진지하게 고려되어지지 않았다. 하지만 시험의 결과에 민감한 국내의 교육적, 사회적 상황을 고려할 때, 안면 타당도의 문제는 결코 가볍다 할 수 없다. 더구나 수험자로부터 얻는 시험에 대한 피드백은 시험의 타당도와 신뢰도를 개선하는 데 중요한 정보가 될 수 있다는 점을 생각할 때(신상근, 2010), cloze의 타당성을 학습자 편에서 묻는 일은 매우 중요하다.

이에 본 연구는 두 가지 측면에서 cloze의 타당성 문제에 접근하고자 한다. 우선, 학습자들의 실제 cloze 점수에 기반한 한 요인분석을 시행함으로써 cloze의 하부요인을 도출하고, 이를 기초로 이 도구의 구인타당도를 논의하고자 한다. 또한 수험자 대상 설문 분석을 통해, 학습자 관점에서의 타당도, 즉 cloze에 대한 안면 타당도를 다각도로 분석해보고자 한다.

Ⅱ. 이론적 배경

1. 언어능력 측정 도구로서의 cloze 테스트

Cloze는 글 속의 누락된 단어들을 적절한 형태로 채워 넣음으로써 글을 완성된 형태로 복원하도록 요구하는 테스트이다. 기본적으로 cloze는 글에서 n번째 단어를(예를 들어 매 7번째 단어마다) 빈칸으로 만드는 규칙빈칸 채우기(fixed-ratio deletion)의 형태를 취하나, 평균 간격은 유지하되 빈칸의 위치는 유동적으로 소폭 조정하는 rational-deletion, 단어의 일부분을 제공하는 C-test, 불필요한 부분을 제거하도록 요구하는 cloze-elide test, 답을 고르도록 하는 선다형 cloze(multiple-choice cloze test)등 목적에 따라 여러 변형된 형태로도 사용되고 있다.

이 도구는 애초에 아동들의 읽기 자료 가독성(readability)을 측정하는 도구로 제안되었지만(Taylor, 1953), 이후 모국어 및 외국어 화자의 다양한 언어능력 측정 도구로서의 유용성이 대두되면서, 외국어 교육 및 평가 분야에서 지속적인 관심의 대상이 되어왔다.

우선 cloze 테스트는 일찍이 읽기 능력의 측정도구로서 관심을 끌어왔다(Bormuth, 1965, 1967; Crawford, 1970; Rankin, 1970; Rankin & Culhane, 1969; Taylor, 1957). Taylor(1957)는 cloze 점수와 읽기시험 점수들에 상관관계가 존재함을 발견하고 이를 토대로 cloze가 영어 산문의 이해력을 측정할 수 있는 도구라고 주장하였다. 후속 연구들에서도 cloze 테스트와 다양한 읽기 시험들 간의 상관관계가 발견되었고 (Anderson, 1976; Bormuth, 1967; Jonz & Oller, 1994; Oller & Jonz, 1994; Rankin & Culhane, 1969; Sadeghi, 2013), 이로써 cloze 테스트는 기존의 읽기 시험을 대체할 수 있는 유용한 대안으로 부상하였다.

나아가 cloze 연구자들은 cloze 점수가 외국어 학습자의 언어능숙도 전체를 대변할 수 있다고 제안하였다(Alderson, 1979; Cohen, 1980; Hale et al., 1989; Hinofotis, 1987; Jonz, 1976; Oller, 1979; Oller & Conrad, 1971; Stubbs & Tucker, 1974). Oller와 Conrad(1971)는 UCLA의 비원어인 학생들의 배치고사 용으로 사용되는 능숙도 검사인 ESLPE를 구성하는 하위 시험들과 cloze 사이에 높은 상관관계를 발견하였고, Jonz(1976) 역시 배치고사 용 능숙도 검사의 하위요소들과 cloze사이에 유의미한 상관관계를 발견하였으며, Hinofotis(1987)는 TOEFL과 cloze 간에 높은 상관관계를 보고하였다. Cloze에 대한 여러 연구들에 대한 종합분석을 바탕으로, Bachman(1982)은 cloze가 거의 모든 언어능숙도 측정도구와 높은 상관관계를 가지고 있다고 요약하고, 이로써 이 도구가 높은 예측타당도(predictive validity) 가진다고 정리하였다. 이처럼 기존 영어능숙도 평가와의 상관관계로 인해 일부 연구자들은 수 시간을 소요하는 기존 능숙도 평가 대신 짧은 시간 안에 시행과 채점이 가능한 cloze를 시행하는 대안을 제안하기에 이르렀다.

언어의 구조적 위계 측면에서 보더라도, cloze는 외국어 학습자의 다양한 언어 측면을 두루 측정할 수 있는 평가도구로 주목받아 왔다. 많은 연구들이 cloze가 단어나 문법과 같은 문장 내적(intrasentential)언어 측면 뿐 아니라 문장들 간(intersentential)의 관계 및 담화와 같은 상위적 측면까지 측정할 수 있다고 제안하였다(Alderson, 1980; Bachman, 1985; Brown, 1983; Chihara et al., 1977; Gaillard & Tremblay, 2016; Jonz, 1987; Lee, 1995; Oller & Conrad, 1971).

Oller 등은 구체적으로, cloze 문항을 풀기 위해서는 음운지식, 단어 지식, 문법 구조에 대한 지식, 의미 파악, 담화 구조에 대한 지식, 앞으로 나올 단어에 대한 예측 능력, 배경 지식, 유추 능력, 읽기 전략 등 언어의 매우 핵심적인 능력들이 동시에 발휘되어야 한다고 설명하였다(Oller, 1979; Oller & Jonz, 1994). 예를 들어 지문 중에 있는 "…an ____ pie…"의 빈칸에 들어갈 단어는 국지적으로만 보아도 모음으로 시작되는 단어여야 하고, 그 중 pie를 수식할 수 있는 형용사나 명사여야 하며, 파이 재료로 쓰일 수 있는 것들 중에는 "apple"이 있다는 등의 지식이 동원되어야 한다는 것이다(Oller & Jonz, 1994). 이들은 나아가, cloze가 단일 항목이 단일 측면을 평가하도록 되어있는 분리평가(discrete-point test)와는 달리 한 항목 안에 언어의 여러 측면이 종합적으로 녹아들어가도록 되어있는 통합평가(integrative test)이며, 이를 통해 학습자의 통합적, 비분리적 언어구사 능력(g-factor)이 제대로 측정될 수 있다고 제안하였다.

실증적 연구들은 cloze 테스트의 이러한 다면성을 입증하고자 하였다(Alderson, 1980; Chihara et al., 1977; Bachman, 1985; Brown, 1983; Gaillard & Tremblay, 2016; Jonz, 1987; Lee, 1995; Oller & Conrad, 1971). Chihara 등(1977)은 영어학습자를 대상으로 한 연구에서 학습자들의 cloze 점수와 해당cloze 본문의 순서잡기 점수간에 높은 상관관계를 발견하고, cloze 가 담화적 차원에서의 일관성 파악능력과 관계가 있다고 제안하였다. 또한 Alderson(1980) 등은 cloze 테스트를 통해 문장간의 응집력(cohesive relationship)에 대한 민감성을 측정할 수 있다고 제안하였다. 한국의 영어학습자들을 대상으로 한 Lee(1997)의 연구와 Kim(2014)의 연구에서는 한국 대학생들의 cloze 점수와 에세이 작문 점수간의 유의미한 상관관계를 보고하였고, Gaillard와 Tremblay(2016)는 유도모방과제(elicited imitation task, 문장을 듣고 구두로 따라하는 과제)가 영어권 국가 체류기간보다 cloze와 오히려 더 높은 상관관계를 보였다고 보고하였다.

