

규칙빈칸메우기 시험의 문항수에 관한 연구

최 용 재

1. 서 론

규칙빈칸메우기 시험(cloze test)은 Wilson Taylor가 1953년에 이른바 “cloze procedure”라는 이름으로 *Journalism Quarterly*에 소개한 이래, 이독성(readability)을 측정하는 데 있어서 새롭고 효과적인 도구로 인식되어 왔다. 원래 이 도구는 측정의 대상인 단락(passage)이 학습자에게 읽히기에 쉬운가 아니면 어려운가를 판단하는 도구였으나, 외국어로서의 영어교육(TEFL)에 적용되면서부터는 단락의 난이도가 아니라 그 단락을 이용하여 학습자의 언어능력을 측정하는 일종의 평가도구로 사용하게 되었다. 이 측정도구는 형태심리학(Gestalt psychology)의 이론적 뒷받침과 측정도구의 제작과 시행이 쉽다는 이점 때문에 큰 호응을 받아왔다. 다만 삭제할 단어의 수에 대한 본격적 연구가 근년에 이르러서야 겨우 시작되었을 뿐이어서(참고: Sciarone and Schoorl 1989) 이에 대한 연구가 크게 요청된다. 따라서 이 논문은 영어가 외국어(EFL)인 상황하에 있는 한국인 학생들을 대상으로 할 때, 규칙빈칸메우기 시험에 있어서 삭제할 단어의 수는 최소한 몇 개인가를 규명하려는 하나의 시도이다.

이독성이라는 것은 워낙 광범한 내용을 포함하고 있는 측정기준으로 독서자의 순수한 언어지식이나 언어능력 밖에도 문화적 상식과 이해는 물론이고 학문적 이해가 함께 작용하는 복합적 독서행위를 가능케 하는 능력이라고 할 수 있다. 이독성에 대한 이런 이해가 옳다고 하면 Wilson Taylor가 소개한 규칙빈칸메우기 시험은 주어진 단락을 읽어 나갈 수 있는 학습자의 복합적 능력을 진단하려는 것임에 틀림이 없다. 특히 이 시험은 단락 자체가 이미 출판된 원문에서 잘라온 것이므로 음성언어능력이 아닌 문자언어능력의 측정이 될 수밖에 없다. 그러나 원문의 의미내용에 따라서는 그 분야의 전문적 지식은 아닐지라도 어느 정도의 상식이 요구되리라는 것은 자명한 일이다. 더구나 언어는 문화와 불가분의 관계에 있기 때문에 문화와는 중립인 전문학술적 내용의 단락이 아닌 한에는 필연적으로 문화적 지식의

바탕 위에 내용을 파악하고 이해해야 한다(최용재 1990). 이런 의미에서 볼 때, 규칙빈칸메우기 시험은 학습자의 종합적 독서능력을 측정하는 것이라고 볼 수 있겠다.

이런 이독성을 측정하는 도구가 EFL현장에 소개되면서, 규칙괄호메우기 시험이 언어능력을 종합적으로 측정할 수 있다는 주장을 낳게 하고 또 그런 믿음을 갖게 만든 과정이나 현실은 누구나 잘 아는 사실이다. 여기서 주의할 것은 EFL현장에서 “종합적”이라는 말이 이독성에서의 종합적이라는 말과는 다른 뜻이라는 점이다. EFL에서 종합적이라는 말은 언어능력 가운데서 지식이나 기능의 한 부분을 측정하는 것이 아닌 언어지식과 능력 및 언어의 네 기능을 모두 한 데에 묶어서 측정한다는 뜻으로 쓰이는 경우가 보통이다. 더구나 몇몇 학자들이 행한 일련의 연구와 주장은 규칙괄호메우기가 언어의 개별기능 시험들과 아주 좋은 상관관계를 보이고 있음을 보고하고 있어서(대표적인 예는 Oller and Conrad 1971), 기존의 TOEFL이나 Michigan Test를 대체할 수 있는 일종의 “super-test”로 인식되는 경향마저 있다. 출제의 용이성과 시험결과의 높은 신뢰도가 이 측정도구의 가장 큰 장점으로 누구나 쉽게 마련하고 사용할 수 있다는 실용성이 뒤따른다는 점은 이미 지적한 바와 같다.

그러나 동전에는 앞과 뒤가 있듯이 이런 장점 가운데에도 단점 또는 문제점이 내포되어 있음을 본다. 무엇보다 먼저 “출제가 쉽다”는 말은 비전문가도 기계적인 절차만 따르면 쉽게 평가도구를 마련할 수 있다는 말인데, 그러다가 보니 문항 하나하나가 언어능력 평가에 얼마나 기여하는지가 의심스럽게 되었다. 다음으로 “신뢰도가 높다”는 것도 전체 언어능력을 부분적 언어기능과 대비하여 상관도를 측정한 것이므로 한 사람이 가지고 있는 언어능력이 기능에 따라서 심한 불균형을 이루고 있지 않는 한 좋은 상관계수를 보이는 것은 당연한 일이다. 끝으로 “실용성의 문제”도 시험을 치르는 사람의 숫자가 적을 때는 문제가 없지만 수험생의 숫자가 천명, 만명 등으로 늘어났을 때에는 수작업에 의한 채점에 의지한다는 것은 비실용적이다. 규칙빈칸메우기 시험의 이런 여러 문제점들 가운데서 맨 먼저 지적한 문제의 해결이 이 시험의 성패를 좌우하는 관건이다. 이 문제를 해결하는 방법은 전통적으로 실시해 온 문항수 늘리기인바, 이 논문이 규명하고자 하는 것이 바로 이 문항수 늘리기의 적정선이다.

2. 배경과 이론

Wilson Taylor가 규칙빈칸메우기 시험을 처음으로 소개한 논문의 맨 처음 문단은 비록 두 개의 문장으로 이루어져 있으나 시사하는 바는 크다.

“Cloze procedure” is a new psychological tool for measuring the effectiveness of communication. The method is straightforward: the data are easily quantifiable; the findings seem to stand up (Taylor 1953).

