

# 英语测试中影响完形填空难度因素的实证研究

华东师范大学 奚翠华

**摘要:**完形填空被认为是一种测试综合语言能力、阅读理解能力的快捷经济的方式。本研究就可能影响完形填空难度的几个变量进行实证探讨,其中包括语篇类型、删词类型及答题方法。本研究以98名高三学生为实验对象,完成3篇填空式完形测试和3篇选择式测试。测试完成后,笔者对实验数据进行收集、分析,探究这些变量对完形填空测试难度的影响,并试图在命题难度的把握上找到一种更为合理、科学的测试方法。

**关键词:**完形填空;难度;语篇类型;删词类型

## 1. 引言

我国外语测试中,到处可见完形填空的身影,但是对完形填空的系统研究却不多见,尤其是对完形填空难度的研究更为少见。李筱菊(1997)指出:"如果把完形填空研究的国内和国外情况进行比较我们会发现一个十分奇怪的情况,中国完形填空使用最多研究最少,而国外的完形填空使用不多,但是研究的确实特别多。"总体上看,国内的完形填空研究还是处于以命题方法介绍、解题策略指导为主,实证研究为辅的格局(毛延生,2008)。鉴于此,笔者希望立足于中国英语学习者的实际情况,以中国的英语学习者为研究对象进行实证研究,通过对中国英语学习者在各种常见的完形填空测试中的表现研究,来考察影响完形填空测试难度的因素及其影响,为今后设计各种完形填空测试题提供掌控难度的方法,更好地为日常教学和测试目标服务。

## 2. 研究背景

完形测试(cloze test)是英国学者Wilson Taylor于1953年提出的,最初此类题型是一种测试母语文章可读性的方法,到七十年代,由于奥勒(Oller)等人的大力提倡,这一形式引起广大外语测试专家的注意与兴趣,并作为一种综合考试的形式用于测量个人的外语水平。完形填空被认为是一种测试语言综合能力和阅读理解能力快捷经济的方式(徐强,1992)。

根据Cohen(1980)的定义,完形填空就是题目设计者将一篇由一个或数个段落组成的短文中的一些单词删除,要求受试者利用语言知识、语篇知识及社会知识将删除的词复原的一种测试方法。删去某些单词的短文看上去形体不整,填空使其恢复形体完整,故其称之为完形填空。完形

填空以格式塔心理学中的完形概念为基础,强调语言综合能力的考察(Alderson,2000)。

纵览完形填空60年的发展历程,研究的核心均围绕其测试效度展开(毛延生,2008)。笔者在研读了国内外专家学者的研究后发现,完形填空的很多研究是针对完形填空的效度和信度进行的。据毛延生(2008)对中国知识资源总库(CNKI)的检索,在有关完形填空的文献类研究论文中,55.3%是对完形填空效度、信度理论性研究及其与阅读、词汇教学结合的研究;45.7%是简介性研究引介以及解题方法介绍。

国外很多学者(Oller,1973;Shohamy,1978;Hinofotis,1980;Mullen,1980)的研究表明,完形填空与听力理解、写作及阅读理解之间有很高的相关。Bachman(1985)认为完形填空在特定的测试语境中显示出具有构想效度,如果采用原词评分法的话,其评分客观、容易。

当然,也有学者对完形填空的信度和效度提出怀疑,如Alderson(1980)的研究结果表明,完形测试很难保证评分的一致性,不同的删词起始点、不同的删词比例都会影响测试的结果。Brown(1993)指出,在以往的研究中,效度最差的为19%,最好的为83%;而信度也在3%到96%之间波动(转引自毛延生,2008)。

国内也有些学者对完形填空做了一些具体的实证研究。喻爱菊(1990)对多项选择答案对完形填空有无影响的问题进行实证研究,揭示了多项选择答案对于完形填空的优缺点。陈晓扣(2001)的实证研究比较了填空式和选择式完形试题,指出命题者必须考虑影响试题难度、区分度和效度的两个重要因素:考点层次和推理因素。柳亚杰(2005)通过对照实验,验证了不同起止点

的定距删词完形填空对测试而言都是等价的。

完形填空的研究绝大多数是从效度和信度的角度切入,对难度的研究相对较少。毛延生(2006)对完形测试难度有所涉及,对影响完形填空难度的文本文体(text style)因素作了实证研究,揭示了不同的文本文体在不同的完形填空生成方式下对完形测试难度产生的不同影响。

国外有些学者对难度也有描述,但总体而言,对难度的研究还是偏少。Alderson(1979)和 Braley(1981)通过实证研究表明,语境的难度(context difficulty)和删词比率(deletion rate)对学生的测试结果会产生影响。Mauranen(1988)发现空格完成所需的信息量是影响完形测试难度的重要因素,需要依据上下文的信息来完成的空格的难度显然大于仅需通过理解句意就能完成的空格的难度。Braley(1994)指出语篇本身的难度(text difficulty)会影响完形测试的难度,太难的语篇会影响所有学生的答题表现,从而导致区分度过低。Farhady(1996)也间接提到了关于影响完形测试难度的因素,如受试者、语篇难度、语篇题材、删词方式、删词比率、评分方式及答题方式。