2. Cloze 의 타당성에 대한 문제제기

1) 준거 타당도

위에서 언급한 연구들은 대부분 타당도가 어느 정도 검증된 기존 평가도구와의 상관관계를 밝힘으로써 cloze의 타당도를 입증하려는 시도였다. 즉, cloze

테스트의 타당도를 입증하려는 노력들은 주로 준거 타당도에 의존하였다. 그러나 준거 타당도는 다른 평가도구와의 상관관계치에 불과하기 때문에, cloze의 어떠한 내용적 특성이 다른 도구와의 관련성을 이끌어내는지, 어떠한 하부요인이 cloze 안에 존재하는지를 보여주지 못한다는 문제점이 많은 학자들에 의해 제기되었다(Abraham & Chapelle, 1992; Bachman, 1982; Brown, 1985, 2002; Sadeghi, 2013). Sadeghi(2013)는 cloze의 타당성을 주장하는 여러 연구들을 종합분석한 연구에서, 타당도를 입증하는 기존 연구들이 상관관계 중심의 준거타당도 연구에 집중되어 있었으며, 실제로 cloze가 내용적 측면에서 타당한지에 대해, 즉 cloze가 측정하고자 하는 구인들에 대해서 입증하는 연구들은 거의 없었음을 지적하였다. 게다가, 상관관계 연구들에 대한 종합분석 연구에서는 cloze의 준거타당도를 나타내는 상관관계 계수가 연구자 간 상당한 차이가 있음이 지적되기도 하였다(Brown, 2002). 준거 타당도 연구의 이 같은 한계점을 고려할 때, cloze 테스트의 타당성은 그 cloze의 속성 자체, 또는 그 구인의 측면에서 입증될 필요가 있다.

2) 구인 타당도

Cloze가 무엇을 측정하는가를 문제삼은 실증적 연구는 많지 않다. cloze의 구인을 추출해 내기 위해 사용되는 가장 기본적인 실험방식은 요인분석이다. Cloze에 대한 대표적 요인분석의 사례로는 Bachman(1982)의 연구를 들 수 있다. 그는 cloze가 "전반적인 언어능력(g-factor)" 뿐 아니라, 그 하위 3 요소로서 절 범위 내의 "통사적 능력", 문장 간의 관계 파악과 관계되는 "담화응집성 능력", 반복되는 패턴과 관계되는 "전략적 능력"을 측정한다는 가설을 세웠다. Bachman은 이 가설에 입각하여 의도적으로 통사적, 담화적, 전략적 능력과 관련된 단어들만 선정하여 빈칸으로 만드는 방식으로 cloze를 설계하여, 미국의 성인 영어학습자들을 대상으로 총 30문항의 빈칸을 채우도록 하였다. 그는 학습자들의 응답을 토대로 확인적 요인분석을 한 결과, 단일 요인으로서의 전반적 언어 능력과 이를 구성하는 하위 요인인 통사적 요소, 담화응집성 요소, 그리고 전략적 요소가 확인되었다.

그러나 계획적으로 기능어, 담화표지어 등을 빈칸으로 만든 후 요인분석을 하게 되면 기능어들은 기능어들끼리, 담화표지어는 담화표지어끼리 상관관계를 이루어 각기 요인으로 도출되는 것은 당연히 예측할 수 있는 결과로서, 일정 간격으로 규칙적 빈칸을 두는 표준적 cloze 절차에 따라 얻은 점수들을 요인분석 하는 경우에도 유사한 결과가 나온다고 보장하기 어렵다.

보다 최근에 이루어진 몇몇 국외 연구들 중에서도 cloze의 구인 관련 연구들을 찾아볼 수 있다(Khodadady, 2007, 2014; Kongsuwannakul, 2017; Saeedi, 2016). Khodadady(2014)는 학습자에게 수 차례의 C-test를 실시하고 각 C-test 총점들과 다른 능숙도 시험들의 총점들을 가지고 요인분석을 실시한 결과, C-test점

수들은 다른 능숙도 점수들과 비슷하게 “능숙도”로 추정되는 제1 요인에 부하량을 가지는 것으로 나타났으며, 이 밖에도 C-test 고유 특성으로 철자 요건 민감성 같은 요소 등이 파악되었다. 그러나 이 연구는 C-test의 각 문항들에 대한 반응(점수)을 중심으로 이 도구의 하부 구인들을 추출해 내는 대신 여러 시험들의 총점만으로 요인분석을 돌렸기 때문에, 근본적으로는 시험 총점 간 상관관계에 기반하는 연구라 할 수 있다. Kongsuwannakul(2017)은 코퍼스를 활용한 용어색인(concordance) 기반 빈칸 채우기인 ConCloze를 개발하고 그 유용성을 입증하는 절차로서 구인타당도를 조사한 결과, 단어의 의미, 연어(collocation) 지식, 배경지식, 문법지식 등이 그 하부 요인으로 추출되었다고 보고하였다. Kongsuwannakul은 이를 근거로 ConCloze가 학습자의 능숙도를 측정할 수 있는 유용한 도구가 될 수 있다고 제안하였다. 그러나 ConCloze는 한 문항의 빈칸에 대해 코퍼스 기반 단어색인이 제공하는 한 줄짜리 맥락을 여러 개 제공하며, 문장간 연결이 결여되었다는 점에서 하나의 연결된 담화에 규칙적으로 빈칸을 두는 표준적 cloze와는 거리가 있다.

국내의 cloze 구인 연구는 더 찾아보기가 힘들다. Cloze의 구인에 대한 국내 연구로는 Lee(2002)를 들 수 있다. Lee는 cloze 테스트에서 학습자들이 활용한 전략들의 분석을 통하여 cloze의 구인을 간접적으로 추정하고자 하였다. 이 연구에서는 선다형 cloze를 시행한 후 학습자들과의 구두면담을 통해 학습자들이 시험 당시 사용했던 전략들을 조사하였다. 면담을 통한 전략 분석을 통하여 Lee는 단어 수준 능력, 문장 수준 능력, 담화 수준 능력, 텍스트 외적 능력과 추론 능력 등을 cloze 의 구인으로 제안하였다. 그러나 이 연구는 전략 사용을 매개로 하는 간접적인 구인 연구로서, 학습자 수행을 바탕으로 한 보다 직접적인 구인의 연구로 그 결과를 확인할 필요가 있다.

3) 안면 타당도

cloze 테스트의 또 다른 잠재적 문제점은 이 평가방식이 다른 기존 테스트와는 사뭇 다른 모습을 하고 있어, 수험자들의 입장에서 그 타당성에 대한 의구심이 생길 수 있다는 점이다(Bormuth, 1968; Brown, 2007). Brown(2007)은 ESL 실험집단 학생들을 대상으로 한 배치고사로 cloze와 받아쓰기(dictation)을 시행했을 때의 경험을 아래와 같이 술회하고 있다(p.449).