이 짝막한 문단에서 발견할 수 있는 주장은 세 가지인데, 이 시험이 (1) 의사소통의 효과를 측정할 수 있고, (2) 채점하기가 쉬우며, (3) 시험결과가 믿을 만하다는 것들이다. 이 측정도구가 TEFL에 소개되면서는 이 세 가지 주장은 각각 (1) 종합적 언어구사능력을 측정할 수 있고, (2) 채점방법에 따라서 결과가 다를 수 있으며, (3) 시험결과가 나타내는 것이 무엇인지 확실하지 않을 수도 있다는 등으로 이해되기에 이르렀다.

2.1. 예측문법과 잉여자질

“종합적 언어구사능력”이라는 말은 원래 의사소통을 목적으로 하는 언어 행위를 종합하는 포괄적인 의미를 가지고 있는 것으로 이해된다. 그러나 여기에서는 문자언어를 통한 의사소통에 국한하므로, 규칙빈칸메우기 시험을 치르는 행위는 원문에 기록으로 나타난 내용과 표현형식을 이해하고 그에 대하여 반응하는 행위가 전부이다. 더욱이 문자언어로 의사소통을 하기 위하여서는 표정이나 몸짓과 같은 비언어적 방법의 도움을 받지 않고서도 언어로만 소통의 목적을 충분히 달성하는 장치가 필요한데, 이를 위하여서는 문자언어 특유의 표현방식이 따로 있을 수 있다. 그러므로 빈칸메우기라는 문제를 해결하기 위하여서는 일반적 언어지식을 동원할 뿐만 아니라 문자언어에 알맞은 표현방식도 동원해야 함은 물론이다. 그러나 이 문제해결은 단순한 언어지식만을 가지고는 불가능할 것임은 글을 소리내어 읽을 수 있다고 해서 국민학생이 성인용의 문서를 이해할 수 없음에서 분명해진다. 따라서 규칙빈칸메우기 시험문제를 이해하기 위하여서는 선행지식(prior knowledge)의 도움이 절대 필요하다.

그러나 규칙빈칸메우기 시험이 EFL에서 언어능력평가로 사용되면서 단순한 언어능력만을 측정하는 것으로 그 범위가 한정되어 언어외적인 선행 지식은 언급도 되지 않을 뿐만 아니라 문자언어와 음성언어의 차이마저도 주목되지 않고 있는 실정이다. 따라서 규칙빈칸메우기 시험에서의 종합적 언어구사능력은 순수언어능력을 지칭하는 것으로 간주되는 경우가 많다. 그런데 문제는 언어능력이라는 것의 정의부터 명확하지가 않아서 학자에 따라 저마다 다른 실체로 이해하고 있다(최용재 1990). 언어를 구사한다는 입장에서 볼 때, 규칙빈칸메우기에서 언급되는 언어능력은 정적인 요소로만 이루어져 있는 Chomsky의 순수언어능력이 아닌 역동적 요소를 포함하고

있는 것이라야 한다. 여기서 “역동적”이라고 하는 뜻은 순수언어능력을 초월하여 그 능력을 능동적으로 수행하여 나간다는 것이다. 결국 규칙빈칸메우기 시험에서는 의사소통능력을 포함한 역동적 언어능력이 작용하는 것으로 간주되어야 한다.

이런 언어능력이 규칙빈칸메우기 시험에서 하는 일은 잉여자질을 활용하여 그 문장의 결손부분을 재생하는 것이다. 이 일을 해내기 위하여서는 결손부분의 전후에 있는 단어와 그에 따른 문법관계를 고려하는 작업이 필수적이라고 생각된다. 사실 결손이 일어난 문장 자체 안에서 문제를 해결하는 일이 70.0%라는 보고가 있으나, 아울러 문장의 한계를 벗어나는 경우도 10.8%나 된다고 한다(Henk et al. 1985). 특히 문제해결을 위하여 문장의 한계를 넘어선다는 것은 단락 전체와 연관하여 판단한다는 증거로 볼 수 있는바, 이렇게 되면 언어외적인 선행지식이 5.2%나 이용된다는 주장이 뒤따르게 된다(Henk et al. 1985). 이를 다시 정리해 보면, 한 문장에 국한하여 문제를 해결할 수 있을 때는 물론 그렇게 하겠지만, 그것이 여의치 아니할 때는 문장의 범위를 넘어서서 이웃하는 문장이나 문단 또는 전체 단락에 연관하여 결손된 부분에 알맞은 낱말을 생각해 내는 것으로 판단된다. 그러나 한 문장 안에서도 결손된 부분이 여러 개 있을 수 있으므로 결손부분의 전후관계를 보아 판단하는 일에도 한계가 있다.

대개의 경우에 수험생은 앞에서부터 차례로 문제를 풀어가므로, 하나의 결손부분이 있다고 할 때 그 결손부분의 뒤보다는 앞을 이용하여 판단하는 일이 더 많을 것임을 쉽게 상상할 수 있다. 다시 말하자면, 문제의 해결은 글을 읽어 나가는 방향으로 나아갈 것이므로 판단의 방향성이 설정될 수 있다는 것이다. Oller(1975)에 의하면 수험생은 예측문법(grammar of expectancy)으로 문제를 해결한다. 읽은 내용을 바탕으로 하여 다음에 올 내용을 예측하고 확인하는 과정이다. Oller는 바로 이 능력을 예측문법이라고 보고, 그 예측문법을 규칙빈칸메우기 시험이 측정한다고 주장한다. 물론 시험대상으로서의 예측문법의 내용이나 구성요소는 분명하지 않지만, Oller나 그의 추종자들이 생각했을 순수언어능력 내지는 언어구사능력만 있는 것이 아니라 선행지식도 포함될 가능성이 크다. 다시 말하자면, Oller의 예측문법도 문법인 이상 내용이 있을 것임에는 틀림이 없겠으나 현재까지 알려진 바는 전혀 없어서, 언어능력 이외에 다른 능력이나 지식이 포함될 것이라는 추측만이 가능하다.