> … Once I administered a dictation test and a cloze test...as a placement test for an experimental group of learners of English as a second language. Some learners were upset because such tests, on the face of it, did not appear to them to test their true abilities in English…

만일 학습자들이 이 시험이 자신의 능력을 제대로 평가하지 못한다고 느낀다면, 자신의 최대치를 이끌어내기도, 결과를 받아들이기도 어려울 것이다. 학습자가 평가도구가 얼마나 타당하다고 생각하는가, 즉 안면 타당도는 다분히 주관적일 수 있으나, 영어 능숙도 측정이 한 개인의 현재와 미래에 미치는 파급력을 생각해 볼 때, 그리고 향후 교수-학습의 방향에 미칠 환류효과(washback)을 고려할 때, 안면 타당도의 문제는 진지하게 고려되어야 한다. 그러나 학습자의 관점에서 cloze의 안면 타당도를 연구 과제로 삼은 사례는 찾아보기가 힘들다.

이상 살펴본 바, cloze 연구의 문제점은 아래 두 가지로 요약될 수 있다. 첫째, 대부분의 연구들은 cloze의 타당성을 주로 준거타당도에서 찾아왔다. Bachman 등의 구인 연구가 있으나, 이는 의도적으로 측정하고자 하는 평가요소를 정하고 각 요소에 해당하는 단어를 빈칸으로 만들어 놓은 조건에서 수행된 확인적(confirmatory) 요인분석에 의한 것이므로, 단지 측정하고자 의도했던 하위요소를 측정했다는 의구심이 일게 된다. 따라서 표준적인 cloze를 절차에 따라 시행된 시험 결과를 토대로 구인을 탐색적으로 추출해내는 작업이 필요하다.

둘째로, cloze에서 얻은 점수를 근거로 학습자들의 언어 능력을 제대로 추론해 내기 위해서는 학습자의 능력이 온전히 cloze에서 발휘되어야 하는데, 이를 위해서는 학습자 관점에서 cloze가 자신들의 능력을 잘 대변해준다고 생각해야 한다. 즉, 안면 타당도의 문제가 해결되어야 한다. 하지만 cloze의 안면 타당도 문제가 제기되기는 했어도 이를 체계적으로 학습자 시각에서 들여다본 사례는 발견하기 힘들다.

III. 연구 방법

Cloze 테스트의 구인 및 액면타당도를 살펴보기 위해 본 연구는 충청지역 한 대학에 재학중인 영어교육 전공생들을 대상으로 cloze 테스트와 설문조사를 실시하였다. cloze 테스트는 2017년 당시 3학년이었던 29명의 학생들과 2018년 당시 3학년이었던 학생들 44명을 대상으로 2017년과 2018년에 각각 실시되었으며, 이들 중 2018년 참여자들(n=44)은 설문조사에도 참여하였다. Cloze 테스트 총 참여자 73명 중 대부분(n=69)은 TOEIC점수 보유자로서, 표 1에서 볼 수 있는 바와 같이 평균점수 780점 가량의 중상급 학습자 집단이라 할 수 있다.[1]

[1] 연구대상자들의 소속학과에서는 재학생들에게 토익점수를 매학년마다 제출하도록 하였는데, 본 cloze 테스트를 본 학생들 중 4명은 토익점수를 제출하지 않았다.

표 1

참여자 TOEIC 점수 분포

	최소값	최대값	평균	표준편차	n
TOEIC 점수	610	970	780.22	82.83	69

1. Cloze 테스트

1) 시험 도구

본 연구에 사용한 cloze 테스트로는 Brown(1980)에서 사용하였던 것을 택하였다. 빈칸을 포함한 본문은 ESL 읽기교재인 “Man and His World”를 출처로 하였으며, 석기시대의 인간의 삶과 도구의 사용에 관한 일반교양 수준의 내용이 담긴 “Man and His Progress”라는 제목의 글이었다. 총 456 단어 분량의 본문 안에 50개 단어를 빈칸으로 만들었으며, 7번째 단어마다 규칙적으로 빈칸으로 만든 고정비율 생략 방식(fixed-ratio deletion)을 취하였다. 빈칸에 들어갈 답의 구성을 보면, 내용어 27개, 기능어 23개로서 내용어-기능어가 비슷한 비율로 정답에 포함된 시험인 것으로 분석되었다(X^2=.320, p=.572). 그림 1에서와 같이, 도입부의 맨 처음 두 문장은 맥락 파악을 돕기 위해 빈칸 없이 주어졌고, 글의 마지막 문장에도 빈칸을 두지 않은 표준적 cloze였다(Brown(1980) 부록 참조).

2) 시험 절차

참여 학생들은 시험 전에 연구자로부터 여러 개의 답이 가능하며, 맥락상 적절한 답이면 모두 정답 처리함을 고지 받은 후, 25분 동안 주어진 글을 읽어나가면서 50개의 빈 칸에 들어갈 단어들을 직접 써 넣었다.

그림 1

Cloze 테스트의 도입부

MAN AND HIS PROGRESS

Man is the only living creature that can make and use tools. He is the most teachable of living beings, earning the name of Homo sapiens. __(1)__ ever restless brain has used the __(2)__ and the wisdom of his ancestors __(3)__ improve his way of life. Since __

ANSWERS
1. _____
2. _____
3. _____
4. _____
5. _____
6.

3) 채점 방식

회수 답지의 채점은 원문의 단어만을 답으로 채택하는 방식(exact answer scoring) 대신, Brown(1980)의 제안을 따라, 가능한 복수의 답을 모두 정답으로 채택하는 방식(acceptable answer scoring)을 취하여 정답에는 1점, 오답은 0점을 부여하였다. 예를 들어 그림 1에 제시된 본문의 첫 번째 빈칸은 원문에서 “his”이지만, 가능 답 목록에는 “man’s”, “our”, “the”들도 함께 제시되어 있어 원문에 쓰인 단어이든 가능 답이든 이 중 하나를 답으로 썼으면 정답으로 처리하였다.

Brown(1980)에 따르면, 여러 채점 방식을 비교해 본 결과, 원문의 답만을 정답으로 채택하면 정답율이 지나치게 낮고, 선다형은 정답율이 높아지는 반면, 가능답을 모두 받아들이는 방식은 적정 난이도인 .50 정도의 정답율을 보였으며, 또한 여러 방식 중 가장 높은 변별력과 신뢰도를 보이는 것으로 드러났다. 무엇보다도 원문에 쓰인 단어만을 정답으로 채택하는 경우, 맥락 상으로 볼 때 충분히 적절한 답임에도 불구하고 오답 처리됨으로써, 그 결과에 대한 수험자들의 의구심과 불만을 가중시킬 수 있다는 점을 고려할 때(Brown, 1987; Hinofotis, 1987; Lee, 1997), 가능한 복수 정답 처리 방식이 가장 적합하다고 판단하였다. 가능 답은 Brown이 제시한 가능 답들로 한정하였다.

2. 설문지 조사

Cloze의 안면 타당도와 신뢰도, cloze 테스트에 대한 학습자들의 체감 난이도 등 cloze 테스트가 학습자 관점에서 어떻게 인식되는지를 알아보기 위해, 2018년에 cloze시험을 치른 44명을 대상으로 시험 직후 설문을 실시하였다. 설문에는 본 시험의 체감 난이도, 체감 타당도, 체감 신뢰도 영역이 포함되었다. 또한 cloze가 어떤 능력을 요구하는지에 대한 학습자들의 생각을 구체적으로 알아보고자, Lee(2002)의 전략 기반 구인 연구를 비롯, 선행 연구들을 참조하여 문항을 설계하였다.