여기서 생각해 볼 것은 시험문제의 표집(sampling)이다. 규칙빈칸메우기 시험은 단어의 규칙적이고 기계적인 삭제에 의존하여 무작위에 의한 표집을 하고 있으므로 어떤 특정한 언어항목을 의도적으로 출제하기는 어렵다. 따라서 예측문법의 어떤 내용이나 어느 부분을 측정하는지가 분명하지 않

고 잉여자질에 의한 추측도 계획적으로 주어지는 것이 아니라서 수험생이 어떤 책략에 의하여 정답을 추측해내는지를 알 길이 없다. 물론 언어사용은 상황에 따라서 다를 수가 있어서 중요하지 않은 언어표현이 되풀이될 수도 있고 필요없는 반복이 있을 수가 있지만, 시험은 경제적이고 집중적이어야 하므로 무계획하고 산만하며 필요없는 반복만 있는 시험을 좋다고 판단하기는 어렵다. 이런 이유로 말미암아 규칙빈칸매우기 시험에서 문항의 수는 되풀이되는 부분을 빼고 나면 그 수가 상당히 줄어들 것이다. 더구나 반복되는 문항의 수가 일정하지 아니하므로, 규칙빈칸매우기 시험이 참으로 믿을 만한 측정도구가 되려면 반복되고도 충분할 만큼 많은 문항, 즉 빈칸의 확보는 필수적이다.

2.2. 채점방법

규칙빈칸매우기 시험의 채점은 아주 쉽다는 것이 정평인 듯하다. 또한 채점시에 경험이 없거나 훈련을 받지 않은 채점자라도 객관성을 유지하기가 쉽기 때문에 채점자의 신뢰도가 완벽하다는 장점이 있다(Henk and Selders 1984). 그러나 출제자가 삭제한 낱말을 그대로 고스란히 재생하지 못했다고 해서 감점을 하는 것은 너무나 가혹한 채점방법이란 말을 듣게 된다. 예를 들면, “a (beautiful) girl”이라고 할 것을 “a (pretty) girl”이라고 답을 할 경우를 놓고 이를 틀렸다고 할 수가 있겠느냐는 말이다. 더구나 언어사용의 실재를 보면 표현하고자 하는 하나의 의미에 대응하는 표현, 즉 언어형태는 다양하다. 그런데 그 가운데서 출제된 단락의 저자가 어찌면 우연히 썼을 수도 있는 특정한 표현만이 올바르다고 주장하는 일을 옳다고 할 수는 없을 것이다. 이런 이유들 때문에 단락에서 삭제된 낱말을 원래대로 재생하도록 하는 지나치게 엄격한 시험채점방법에 대한 비판이 높아지자, 채점에 있어서 판정의 유연성이 요구되기에 이르렀다.

삭제된 낱말을 원상복구토록 하는 경직된 채점방식을 개선하는 방법은 간단하다. 문장의 맥락과 의미내용 그리고 전체 단락에 비추어 보아서 주어진 빈칸에 들어갈 수 있는 낱말들은 모두 옳다고 간주하면 된다. 물론 이 방법에도 장단점이 있다. 우선 장점을 들어보면, 주어진 빈칸에 들어갈 수 있는 여러 가지 낱말들이 있을 수 있다는 사실을 인정함으로써 언어사용상의 우회적 표현현상을 인정할 뿐만 아니라, 낱말이 삭제된 문장의 범위를 넘어서 전체 단락에 대한 다양한 해석을 허용하여 시험 자체가 역동적이 된다. 더욱이 이렇게 될 경우에는 수험생은 자신의 경험에 입각하여 해답을 하고 그에 대한 점수를 받기 때문에 어느 모로 보거나 공정하고 바람직하다(Henk and Selders 1984). 단점으로는 삭제된 낱말 하나에 여러 개의 낱말

이 정답으로 나올 수 있으므로 잘 훈련된 채점자라도 판정에 있어서 어려움을 겪게 된다. 따라서 반드시 전문가가 채점을 하여야 하는 어려움이 있을 뿐만 아니라 채점에 너무 많은 시간이 소요된다는 비실용적인 면이 있다.

위에서 든 두 가지 채점방법을 정리해 보면, 삭제된 낱말을 정확하게 재생하도록 하는 정확단어 채점방법(exact word scoring method or verbatim scoring method)과 맥락이나 의미내용에 알맞으면 옳다고 간주하는 허용단어 채점방법(acceptable word scoring method or synonymic scoring method)이 된다. 이 두 가지 채점방법 가운데서 어느 쪽이 더 바람직한지는 실험을 통하여 증명되고 있다. Oller의 다음 두 인용은 시사하는 바가 크다.

Moreover, the data show that the acceptable word scoring method is superior in terms of item discrimination and validating correlations regardless of the level-of-difficulty of the test (Oller 1972).

The data supported the conclusion that with non-native speakers the method of allowing any contextually acceptable response is significantly superior to the exact word scoring technique (Oller 1973).

여기서 주목할 것은 본토박이가 아닌 수험생에게는 허용단어 채점방법이 훨씬 더 좋다는 주장이다. 상식적으로 생각해 보아도 본토박이도 잘 알아맞히지 못하는 답을 본토박이가 아닌 수험생에게 요구한다는 것은 무리이다.

“본토박이가 아닌 수험생에게는 허용단어 채점방법을 사용한다”는 주장은 당연히 다른 학자들의 호응을 얻기 마련이다(Hinofotis 1980). 그러나 Irvine et al.(1974)에 의하면 정확단어 채점방법과 허용단어 채점방법 사이의 상관계수는 .94의 높은 수치를 보여, 두 가지 채점방법 가운데 어느 것을 택하여도 좋다는 해석을 가능하게 한다. 만일 이 상관계수가 믿을 만하다면, 구태여 두 가지 다른 채점방법을 놓고 어느 방법을 택할지 걱정할 필요가 없을 것이다. 다만 바로 위에서 든 바와 같은 이유들은 높은 상관계수가 보이는 수치와 무관한 것들로 어떻게 하든 설명되거나 개선되어야 할 점들이다. 실제로 Irvine et al.의 실험에서도 두 가지 채점방법이 받아쓰기 시험과의 상관관계에서 “정확단어 .69”와 “허용단어 .75”라는 차이를 보여서 무엇인가 서로 차이가 있음을 시사하고 있다. 더구나 Irvine et al.의 이 실험이 50개라는 한정된 문항수의 범위 안에서 행하여진 것이라면 더 적은 문항수에서나 더 많은 문항수에서는 어떻게 될지 의문이다.