표 2는 설문 문항들을 영역별로 분류하여 제시한 것이다. 설문 참여 학생들은 각 문항에 5점 리커트 척도(1=전혀 그렇지 않다, 5=매우 그렇다)로 응답하였다.

표 2

설문지 문항의 영역별 분류

영역	문항 내용	번호
전반적 타당도	이 테스트는 내 영어실력을 잘 대변해 준다.	1
	이 테스트가 무엇을 알아보려는 것인지 알기 어렵다.	2
	이 테스트 점수는 나의 공인영어시험 점수와 상관관계가 있다.	5

	토익점수가 내 실력을 더 잘 반영한다.	6
	앞으로 이런 형식의 평가를 도입하는 것이 필요하다.	7
난이도	이 테스트는 어렵다.	3
	이 테스트는 나에게 익숙하다.	4
	이 테스트에 사용된 글은 어려웠다.	9
	이 글의 주제는 생소하였다.	11
	더 쉬운 글이었다면 실력이 더 정확히 반영되었을 것이다.	10
	더 익숙한 주제였다면 실력이 더 정확히 반영되었을 것이다.	12
세부	단어 실력이 요구된다.	13
요인	빈칸 단어의 문장안에서의 문법적 역할을 따져봐야 한다.	14
	문장의 의미를 파악하는 것이 요구된다.	15
	앞뒤문장들에서 어떤 내용이 나왔는지 참고해야 한다.	16
	다른 문장들에서 쓰인 단어들을 눈여겨보아야 한다.	17
	앞뒤 문장과의 연결관계를 살펴보아야 한다.	18
	글 전체 맥락을 살펴보아야 한다.	19
	듣기/말하기/읽기/쓰기 중 시험과 관련성이 큰 순서로 쓰세요.	8
신뢰도	이 테스트에서 받은 점수를 신뢰할 수 있다.	22
	답을 적어넣을 당시, 채점자가 제대로 채점할지 걱정이 되었다.	23
	여러개의 답(복수답)이 가능한 문항들이 많다.	21

3. 자료 분석

1) Cloze 결과 분석

본 연구의 모든 통계작업에는 SPSS 버전 23을 사용하였다. 우선 참여자들의 cloze 항목별 점수들을 토대로 탐색적 요인분석을 실시하여 구인을 추출하였다. 시험이 지나치게 어렵거나 쉬운 경우 문항이 제대로 변인 역할을 하지 못한다는 일부 연구자들의 우려를 고려하여(Brown, 2002), 시험 난이도 및 항목별 난이도가 적정한지도 확인하였다. 아울러, 학생들의 TOEIC과 cloze 점수의 상관관계를 토대로 cloze의 준거 타당도를 가늠하였다.

2) 설문 결과 분석

참여자들의 설문 응답을 토대로 cloze 테스트에 대한 학습자들의 체감 타당도와 체감 신뢰도를 알아보고, 이들과 상호관련성을 보이는 항목들을 근거로 cloze의 어떤 측면들이 학습자 관점에서의 시험 타당도 및 신뢰도에 영향을 미치는 변수로 작용하는지 분석하였다. 또한 cloze 테스트의 구체적 측정내용

과 관련된 설문항목들의 분석을 토대로 학습자들이 cloze를 어떤 능력에 대한 평가로 인지하는지 알아보았다.

Ⅳ. 연구 결과 및 논의

1. Cloze의 요인분석

1) 시험 난이도와 항목 난이도

학생들은 평균적으로, 50점 만점의 시험에서 약 반 정도를 맞춘 것으로 나타났다 (M=24.67, SD=6.90). 표 3에서 보여주듯이, 최고점자는 38점을, 최저점자는 7점을 받았다. 평균 문항난이도(IF)는 0.493으로서, 근50%의 이상적인 정답율을 보였다. 표준편차(SD)는 0.211로, 전체 문항의 70%가량이 0.28~0.70의 난이도 범위 안에 고루 분산되어 있었다. 가장 쉬웠던 항목(11번, 정답: “of”)과 가장 어려웠던 항목(15번, 정답: “often”)은 표에서와 같이 각각 0.88과 0.07의 정답율을 보였다. 적정한 평균 난이도와 고르게 분산된 항목 난이도를 볼 때, 이 시험 결과를 토대로 한 요인분석에서 많은 문항들이 변인으로서 제 역할을 수행할 수 있는 조건이 대체로 만족되었다 할 수 있다.

표 3

Cloze 테스트 점수 및 항목 난이도 분포

	최소값	최대값	평균	표준편차
Cloze 총점	7	38	24.67	6.90
문항 난이도(IF)	.07	.88	.493	.211

2) 공인시험점수와의 상관관계

학습자들이 보유한 TOEIC 점수와 cloze 점수간의 Pearson 상관관계 검증 결과, 두 시험은 표 4에서와 같이 .05 수준에서 유의미한 상관관계를 보였다. 공인시험인 TOEIC을 학습자 영어능숙도의 기준으로 본다고 가정한다면, cloze는 높지는 않지만 어느 정도의 준거타당도를 보인다고 할 수 있다.

표 4

Cloze와 TOEIC 점수의 상관관계

	상관계수	유의도	n
TOEIC 총점	.259*	.033	68
듣기 점수	.245	.128	68
읽기 점수	.225	.163	68

한편 TOEIC을 듣기와 읽기로 따로 나누었을 때, 각각의 점수와 cloze 점수 간 상관관계는 유의미한 수치에 이르지 못했다(듣기: r= .245, p = .129; 읽기: r=.225, p =.163).

3) 요인분석

참여자 집단의 점수를 토대로 탐색적 요인분석을 한 결과 아래와 같은 주성분들이 검출되었다. 표5는 각 성분별로 부하된 항목들 중 0.3 이상의 부하량을 가진 항목들만을 표시한 것이다(Hatch(1991) 참조).

요인분석표에서 가장 눈에 띄는 것은 테스트 전반에 대다수의 문항이 첫번째 요인에 부하되었다 점이다. 이들은 하나의 특성을 중심으로 수렴되기 어려운 다양한 성격을 가지고 있었다. 내용어(n=12)와 기능어(n=14)가 골고루 분포되어 있었으며, 담화 관계를 나타내는 같은 대명사, 문장 내 의미 관계를 표시하는 전치사와 접속사, 담화 앞부분에서 이미 나온 단어들, 담화에서는 한번도 나오지 않은 단어들이 함께 요인 1에 실린 것으로 나타났다.

이처럼 다양한 문항들이 상호 관련 지어졌다는 것을 볼 때, 요인 1은 Oller (1979)가 주장하는 바, 학습자의 비분리적이고 총체적인 언어능력, 즉 g-factor가 발현된 것으로 볼 수 있다. 주목할 만한 것은 cloze 총점이 매우 높은 부하량을 보이면서 이 요인의 주요 변인지표로 실렸다는 점이다. 이로 볼 때 이 요인 1은 cloze의 본질적 성격으로 제안되어 온 학습자의 “전반적 언어능숙도”를 나타내는 것으로 분석된다.

표 5

Cloze 테스트 요인분석

	1	2	3	4	5	6	7	답
cloze	.984							-
27	.613							it
42	.552							in
6	.528				.356			been
11	.504							of
45	.500							women
41	.481					.389		cave
30	.478							stick
10	.458					.339		periods
21	.353							dig
39	.358							and
49	.402							and
18		.612						the
17		.556						the
16		.419						the

34	.323	.373					tool
28			.613				man
4			.437				man
38	.309		.415				could
23			.385			.397	sharp
35	.374						development
7				.590			tools
29				.482		.362	sharp
8	.387			.480			during
13		.302		.474			fishing
31	.337						had
5	.318						on
36	.322						have
46	.406						or
9	.357				.404		which
12	.367				.379		all
33					.339		in
19	.342						man's
2						.301	knowledge
44						.498	home
26	.372		.351				one
22	.328						bone
40	.383						draw
48	.369						made

두 번째 성분에는 .3 이상의 부하량을 가진 네 문항이 실렸는데, 34번 “tool”을 제외한 나머지 세 개의 문항은 모두 “the”가 정답이다(the women, the women, the men). cloze 전체에서 정관사 “the”가 정답인 항목은 총 5개이지만, 요인 2에 실린 “the”는 모두 총칭을 나타내는 “the+명사” 표현으로서, 문장 안에서의 관계를 굳이 파악하지 않더라도 구(pharase) 범위 안에서 해결 가능한 매우 국지적 문법과 관련된다. 따라서 두 번째 요인은 “구 범위 내의 관사 지식”으로 정의할 수 있다.

세 번째 성분에서 .3 이상의 부하량을 가진 요소 중 높은 부하량을 가진 두 문항은 모두 “man”을 정답으로 한다. 한편 좀더 낮은 부하량을 가진 “could”와 같은 문항들도 요인 3에 포함된 것을 잘 설명하기는 어렵다. 그러나 cloze 항목의 비분리적 성격을 고려할 때, 빈칸을 인위적으로 조정하지 않은 cloze에서 Bachman(1982)에서처럼 관련 변인들만이 깔끔하게 정리되어 나오는 것은 사실상 비현실적이다. 이런 점들을 고려할 때 이 성분은 본문 주제인 “man”과 관련된 한 “관점 일관성”과 관련된 담화 능력이 아닌가 추정된다.

네 번째 성분을 구성하는 단어들은 내용어들을 중심으로 구성되어 있는데, 이들은 본문에서 비교적 빈도가 높은 단어들로, 글의 앞뒤 문장들에서도 발

견된다. 수험자는 담화 내에서 앞선 문장들에 나온 표현들을 참조하여 답을 쓸 수 있다. 따라서 요인 4는 “담화 내 응집성 기반 의미 구축”과 관련되는 것으로 추정된다.

요인 5는 완료수동태의 ‘been’, 관계대명사 ‘which’, 양화사 ‘all’, 전치사 ‘in’ 등 문장 안에서의 문법적-내용적 관계를 고려하여 해결할 수 있는 기능어 항목들로 이루어져 있었다. 따라서 요인 5는 “문장내 문법-의미 관계”로 대변될 수 있는 것으로 보인다.

요인 6에 실린 문항들은 모두 명사 내용어들이었는데, 이들은 앞뒤의 의미적 맥락을 살펴 본문에 없는 단어를 일반 지식과 추론을 활용하여 창의적으로 써 넣어야 하는 문항들이었다. 이러한 점에서 이 요인에 실린 내용어들은 빈칸의 단어가 본문의 다른 곳에 이미 나와있는 항목으로 구성된 요인 4의 내용어들과는 차별화된다. 따라서 이 요인은 “담화 내 추론 기반 의미 구축”으로 정의될 수 있다.

마지막 성분(요인 7)은 3개의 연속된 문장 속에 배치된 (23), (29)번 문항과 관련되는데, 두 문항 모두 “sharp”를 원문 답으로 하며, “…a (23) stone…a (29) piece of stone…”과 같이 유사병렬적 구절 안에 들어 있다. 또한 그 사이에 “stone was...until it has a sharp edge”와 같은 “sharp”과 “stone”이 모두 들어가는 설명적 문장이 나오므로, 이를 참조하여 답할 수 있는 문항이다. 따라서 이 요인은 “근접 맥락 정보 활용”과 관련된 것으로 추정되었다.

추출된 요인들을 종합해 볼 때, 본 cloze 는 학습자의 전반적이고 통합적인 언어능력을 이끌어내는 많은 문항들을 포함하고 있고, 또한 국지적인 수준에서부터 담화 수준에 이르는 언어적 구인들을 그 하부 구조로 내포하고 있다고 결론 내릴 수 있다.

3. 설문지 분석 결과

1) Cloze 의 체감 난이도

이 테스트가 어려웠는지 묻는 문항(문항 3)에 대하여 참여자들은 5점 척도에서 약4점의 평점을 주었다(M=4.09). 68.2%의 응답자가 “그런 편이다(=4)”라고 답하였으며, “매우 어려웠다(=5)”고 답한 응답자도 20.5%에 달했다. 테스트가 익숙했는지를 묻는 문항에는 2점에 가까운 낮은 평점을 주었다(문항3, M=1.98). 이는 대체적으로 cloze가 이들에게 어렵고 낯선 시험이었음을 알려준다. 그러나 참여자들의 체감 난이도는 cloze 총점 및 TOEIC 총점과 유의미한 상관관계를 보이지 않아(cloze와의 상관관계: r=.20, TOEIC과의 상관관계: r=.24), 체감 난이도와 실제 학습자 수준은 별개인 것으로 나타났다.

시험이 어려웠던 것에 반하여, 본문 난이도는 수험자 집단에게 그리 심각하지 않았으며(문항9, M=2.61), 글의 주제도 생소하지 않은 편이었던 것으로

나타났다(문항 11, M=2.32). 그러나 글이 어려웠다고 한 응답자와 소재가 생소하다고 한 응답자도 각각 전체의 25% 및 24%에 달했다. 표 6은 설문 응답자들의 체감 난이도 관련 항목들의 평균을 정리한 것이다.

체감 난이도 관련 항목들의 상관관계 분석 결과, 표 7에 나타난 바와 같이 학습자들의 cloze에 대한 총체적 체감 난이도를 묻는 문항(문항 3)은 시험 자체의 친숙도과 관련된 문항(문항4)과만 강한 부적 상관관계를 보였고(r=-.549, $p < 0.01$), 글이나 주제 난이도와는 관련되지 않았다.

표 6

Cloze 관련 체감 난이도

번호	문항내용	평균	표준편차
3	이 테스트는 어렵다	4.09	.56
4	이 테스트는 나에게 익숙하다	1.98	.82
9	이 테스트에 사용된 글은 어려웠다.	2.61	.92
11	이 글의 주제는 생소하였다.	2.32	1.14

표 7

체감 난이도 관련 항목들의 상관관계

번호	문항내용	3	4	9
3	이 테스트는 어렵다			
4	이 테스트는 나에게 익숙하다	-.549**		
9	이 테스트에 사용된 글은 어려웠다.	.294	-.135	
10	이 글의 주제는 생소하였다.	-.098	.139	.412**

이로써 본 학습자들에게 이 시험이 어려웠던 이유는 글의 난이도보다는 생소한 형태의 시험이 요구하는 과제의 난이도 때문임을 짐작할 수 있다. 한편, 글 자체와 관련된 문항들(문항 9, 10)간에는 높은 상관관계가 나타났는데, 이는 주제 친숙도에 대한 개인차가 글의 체감 난이도의 개인차로 이어질 수 있음을 보여준다.