2.3. 시험문항

규칙빈칸메우기 시험의 결과가 나타내는 것이 무엇인지에 대한 질문은

곧바로 이 시험이 측정하는 바가 무엇인지에 대한 질문으로 이는 타당성의 문제이다. 문항마다 별도의 측정목표가 있을 수 있는 분리시험(discrete-point test)에 익숙해 있는 사람에게는 종합시험(integrative test)이라는 주장과 함께 소개된 규칙빈칸메우기 시험의 문항 하나하나가 구체적으로 무엇을 측정하며 그 시험의 결과가 전체적으로 무엇을 뜻하는지를 이해하기가 어려울 것이다. 더구나 낱말을 규칙적 간격을 두고 기계적으로 삭제해서 만든 시험의 문항들이 언어기능이나 언어항목을 균형있게 측정하리라고 기대하기는 어렵다. 좀 길지만 다음의 인용이 이를 잘 설명해 줄 것이다.

When, following the cloze procedure, every *n*th word deleted from a piece of text, the resulting blanks may not be equally appropriate as items in a language-proficiency test. The major contaminating factor in this connection is the peculiar frequency distribution of linguistic elements, with few elements occurring very frequently in arbitrary texts and many elements occurring only rarely, and only in specific texts.

In short, mechanical application of the cloze procedure to an arbitrary piece of text may well produce a set of items that would be both too small and too selective to constitute a representative sample of the total set of grammatical and lexical elements whose mastery one seeks to measure (Sciarone and Schoorl 1989).

바로 이런 이유 때문에 규칙빈칸메우기 시험이 맨 처음 소개될 때부터 그 개선방법에 대한 의문이 제기되었다. 또한 이 논문에서도 이를 이미 지적한 바 있다.

Wilson Taylor(1953)는 전체 빈칸이 “모든 종류의 낱말들을 균형있게 포집할 수 있을 만큼 충분한 삭제가 필요하다”면서 “How many blanks are enough?”이란 질문을 던졌다. 그는 1956년에 발표한 논문에서 50개의 빈칸이면 대체로 충분할 것이라고 결론짓고 있는데, 그의 이 주장은 규칙빈칸메우기 시험을 외국어 특히 한국어에 적용해 보고 도달한 것이라는 점에서 EFL에서도 받아들여질 수 있는 주장인 듯하다. Taylor의 이 주장은 뒤에 TEFL에 적용되면서도 받아들여져 50개의 빈칸이 표준인 것처럼 되었다. (예로서 Oller and Conrad 1971, Aitkin 1977을 들 수 있다.) 그러나 Sciarone and Schoorl(1989)는 채점방법에 따라서 최소한으로 필요한 빈칸의 수를 제시했는데, 정확단어 채점의 경우에는 적어도 100개 단어를, 그리고 허용단어 채점에서는 적어도 75개 단어를 삭제해야 한다는 주장이다. 만일 Sciarone and Schoorl의 주장이 옳다면 50개 빈칸을 표준으로 삼았던 때

에 비하여 더 긴 단락의 시험문제로 더 긴 시간에 걸쳐서 측정을 해야 한다는 점에서 규칙빈칸매우기 시험의 경제성에 어느 정도의 흠이 된다고 하겠다.

지금까지 서술한 내용을 종합해 보면, 규칙빈칸매우기 시험은 TEFL에 처음 소개될 때의 기대이상으로 단순한 언어능력을 초월한 더욱 더 광범한 의미의 의사소통능력을 측정하는 것 같다. 다만 문장에서 삭제된 부분들이 분리시험에서와 같이 전체적 문항의 의도적인 균형을 이룰 수가 없어서 반복되는 문항이 많으므로 Taylor의 말을 따라 균형이 있는 출제를 하려고 하면 상당히 많은 수의 낱말들을 삭제해야 한다는 번거로움이 있다. 따라서 이 번거로움을 해소하려는 노력의 하나로 삭제할 단어의 수가 최소한 몇 개 일지를 알아낼 필요가 있다. 물론 이미 어느 정도 공인된 것이나 다름이 없는 숫자가 있기는 하지만, 더욱 최근에 나온 Sciarone and Schoorl의 주장이 상당한 설득력이 있는 반면에 Irvine et al.의 주장과 맞지 않은 면도 있다. 따라서 Sciarone and Schoorl 그리고 Irvine et al.의 주장들을 검증해 볼 필요가 있고, 그런 과정에서 이 논문 나름대로의 새로운 발견과 결론에 도달할 수 있을 것이다.

3. 실험방법

3. 1. 실험대상인원

실험에 자발적으로 참가하려는 학습자를 많이 확보하기가 매우 어려워서 겨우 확보한 인원은 25명에 불과했다. 이들은 모두 동국대학교 영어영문학과 3학년생들로 남여가 절반씩 정도로 섞여 있었으나, 성별의 구분은 큰 의미가 없으므로 통계상으로는 무시하기로 했다. 이들 모두는 대학의 3학년에 재학중이므로 영어교육을 도합 8년 정도를 받은 배경을 가지고 있었다. 이들이 규칙빈칸매우기 시험에 처음으로 노출된 것은 이 실험에 참가하기보다 한 학기전에 한 번 있었고 그 뒤로는 별다른 연습을 한 바가 없었으며, 이 시험을 위해서는 별다른 준비를 한 바가 없었다.

3. 2. 실험자료

실험을 위하여 원문의 단락은 어휘와 문장이 점점 어려워지는 것들로 여섯 개를 추출하되 한 단락에 25개의 낱말을 삭제하는 것으로 표준을 삼았다. 낱말의 삭제는 여섯 단어마다 기계적으로 한 단어씩 삭제하되 각 단락의 시작과 끝에 있는 문장들을 길이에 따라서 한 개 또는 두 개 정도 손을 대지 않고 그대로 두어 수험생이 삭제된 낱말들을 재생하는 데 도움이 되게

했다. 그러나 막상 낱말들을 삭제해 놓고 보니, 기능어가 상당히 많은 수를 차지했다. 삭제된 내용은 다음 <표 1>과 같다.