2) 전반적 체감 타당도

본 cloze가 자신의 실력을 잘 대변해 주는가를 묻는 문항(문항1)에 대한 긍정적 응답은 59.0%, 부정적 응답은13.6%으로, 전체적으로는 약한 긍정의 경향을 보였다(M=3.50). 응답자들은 “이 테스트가 무엇을 측정하려 하는지 알수 없다”라는 문항(문항2)에 대해 미약한 부정(M=2.70)의 경향을, 이런 형식의 평가를 도입하는 것에 대해서는 중립적인 입장을 보였다(M=2.95).

한편 공인시험인 TOEIC에 대한 타당도 평가는 cloze의 경우보다 더 낮은 것으로 나타났다. TOEIC점수가 cloze보다 실력을 더 잘 반영한다고 생각하는

지를 묻는 문항에 응답자들은 다소 부정적 입장을 취했다(M=2.75). 이 응답들을 종합해볼 때, 학습자들은 cloze 테스트의 타당성에 대해 미약하나마 긍정적인 평가를 내렸다고 할 수 있다.

표 8은 응답자들의 cloze 타당성에 대한 전반적 인식과 관련된 문항의 응답 경향을 정리한 것이다. 이 표에서 볼 수 있듯이, 체감 타당도가 글의 어려움이나 주제의 생소함으로 인한 것인가에 대한 질문(문항 10, 12)에 대한 응답은 특이한 경향을 보이지 않았는데, 이는 본 cloze의 글이나 소재 자체가 그리 어렵지 않았던 것과 관계가 있어 보인다.

표 8

안면 타당도 관련 학습자 평가

번호	문항내용	평균	표준편차
1	이 테스트는 내 영어실력을 잘 대변해 준다	3.50	.79
2	이 테스트가 무엇을 알아보려는 것인지 알기 어렵다	2.70	.99
5	이 점수는 나의 공인영어시험 점수와 상관관계가…	2.95	.84
6	토익점수가 내 실력을 더 잘 반영한다	2.75	1.04
7	앞으로 이런 형식의 평가를 도입하는 것이 필요하다	2.95	.98
10	쉬운 글이이라면 실력이 더 정확히 반영되었을 것이다	3.11	1.04
12	익숙한 주제라면 실력이 더 정확히 반영되었을 것이다	3.05	1.26

안면 타당도 관련된 문항인 문항1, 2, 7과 타 문항과의 주요 상관관계를 표 9에 제시하였다. 우선 cloze의 체감 타당도(문항1)는 cloze의 평가 내용에 대한 의구심(문항2)과 강한 부적 상관관계를 가지는 반면, cloze 도입 찬성 입장(문항7)과는 정적 상관관계를 보였다. 바꾸어 말하면, cloze 가 자신의 실력을 대변하지 못한다고 본 사람들은 이 시험이 무엇을 알아보려는지도 모르겠고 이런 평가가 바람직하지 않다고 보는 경향이 있었음을 의미한다.

표 9

안면 타당도와 상관관계를 보이는 문항들

번호	2	7	11	22	23
1	-.533**	.565**	-.336*	.391**	-.196
2	1	-.336*	.530**	-.379*	.196
7	-.336*	1	-.028	.506**	-.419**

1. 실력을 잘 대변함 2. 무엇을 알아보려는지 모르겠음 11. 주제 생소함
22. 점수 신뢰 23. 제대로 채점할지 걱정됨
*: $p < .5$, **: $p < .01$

또한 눈여겨볼 만한 점은 cloze 타당도에 대한 질문(문항 1)은 시험 난이도, 본문 난이도, 시험 친숙도, 주제 친숙도 등 여러 변인 중 주제 친숙도 관련 문항(문항 11)과만 유의미한 상관관계를 보였다는 것이다($r=-.336^*$, $p<.05$) 이는

낮은 주제 친숙도로 인해 자신의 실제 영어실력과 상관없이 cloze 점수가 좌우되는 것에 대한 학습자의 우려를 시사한다. 학습자 집단 전체로 보았을 때는 주제가 대체로 평이했지만, 이 주제에 익숙하지 않았던 일부 학습자들에게 있어서는 전체 문항을 한가지 주제를 중심으로 구성할 수 밖에 없는 cloze의 특성이 체감 타당도 감소의 원인이 될 수 있음을 보여준다.

한편, 안면 타당도 항목들은 체감 신뢰도 관련 문항들(22, 23)과도 유의미한 상관관계를 보였는데, 응답자들의 체감 신뢰도와 관련해서는 아래에 논의하도록 하겠다.

3) 체감 신뢰도

응답자들은 cloze의 점수를 신뢰할 수 있는가에 대한 질문에 대하여도 중립적 입장을 취하였다(M=3.16). 이들은 대체로 복수 답이 존재하는 많다고 인지하였으나(M=4.14). 채점자가 채점을 제대로 할지에 대한 걱정은 그리 높지 않았다(2.66), 이는 복수의 답이 가능하며 문맥에 적절한 답이면 정답으로 처리될 것임을 수험자가 미리 알았던 것과 관계있으리라 짐작된다. 표 10은 학습자들의 체감 신뢰도 영역 문항들의 응답 경향을 정리한 것이다.

표 10

체감 신뢰도 관련 학습자 평가

번호	문항내용	평균	표준편차
21	여러개의 답(복수답)이 가능한 문항들이 많다	4.14	.63
22	이 테스트에서 받은 점수를 신뢰할 수 있다	3.16	.71
23	답을 적을 때, 채점자가 제대로 채점할지 걱정이 되었다	2.66	1.27

상관관계 분석 결과, 체감 신뢰도(문항 22)과 유의미한 상관관계를 가진 항목들은 표 11과 같이 나타났다. 상관관계 분석에 따르면 시험 점수에 대한 신뢰는 채점에 대한 염려(문항 23)가 적을수록 커짐을 알 수 있는데, 이는 채점의 객관성과 일관성이 신뢰도의 핵심 내용을 구성한다는 점을 평가이론이 아닌 일반 수험자의 입장에서도 재확인해주는 것이라 하겠다.

표 11

체감 신뢰도 문항과 상관관계를 보이는 문항들

번호	1	2	7	10	11	12	15	23	Cloze
22	.391**	-.379*	.506**	-.495**	-.351*	-.371*	.371*	-.425**	.331**

1. 실력을 잘 대변함 2. 무엇을 알아보려는지 모르겠음
7. 이런 평가 도입 필요 10. 쉬운 글이라면 실력이 더 잘 반영될 것임
11. 주제 생소함 12. 익숙한 주제라면 실력이 더 잘 반영될 것임
15. 문장 의미 파악 요구됨 23. 제대로 채점할지 걱정됨

* $p < .05$, ** $p < .01$

또한 이 상관관계표는 앞서 간략히 언급했듯이, 체감 신뢰도가 타당도 관련 항목들과 강한 상관관계에 있음을 보여주고 있다. cloze 점수를 신뢰할 수 있다고 한 응답자들은 이 도구가 본인의 실력을 잘 대변해준다고 판단하였고, 차후 cloze로 평가받기를 원하는 것으로 나타난 반면, cloze의 결과를 신뢰할 수 없는 학습자는 이 도구가 무엇을 평가하는지 모르겠고, 자신의 능력을 잘 발휘하기에는 글이 어렵고 주제가 생소하다고 느끼는 경향을 보였다. 학습자의 체감 신뢰도와 타당도의 밀접한 상관관계는 신뢰도가 타당도의 필수 선제조건이라고 하는 평가분야 학자들의 주장과도 일치한다(Bachman, 1990).

4) Cloze가 요구하는 능력에 대한 인식

이 시험이 요구하는 능력을 내용적으로 살펴본 문항들은 단어, 문장, 담화 수준의 능력에 대한 문항들로 구성되어 있었는데, 응답자들은 표 12에서 보여주는 바와 같이 이 영역 모든 문항들에 높은 평점을 주었다. 이는 Lee(2002)에서 제안한 cloze 수험 전략들에 대한 양적 증거를 제공해 주는 셈이다.

표 12
cloze에서 요구되는 언어적 능력 인식

번호	문항내용	평균	표준편차
13	단어 실력이 요구된다.	4.30	.70
14	문장 내 문법적 역할을 따져봐야 한다.	4.20	.70
15	문장의 의미를 파악하는 것이 요구된다.	4.64	.49
16	앞뒤문장들에서 어떤 내용이 나왔는지 참고해야 한다.	4.57	.55
17	다른 문장들에서 쓰인 단어들을 눈여겨보아야 한다.	4.39	.62
18	앞뒤 문장과의 연결관계를 살펴보아야 한다.	4.55	.50
19	글 전체 맥락을 살펴보아야 한다.	4.23	.74

한편, 문장, 담화 수준의 항목들은 주로 동일 수준 내에서만 긴밀한 관계를 가지는 것으로 발견되었다. 표 13은 이 시험에 필요한 다양한 능력들에 대한 민감도 간의 상관관계를 보여주고 있다.

상관관계표에 따르면, 단어수준 관련 문항(문항 13)은 문장이나 담화 수준 관련 문항과 상관관계가 매우 낮게 나타난다. 이는 cloze의 단어적 측면을 인식하는 학습자들이 문장이나 담화수준에 대한 내용적 타당도를 인식하지 못할 수 있음을 의미한다. 한편 담화수준 관련 문항들(문항 16-19)은 서로 매우 강한 상관관계를 보였는데, 이는 담화수준에 대한 총체적 인식이 존재함을 보여준다. 문장 수준 능력과 관계된 두 문항(문항 14, 15) 간에도 역시 강한 상관관계를 보였는데, 흥미로운 점은 문장의 문법 차원에 대한 의식(문항14)은 담화 차원에 대한 의식들과 유의미한 상관관계를 보이지 않았던 반면, 문장의 의미 차원에 대한 의식(문항 14)은 담화 차원의 모든 문항들과 .05 유의

수준에서 상관관계를 보였다는 것이다. 이는 담화의 일관성이 의미의 흐름을 기반으로 하며(Oller & Jonz, 1994), 의미의 흐름을 파악하는 것은 문장의 의미를 파악하는 것으로부터 시작하기 때문인 것으로 해석된다.

표 13
Cloze에서 요구되는 언어적 능력 인식 간의 상관관계

	13	14	15	16	17	18
14	.016					
15	.05	.495**				
16	-.023	-.128	.358*			
17	-.108	.028	.323*	.437**		
18	-.006	.138	.354*	.454**	.502**	
19	-.087	.221	.298*	.477**	.361*	.593**

13. 단어 실력 14. 문장내 문법 역할 파악 15. 문장 의미 파악
16. 앞뒤문장 내용 참고 17. 타문장 단어 참고 18. 앞뒤 문장과의 연결
19. 글 전체 맥락 파악

V. 요약 및 결론

본 연구는 한국 영어학습 맥락에서 cloze 테스트의 타당성을 구인 타당도 및 안면 타당도의 측면에서 검증하고자, 학습자의 cloze 수행 결과에 기반한 요인분석과 학습자 대상 설문 분석을 시행하였다. 그 결과는 다음과 같이 요약될 수 있다.

첫째, 요인분석 결과, 다음과 같은 요인들이 추출되었다. 우선 다양한 종류의 다수 문항들이 하나의 주요 성분에 부하되었는데, 이는 cloze가 학습자의 비분리적이고 전반반적 영어능력인 g-factor를 측정한다는 선행 연구들을 한국의 영어 학습 맥락에서도 확인해 주는 결과이다.

둘째, cloze전반적 영어능력 외에도 “구 범위 내 관사 사용”과 같은 국지적 능력으로부터 “문장 내 문법관계 파악”, “담화 응집성 기반 의미 구축”, “담화 내 추론 기반 의미 구축”, “관점 일관성”과 같이 문장 수준, 담화 수준의 요인들이 추출되었는데, cloze 는 문장 수준에서의 언어능력만을 측정한다고 주장한 일부 연구들과는 달리, 전반적 언어능력과 구, 문장, 담화 수준의 언어 능력을 측정하기에 타당한 통합적 평가도구임을 보여준다.

셋째, 설문 분석 결과 cloze테스트에 대한 전반적인 체감 타당도는 다소 애매하게 나타났다. 전반적인 cloze의 체감 타당도는 그리 높지 않았지만, TOEIC과 같은 기존 시험에 비해 낮지도 않은 것으로 드러났다. 이 시험의 타당도를 낮게 평가한 응답자들은 신뢰도 역시 낮게 평가하는 경향을 띠었는데, 신뢰도와 타당도에 모두 연결된 시험 요소는 “주제의 생소함”이었다. 이는 단

일 글은 단일 주제에 편중될 수 밖에 없다는 cloze의 본질적 속성과 관련된 문제로서, 하나의 주제를 중심으로 한 일관성의 구축은 cloze가 가지고 있는 큰 장점이지만, 다른 한편으로는 이로 인해 주어진 주제에 친숙하지 않은 학습자가 자신의 언어적 능력을 제대로 발휘할 수 없다는 것을 의미하며, 주어진 cloze 테스트가 무슨 내용을 주제로 하는가에 따라 학습자의 점수가 민감하게 달라질 수 있는 위험 요소가 내재되어 있음을 보여준다. 따라서 학습자의 능력이 제대로 발휘될 수 있도록 하기 위해서는 cloze 본문 선정에 있어, 학습자 집단의 특성, 관심, 선호하는 주제 등을 세심하게 고려하여야 할 것으로 보인다.

넷째, 학습자들의 설문 응답은 cloze 테스트가 단어, 문장, 담화 수준의 능력을 측정한다는 것을 학습자들이 인지한다는 것을 시사하였는데, 이는 cloze 점수를 토대로 한 본 연구의 요인분석 결과와 수렴되는 결과라 할 수 있다. 그러나 본 설문에서는 어떤 능력이 요구되는지를 직접 서술하도록 하는 개방식 문항 대신, 구체적 진술을 주고, 이에 대한 인지 여부만 표시하도록 하는 형식을 취하였으므로, 만일 학습자 스스로가 측정되는 능력이 무엇인지 직접 말하도록 요구했다면 동일한 구인이 표출되지 않았을 수도 있다. 이러한 점은 본 연구의 한계점이며, 결과 해석에 주의를 기할 필요가 있다.