<표 1> 시험 문항구성

문 제	소계 누계	문단수	문장수	단어수	문 항 수		
					기능어	내용어	합 계
1		3	22	216	10	15	25
		3	22	216	10	15	25
2		3	18	196	10	15	25
		6	40	412	20	30	50
3		2	16	212	11	14	25
		8	56	624	31	44	75
4		3	15	198	10	15	25
		11	71	822	41	59	100
5		2	11	203	8	17	25
		13	82	1025	49	76	125
6		3	12	208	9	16	25
		16	94	1233	58	92	150

3.3. 실험절차

시험은 오전중에 90분에 걸쳐 모두 한꺼번에 실시했다. 시간이 부족하여 문제를 다 풀지 못한 것 같지는 않았고 두 사람이 끝부분을 마무리 짓지 못했는데 시간이 없어서라기보다는 아마 자신이 없어서 포기했던 것 같다. 채점방법은 정확단어 채점방법과 허용단어 채점방법을 다 사용하였다. 두 가지 채점방법 모두에서 철자상의 오류는 용인하였고, 허용단어 채점방법에서 허용단어의 범위를 단락 전체의 의미에 크게 어긋나지 않은 것만을 정답으로 했다. 결과의 통계는 시험에 제시된 차례에 따라서 첫째 단락부터 둘째, 셋째 등으로 나아가면서 누계로 처리했다. 통계는 신뢰도의 경우는 Kuder-Richardson formula를 그리고 정확단어 채점결과와 허용단어 채점결과와의 상관계수는 Spearman rank order correlation을 사용하여 계산했다.

4. 결과와 논의

채점결과를 놓고 보면, 수험자의 수가 워낙 적기 때문에 점수의 분포가 약간 고르지 못하고 집중치(mode)가 이중으로 나오는 것을 볼 수 있었다. 특히 시험문항수가 적을수록 이런 경향이 뚜렷했다. 다음 <표 2>는 시험결

과의 집중경향(central tendency)과 문항수의 누진에 따른 신뢰도(reliability)를 보인 것이다.

〈표 2〉 집중경향 및 신뢰도

문항수	25		50		75		100		125		150	
채점방법	EW	AW	EW	AW	EW	AW	EW	AW	EW	AW	EW	AW
범위	8	10	11	16	18	29	28	40	39	44	48	61
중앙치	11	19	20	35	28	46	39	62	47	70	52	81
산술평균	11	19	20	34	29	47	40	61	46	69	52	79
표준편차	2.2	2.3	2.9	3.6	4.8	6.1	6.8	8.6	9.1	11.4	11.9	15.3
신뢰도	-.23	-.17	-.49	.17	.23	.53	.48	.68	.65	.76	.76	.84

(주의: EW=Exact Word Scoring Method, AW=Acceptable Word Scoring Method)

범위(range)를 보면, 25문항 시험의 경우에 8점에서 10점 사이로 전체 문항수의 1/2 수준이고, 150문항 시험의 경우는 48점에서 61점 사이로 대략 1/3 수준이다. 그러므로 문항수가 적으면 적을수록 범위가 넓어지고 문항수가 많으면 많을수록 범위가 좁아지는 경향을 보인다고 말할 수 있다. 물론 50문항 시험의 경우에 범위가 전체 문항수의 1/5과 1/4의 수준이어서 의외라고 지적할 수도 있겠다. 그러나 문항수가 적을 때에는 문항의 난이도가 변인으로 작용하여 그런 불규칙한 결과를 보일 수 있으나 문항수가 많아지면 난이도의 평준화가 이루어져서, 전체적으로 볼 때 문항이 많으면 많을수록 범위는 안정된다. 정확단어 채점결과와 허용단어 채점결과를 서로 비교해 볼 때 허용단어 채점결과 범위가 더 넓다. 이 경향은 중앙치, 산술평균, 표준편차 등의 모든 집중경향에서 보이고 있는바, 채점방법에 따라서 점수에 엄격하고 유연한 차이가 있으므로 허용단어 채점의 결과가 더 후한 점수를 내보인 것으로 판단된다.

중앙치(median)와 산술평균(arithmetic mean)은 거의 일치하고 있어서 전체 12개의 짝 가운데서 4개가 완전히 일치하고 나머지 7개의 짝은 1점의 차이가 있으며, 마지막 1개의 짝만이 2점의 차이가 있다. 더구나 표준편차도 문항수에 따라서 꾸준히 증가하고 있을 뿐만 아니라 수치도 적정선을 보이고 있다. 중앙치 산술평균 표준편차 등의 이런 경향은 이 실험의 성적분포가 고르다는 사실을 반영한다고 볼 수 있다. 다만 표준편차의 경우에 어떤 특정수의 문항을 정확단어 채점방법으로 채점한 결과와 그보다 25개의 문항이 적은 문항수를 허용단어 채점방법으로 채점한 결과가 엇비슷하다는 흥미있는 사실을 발견할 수 있다. 예를 들면 150문항 시험을 정확단어 채점방법으로 채점한 결과와 125문항 시험을 허용단어 채점방법으로 채점한 결

과의 표준편차는 각각 11.89와 11.37로 비슷하다. 이는 정확단어 채점의 결과가 산술평균치에 더욱 근접하여 있다는 것을 나타내는 것인바, 표준편차 사이의 이런 현상은 범위에서 더욱 뚜렷하게 나타남을 볼 수 있다.

이 시험의 신뢰도를 보면, 25문항 시험과 50문항 시험의 경우에는 두 가지 채점방법이 모두 바람직하지 않은 것으로 나타났다. 따라서 “50개 이하의 문항으로 된 규칙빈칸메우기 시험은 믿을 만한 것이 못된다”는 결론을 내릴 수 있다. 두 가지 채점방법 모두에서 신뢰도가 꾸준한 증가를 보이며 우량한 것으로 나타나는 것은 75문항 시험부터서이다. 물론 75문항 시험에서 정확단어 채점방법의 신뢰도는 .23으로 바람직하지 못하지만 그 이후부터 계속하여 상승세를 보인다는 데에 의의가 있다고 하겠다. 위에 보인 통계는 문항수가 늘어남에 따라 신뢰도가 향상됨을 보여 준다. 더욱이 같은 문항수의 시험이라도 채점방법에 따라서 신뢰도가 차이가 나는데, 허용단어 채점방법이 정확단어 채점방법보다 신뢰도가 높음을 알 수 있다. 결국 신뢰도는 문항수가 75개 이상인 시험을 허용단어 채점방법으로 채점할 때가 좋다고 할 수 있다.

다음으로 이 실험의 핵심인 문항수 차이에 의한 상관계수를 보자. 이 실험은 삭제할 낱말의 적정수를 발견하려는 것이므로 다음은 25개의 문항을 간격으로 하여 인접한 문항수의 시험과의 상관계수를 연속적으로 계산한 것이다.

〈표 3〉 문항수대비 상관계수

문항수대비	25 : 50	50 : 75	75 : 100	100 : 125	125 : 150
정확 단어	.77	.63	.92	.92	.99
허용 단어	.44	.77	.93	.94	.98

위 표를 보면, 정확단어 채점의 경우에 “20 : 50”의 .77과 “50 : 75”의 .63은 도치되어 있어서 의심이 가는 만큼 둘 다 채택하기가 어렵다. 반대로 허용단어 채점의 경우에 “50 : 75”에서부터 우량한 상관계수를 보이고 있으나, “75 : 100”의 경우와 비교해 볼 때 확신하기는 어렵다고 하겠다. 따라서 “75 : 100”부터 두 가지 채점방법 모두가 완벽에 가까운 계수인 .90 이상의 수치를 보이고 있으므로 규칙빈칸메우기 시험에서 빈칸의 적정수는 최소한 75개는 되어야 한다는 해석이 가능하다.

5. 결 론

이 실험은 규칙빈칸메우기 시험에서 빈칸의 적정한 수를 발견하려는 것

으로 실험을 통하여 몇 가지 흥미있는 사실을 발견했다. 우선 규칙빈칸메우기 시험이 측정도구로서 믿을 만한 것이 되기 위하여서는 문항수가 상당히 많아야 한다는 것을 알 수가 있다. 집중경향에서부터 그러한 조짐을 보이고 있으며, 신뢰도와 상관계수에서도 마찬가지로 빈칸이 많은 것이 바람직한 것으로 나타난다. 신뢰도의 경우만을 놓고 보면 허용단어 채점방법의 경우에 75문항 시험에서 .53이고 정확단어 채점방법의 경우에 100문항 시험에서 .48로 각각 양호한 편의 수치를 보이고 있어서, 앞에서 이미 든 Sciarone and Schoorl의 “허용단어 채점방법으로는 75개, 정확단어 채점방법으로는 100”이라는 주장을 뒷받침하는 듯하다.

그러나 상관계수의 경우만을 놓고 보면, 삭제한 단어의 수와는 무관하게 가장 낮은 수치라고 해도 .44이어서 판단을 어렵게 한다. 그러나 이 실험결과를 전체적으로 볼 때 많은 빈칸을 요구하는 이유가 다음과 같이 충분하다. 첫째, 빈칸은 많으면 많을수록 잉여자질과 맥락에 의한 예측이 용이해져서 문항의 난이도가 평준화된 것은 물론이고 집중경향이 안정되어 좋다는 일반적이며 절대적인 경향이 있다. 둘째, 신뢰도에 있어서는 50문항까지도 부정적이므로 반드시 그 이상은 되어야 한다. 셋째, “75:100”에서 정확단어 채점방법과 허용단어 채점방법이 근접한 상관계수를 보이고 있다. 따라서 이 모든 점을 종합하여 판단하면, 규칙빈칸메우기 시험의 빈칸은 최소한 75개는 되어야 하고 그 이상으로 넘어가서 더 많을수록 좋다는 결론에 도달할 수 있다.

참 고 문 헌

- 최용재(1981) ‘Cloze vs. Supply Type Tests,’ *Dongguk Review* 8-9(동국대학교), 155-166.
- 최용재(1990a) ‘외국어 교습과 문화,’ 언어논총 8(계명대학교), 61-79.
- 최용재(1990b) ‘언어능력의 측정가능성에 대하여,’ 논문집 29(동국대학교), 27-58.
- Aitkin, Kenneth G. (1977) ‘Using Cloze Procedure as an Overall Language Proficiency Test,’ *TESOL Quarterly* 11(1), 59-67.
- Chavez-Oller, Mary A., Tetsuro Chihara, Kelley A. Weaver and John W. Oller, Jr. (1985) ‘When Are Cloze Items Sensitive to Constraints across Sentences?,’ *Language Learning* 35(2), 181-206.
- Goodman, Kenneth (1970) ‘Reading: A Psychological Guessing Game,’ In Harry Singer and Robert B. Ruddell (eds.) *Theoretical Models and Processes of Reading*, Newark, Delaware: International Reading Association, 259-72.

- Henk, William A., John P. Helfeldt, Steven D. Rinehart (1985) 'A Metacognitive Approach to Estimating Intersentential Integration in Cloze Tests,' *National Reading Research Conference Yearbook* 34, 213-18.
- Henk, William A. and Mary L. Selders (1984) 'A Test of Synonymic Scoring of Cloze Passages,' *Reading Teacher* 38(3), 282-87.
- Hinofotis, Frances Butler (1980) 'Cloze as an Alternative Method of ESL Placement and Proficiency Testing,' In John W. Oller, Jr. and Kyle Perkins (eds.) *Research in Language Testing*, Rowley, Massachusetts: Newbury House, 121-28.
- Irvine, Patricia, Parvin Atai and John W. Oller, Jr. (1974) 'Cloze, Dictation, and the Test of English as a Foreign Language,' *Language Learning* 24(2), 245-52.
- Oller, John W., Jr. (1972) 'Scoring Methods and Difficulty Levels for Cloze Tests of Proficiency in English as a Second Language,' *Modern Language Journal* 56(3), 151-58.
- Oller, John W., Jr. (1973) 'Cloze Tests of Second Language Proficiency and What They Measure,' *Language Learning* 23(1), 105-18.
- Oller, John W., Jr. (1974) 'Expectancy for Successive Elements: Key Ingredient to Language Use,' *Foreign Language Annals* 7(4), 443-52.
- Oller, John W., Jr. and Christine A. Conrad (1971) 'The Cloze Technique and ESL Proficiency,' *Language Learning* 21(2), 183-95.
- Sciarone, A. G. and J. J. School (1989) 'The Cloze Test: Or Why Small Isn't Always Beautiful,' *Language Learning* 39(3), 415-38.
- Taylor, Wilson L. (1953) "'Cloze Procedure": A New Tool for Measuring Readability,' *Journalism Quarterly* 30, 415-33.
- Taylor, Wilson L. (1956) 'Recent Developments in the Use of "Cloze Procedure",' *Journalism Quarterly* 33, 42-48, 99.