본 연구의 결과들을 종합해 볼 때, cloze 테스트는 구인 타당도와 안면 타당도 면에서 어느 정도 타당한 평가 도구라 결론 내릴 수 있다. 비용과 시간, 제작의 용이성 등 많은 장점을 가진 cloze 테스트가 그 하부 구조뿐 아니라 학습자의 시각에서 볼 때에도 어느 정도 타당하다면, 이 평가도구에는 적지 않은 잠재력이 내재되어 있다고 하겠다. 그러나 cloze 테스트로 공인시험을 대체하는 것은 쉽지 않을 것이다. 영어 능숙도의 중요한 요소인 발음이나 유창성, 상호작용 능력, 현장성과 같은 요소들은 일정 시간을 주고 지필로 측정하는 cloze로는 측정하기가 어렵기 때문이다. 따라서 교육 및 연구에서 이 평가도구를 활용함에 있어 이러한 제한점을 고려하여야만 할 것이다.

참고문헌

신상근 (2010). *외국어 평가의 이론과 실제*. 서울: 한국문화사.

Abraham, R. G., & Chapelle, C. A. (1992). The mesning of cloze test scores: An item diffiulty perspective. *The Modern Language Journal, 76*(4), 468-479.

Alderson, J. C. (1979). The cloze procedure and proficiency in English as a foreign language. *TESOL Quarterly, 13*(2), 219-227.

Alderson, J. C. (1980). Native and non-native speaker performance tests. *Language Learning, 30*(1), 59-76.

Anderson, J. (1976). *Psycholinguistic experiments in foreign language teaching.* St. Lucia, Queensland, Australia: The University of Queensland Press.

Bachman, L. F. (1982) The trait structure of cloze test scores. *TESOL Quarterly, 16*(1), 61-70.

Bachman, L. F. (1985). Performance on cloze tests with fixed-ratio and rational deletions. *TESOL Quarterly, 19*(3), 535-555.

Bachman, L. (1990). *Fundamental consideration in language testing.* Oxford: Oxford University Press.

Bormuth, J. R. (1965). Validities of grammatical and semantic classifications of cloze test scores. In J. A. Figurel (Ed.), *Reading and inquiry* (pp. 283-285). Newark, DE: International Reading Associates.

Bormuth, J. R. (1967). Comparable cloze and multiple-choice comprehension tests scores. *Journal of Reading, 10*, 291-299.

Bormuth, J. R. (1968) The cloze readability procedure. *Elementary English, 45*(4), 429-436.

Brown, H. D. (2007). *Teaching by principles: An interactive approach to language pedagogy* (3rd ed.). White Planes, NY: Pearson Education.

Brown, J. D. (1980). Relative merits of four methods for scoring cloze tests. *The Modern Language Journal, 64*(3), 311-317.

Brown, J. D. (1983). A closer look at cloze: Validity and reliability. In J. W. Oller (Ed.) *Issues in language testing research* (pp. 237-250). Rowley, MA: Newbury House.

Brown, J. D. (2002). Do cloze test work? *Second Language Studies, 21*(1), 79-125.

Chihara, T., J., Oller, K. W., & Chavez-Oller, M. (1977). Are cloze items sensitive to constraints across sentences? *Language Learning, 27*(1), 63-70.

Cohen, A. D. (1980). *Testing language ability in the classroom*. Rowley, MA: Newbury House.

Crawford, A. (1970). *The cloze procedure as a measure of reading comprehension of elementary level Mexican-American and Anglo-American children.* Unpublished doctoral dissertation, University of California Los Angeles, Los Angeles.

Gaillard, S., & Tremblay, A. (2016). Linguistic proficiency assessment in second language acquisition research: The elicited imitation task. *Language Learning 66*(2), 419-447.

Hale, G. A., Stansfield, C. W, Rock, D. A., Hicks, M. M., Butler, F. A., & Oller, J. W. (1989). The relation of multiple-choice cloze items to the Test of English as a Foreign Language. *Language Testing, 6*(1), 47–76.

Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics.* New York: Newbury House.

Hinofotis, F. B. (1987). Cloze testing: An overview. In M. H. Long & J. Perking (Eds.), *Research in language testing* (pp. 121-128). Rowley, MA: Newbury House.

Jonz, J. (1987). Textual cohesion and second language comprehension. *Language Learning, 37*(3), 409-38.

Jonz, J. (1976). Improving on the basic egg: The MC cloze. *Language Learning, 26*(2), 255–265.

Khodadady, E. (2007). C-tests method specific measures of language proficiency. *Iranian Journal of Applied Linguistics, 10*(2), 1-26.

Khodadady, E. (2014). Construct validity of C-tests: a factorial approach. *Journal of Language Teaching and Research, 5*(6), 1353-1362.

Kim, P. (2014). A study on EFL college students' reading and integrated writing ability and their reading strategy use. *English Language & Literature Teaching, 20*(3), 119-137.

Kongsuwannakul, K. (2017). *Investigating the construct validity of a concordance-based cloze test: a mixed-methods study*. Unpublished doctoral dissertation. University of Leicester, Leicester.

Lee, J.-W. (2002). An analysis of test-taking strategies for the cloze. *English Teaching 57*(1), 213-237.

Lee, S. (1997). Cloze test as a measure of EFL writing proficiency. *English Teaching, 52*(3), 151-172.

Oller, J. W. (1979). *Language tests at school: A pragmatic approach*. London: Longman.

Oller, J. W., & Conrad, C. A. (1971). The cloze technique and ESL proficiency. *Language Learning, 21*(2), 183–195.

Oller, J. W., & Jonz, J. (1994). Why cloze procedure. In J. W. Oller, Jr., & J. Jonz (Eds.), *Cloze and coherence* (pp. 1-20). Cranbury, New Jersey: Bucknell University Press.

Rankin, E. F. (1970). The cloze procedure: Its validity and utility. In R. Farr (Ed.), *Measurement and evaluation of reading* (pp. 237–253). New York: Harcourt, Brace and World.

Rankin, E. F., & Culhane, J. W. (1969). Comparable cloze and multiple-choice comprehension scores. *Journal of Reading, 13*(3), 193–198.

Robinson, E., & Dicken, P. (1979). Closure procedure and cognitive mapping. In *Environment and Behavior, 11*(3), 351-373.

Saeedi, M. (2016). Construct validity of multiple-choice cloze test and cloze-elide test in testing reading comprehension among Iranian EFL learners. *International Research Journal of Applied and Basic Sciences, 10*(6), 632-637.

Stubbs, J. B., & Tucker, G. R. (1974). The cloze test as a measure of English proficiency. *Modern Language Journal, 58*(5), 239–241.

Taylor, W. L. (1953). Cloze procedure: a new tool for measuring readability. *Journalism Quarterly, 30*(4), 414-438.

Taylor, W. L. (1957). Cloze readability scores as indices of individual differences in comprehension and aptitude. *Journal of Applied Psychology, 41*(1), 19–26.

예시언어(Examples in): English, Korean
적용가능 언어(Applicable Language): English
적용가능 수준(Applicable Levels): Secondary, tertiary

한혜령
서원대학교 사범대학 영어교육과
28674 충북 청주시 서원구 무심서로 377
Email: hyerhahn@seowon.ac.kr

Received in October 16th, 2018
Reviewed in November 27th, 2018
Revised version received in December 5th, 2018