ABSTRACT

On the Number of Blanks in Cloze Procedure

Yongjae Choe

This paper examines the ideal number of blanks in cloze procedure. 25

subjects who were English major juniors volunteered to participate in the experiment and took six cloze tests each of which consisted of 25 blanks. Both exact word scoring method and acceptable word scoring method were employed and the results of both scoring methods seemed to favor a test of more than 75 blanks. Especially, the K-R reliability excluded the employment of a cloze test of less than 50 blanks. Spearman rank order correlation between "75 blanks vs 100 blanks" marked .92 by exact word scoring method and .93 by acceptable word scoring method. In the light of all these results and their interpretations, it may be concluded that an ideal number of blanks of a cloze test should be at least 75 or more.

〈부록 1〉 규칙빈칸메우기 시험문제

Fill in each blank with a suitable word.

In the Station

Ted is in a large room. It has a high ceiling and a stone floor. On the left there is a row of windows. A sign over each window says, "Tickets and information." On the right there is (A) counter. A woman sells candy (AND) magazines there. Near the counter (THERE) are three rows of square (LOCKERS). A sign over them says, "(CHECK) your luggage here." In the (MIDDLE) of the room there are (SOME) rows of benches. People sit (ON) the benches while they wait (FOR) the train.

There are many (PEOPLE) in the room. All of (THEM) are carrying bags and suitcases (AND) are hurrying. A big fat (MAN) is carrying an armful of (PACKAGES). He is walking rapidly and (IS) puffing loudly. A girl is (LOOKING) at a big chart on (THE) wall. The chart tells when (THE) trains leave and arrive. A (WOMAN) is running and pulling a (CHILD) with her. The child is (HOLDING) a doll and is crying.

(TED) is standing by a large (DOUBLE) door at the end of (THE) room. He is wearing a (BLUE) suit and a red necktie and looks happy. The big clock over the door says five o'clock. It is time for a train to arrive, and many people are coming through the door.

Rail roads – A Great Achievement

More and more people travel by plane every year. Yet railroads are important in America. Railroads built the nation. Jacques Barzun, a noted American (WRITER) and educator, recently wrote an (ARTICLE) about the importance of railroads. (HE) is the Dean of Faculties (OF) Columbia University in New York (CITY). He writes on a variety (OF) subjects. This article appeared in (HOLIDAY) magazine. Here is a summary (OF) it.

Railroads, Mr. Barzun says, represent (THE) first great attempts at modern (ORGANIZATION). They coordinate space, time, matter, (AND) men. Before railroads (at the (END) of the eighteenth century) men (COORDINATED) their efforts only in war, (IN) religious ceremonies, and in the (ADMINISTRATION) of law. These coordinated efforts (LASTED) over only short period of (TIME) and small amounts of space. (A) railroad, however, may require the (COMBINED) effort of workers for a (PERIOD) of years and may stretch (ACROSS) a nation.

When a railroad (IS) built, people can depend upon (THE) train schedule. The train arrives (IN) one town at a certain (HOUR). It arrives in other towns at other fixed times. These times are quite predictable. The train schedule is like the schedule of sun, moon, and stars.

Why the Colonists Came

People have come to American from many nations and for many reasons. Some came for economic advancement, other for adventure, and still others for religious freedom. Some men came for political (REASONS), though few were political refugees. (A) few immigrants came merely out (OF) curiosity. Spaniards who were searching (FOR) the fountain of youth made (THE) first permanent settlement in North (AMERICA). They came to Florida in (THE) late sixteenth century.

The first (ENGLISH) colonists came on a wave (OF) exploration and adventure during the (REIGN) of Queen Elizabeth I. They (LANDED) in Virginia in the late (SIXTEENTH) century, but they had little (IDEA) of living conditions in the (NEW) world. Without any supplies from (ENGLAND), they soon starved. The first (PERMANENT) British colony was Jamestown, Virginia. (IT) started in 1607 and after (A) little time it

became a (SUCCESSFUL) agricultural community. Leaders of religious (GROUPS) from England established colonies in (NEW) England in the 1620s and (THE) 1630s. Englishmen also settled Maryland (AND) Pennsylvania as social experiments Debtors (FROM) England settled in Georgia. If (A) person had no money, another colonist often paid his passage in return for his services for several years. Such workers were called indentured servants and were almost like slaves until they had paid for their passage.

Imagination in Fact and Fiction

We often need to distinguish between fact and fiction. We use facts in making practical decisions. The temperature on Tuesday morning (IS) ten degrees below freezing. We (PUT) on heavier clothing. It is (TWELVE) o'clock. We go home to (LUNCH).

We also use language for (THE) pleasure that we can get (OUT) of it. We can play (WITH) ideas. We can play with (THE) form of expression of ideas. (WE) can play with the sounds (OF) language. Thus we have literature : (OUR) plays, our poetry, our novels, (OUR) short stories, and our legends. (WE) sometimes refer to these forms (AS) fiction, because the characters never (REALLY) lived and the stories never (REALLY) happened. Tall tales; or exaggerated (STORIES), are one of the forms (OF) fiction. In the United State (THE) tall tales of legendary heroes (HAVE) become part of the literature (OF) the country. One of these (IS) the story which you have (JUST) read about Paul Bunyan.

Imagination (PLAYS) a big part in both (THE) practical use of language and the artistic use of language. In the story of Paul Bunyan we saw how the men who cleared the land for farming had fun telling imaginative stories.

Oscar the Fish

Americans like to tell fish stories. Fishermen who like to brag about the big fish they have caught may have started the habit. Each one tells of a (FISH) that was in some way (LARGER) or better or more difficult (TO) catch than any other fish, (OR) that was caught in some (UNUSUALLY) clever way. If you tell (A) story that is just a (LITTLE) bit too good to be

(TRUE), someone will say, “There’s something (FISHY) about that.” Here is a (FISH) story that is slightly different (FROM) the usual kind.

Once out (WEST) in Colorado there lived an (INDIAN) who had a pet fish (NAMED) Oscar. He kept it in (A) tub of water behind his (TENT). This Indian was so lazy (THAT) he got tired of putting (FRESH) water into the tub every (DAY). He decided it might be (EASIER) to teach Oscar to live (WITHOUT) water, and he wondered if (HE) could find a way to (DO) this. He began by keeping (THE) fish out of water for (A) few minutes each day. Gradually he increased the time so that finally Oscar could stay out of water all night if he lay in the wet grass by the tub.

The Settlement of the West

Since the founding of the first colonies in North America the population has been constantly moving westward. Most of the movement has been gradual. In the early days, after (EXPLORER) and traders had shown the (WAY), settlers came slowly to claim (THE) land and build new homes. (TWO) spectacular migrations, however, were the (GOLD) rush and the land rush.

(IN) 1848 gold was discovered in (CALIFORNIA). By that time the Spanish (HAD) explored the region and had (GIVEN) its name, which means “heat (OF) the ovens.” They had established (MISSIONS) along the coast, but they (HAD) not built cities, and few (AMERICANS) had settled there. In the (YEAR) after the discovery of gold (HOWEVER), men cut wagon trails across (THE) prairies and crossed the mountains (ON) horseback and with ox teams. (THEY) went to make their fortunes (IN) California. Many of these adventurers, (OR) “forty-niners” as they were called, (FOUND) the fortunes they had gone (TO) seek, but others lost theirs (AS) soon as they had made (THEM) or they did not make (THEM) at all.

The land rush was a westward movement of another kind. In 1889 Congress declared that land in the recently acquired territory of Oklahoma was open for settlement at noon on April 22.

〈부록 2〉 문항수별 득점 및 순위

성명	25EW	순위	50EW	순위	75EW	순위	100EW	순위	125EW	순위	150EW	순위
	AW	순위	AW	순위	AW	순위	AW	순위	AW	순위	AW	순위
강민	9	19	19	14	27	16	39	13	50	9	57	10
	17	20	35	11	49	10	68	6	83	5	97	4
구은	14	3	21	8	31	7	37	18	41	19	46	19
	32	2	35	11	49	10	57	19	61	20	68	21
김정	10	17	19	14	25	20	36	21	39	21	44	21
	23	1	38	3	50	8	62	11	69	15	76	17
문병	7	24	18	18	26	18	37	18	37	22	37	23
	16	23	36	6	44	19	59	17	59	22	59	22
박민	9	19	18	18	28	14	40	11	40	20	46	19
	17	20	31	20	46	13	60	15	60	21	69	20
박종	11	11	17	21	32	6	44	6	54	4	66	4
	20	6	39	2	60	1	80	1	90	1	107	1
변문	12	6	20	11	29	11	43	8	61	8	59	8
	19	11	32	17	45	16	62	11	74	8	82	12
성일	8	22	15	24	23	23	26	24	28	24	28	24
	14	24	28	24	39	23	46	23	48	24	48	24
심우	14	3	24	2	38	2	48	4	52	7	60	7
	21	3	36	6	53	4	68	6	72	10	81	13
안승	11	11	21	8	30	9	41	10	47	12	52	14
	19	11	33	15	45	16	59	17	65	16	74	19
오경	10	17	18	18	31	7	44	6	54	4	62	6
	20	6	36	6	54	3	71	3	85	2	97	4
유재	7	24	16	23	28	14	37	18	43	17	52	14
	18	16	32	17	49	10	62	11	71	12	83	10
이민	15	1	24	2	38	2	49	3	59	3	70	2
	20	6	32	17	49	10	62	11	74	8	90	6
이수	12	6	23	4	29	11	38	15	43	17	51	16
	20	6	35	11	44	19	55	22	64	18	76	17
이혜	11	11	20	11	28	14	40	11	47	12	57	10
	18	16	32	17	46	13	63	9	72	10	84	9
장병	8	22	20	11	34	5	46	5	53	6	66	4
	19	11	40	1	56	2	73	2	84	3	101	2
장은	11	11	18	18	25	20	37	18	43	17	50	18
	18	16	31	20	43	21	57	19	64	18	76	17
정수	11	11	23	4	30	9	43	8	49	10	58	9
	18	16	36	6	45	16	61	14	70	13	83	10

성명	25EW	순위	50EW	순위	75EW	순위	100EW	순위	125EW	순위	150EW	순위
	AW	순위	AW	순위	AW	순위	AW	순위	AW	순위	AW	순위
정지	8	22	15	24	21	25	26	24	26	25	26	25
	18	16	29	22	39	23	46	23	46	25	46	25
조미	11	11	22	6	36	4	51	2	65	1	74	1
	17	20	35	11	52	6	69	4	84	3	97	4
조성	13	5	26	1	39	1	54	1	62	2	69	3
	17	20	37	4	53	4	69	4	78	6	86	7
최성	10	17	17	21	23	23	29	23	37	22	41	22
	13	25	24	25	31	25	40	25	49	23	54	23
최필	11	11	20	11	26	18	37	18	45	14	51	16
	19	11	35	11	45	16	60	15	70	13	77	14
홍성	11	11	18	18	27	16	39	13	47	12	55	12
	19	11	29	22	39	23	56	21	65	16	77	14
홍희	25	1	22	6	24	22	36	21	44	15	52	14
	21	3	35	11	51	7	65	8	75	7	85	8

110-715

서울시 중구 필동 3가 26

동국대학교 문과대학

영어영문학과