总体来说,以往对完形填空测试难度的研究较少,且对影响难度的因素的研究比较单一,如对语境难度和删词比率或语篇难度等一个或两个影响因素的研究和对不同文体之间、不同比率删词之间的比较研究少有关注。事实上,多个因素可能同时影响完形测试的难度,并且在不同的情况下产生不同的影响。

难度(difficulty 或 item facility)是指试题的难易程度,它是经典测试理论和试题反应理论的重要组成部分。它对试题的质量有很大的影响。如果一个测试对某一群受试者的水平来说太难或太易,其信度通常低,但是同样的测试用于中等水平

或能力水平适合于该难度的群体,则试卷可信度会很高(Henning, 2001)。试题的难度又总是与另一个重要指标区分度有一定关系,太难或太易的试题往往无法体现很好的区分度,从而无法保证考试的效度和信度(Davies, 1999)。难度是一个心理指标,受到多个因素的影响,如受试者的年龄、性别、个性特征和受试环境等;它同时又是一个相对指标,不同的受试者可对同一试题反映出不同的难度。难度受到受试者的影响,包括个性特征(personal characteristics)、背景知识(topical knowledge)、情感图式(affective schemata)及语言能力的影响(Bachman & Palmer, 1996)。

Halliday & Hasan(1976)根据话语分析,将语篇的理解划分为三个层次:1)词的层面(micro-level),即词的选择和词与词之间的关系;2)句的层面,即句子的功能和结构;3)语用的层面(pragmatic level),即读者语篇外的知识。

Bachman(1985)依据还原词的信息来源,把完形测试的空格类型划分为:1)信息源于空格所在分句(clause);2)信息源于空格所在句子(sentence);3)信息源于空格所在上下文(across sentences but within the text);4)信息源于语篇外知识(extra-textual information)。本研究把1)和2)归为一类,即句中理解,将其他两类分别归为句际理解和语篇外理解。

一些研究者(Brown, 1988; Harrison, 1983)依据词的特征(word characteristic),把空格划分为语义(semantic)和功能(functional)空格。

毛延生(2006)的实证研究表明,答题方式对完形测试的难度亦有影响。

综合国内外已有的研究,影响完形填空难度的因素可归纳如下(见图1):

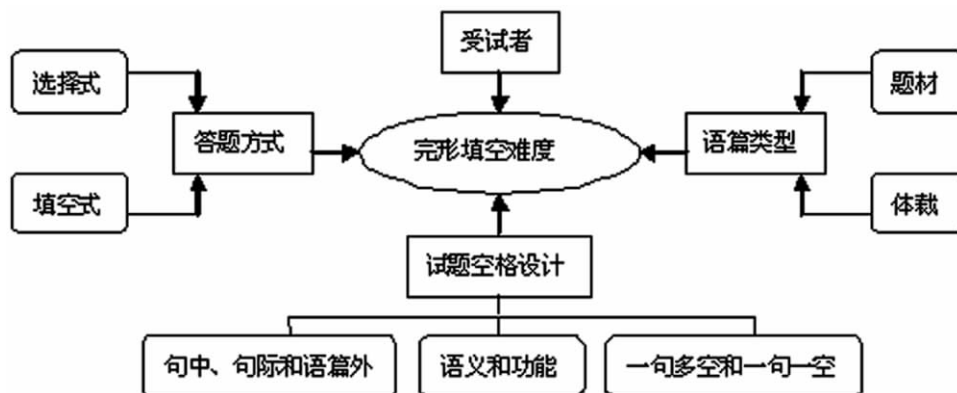


图 1. 影响完形填空难度的因素

从图1可以看出,完形填空的难度受到受试者个性特征、语篇类型、试题空格设计和答题方式的影响,其中语篇类型包括不同的题材和体裁,答题方式包括选择式和填空式两种,试题空格设计包括三种类型。

本研究中的实验对象年龄、背景相近,理解试的语篇无需特殊的背景知识和情感图式,所以本实验不把受试者的个性特征作为研究重点。毛延生(2006)对文体因素结合答题方式和删词方式对难度的影响作过实证研究,指出文体因素对填空式的完形测试难度有较大影响。本研究将从另一个视角研究在不同的语篇类型和答题方式下,空格设计对完形测试难度的影响。

本研究主要讨论语篇类型为故事、含文化背景的说明文和科普说明文;答题方式包括选择式和填空式两种。空格的设计是本研究的一个重点,包括三类:1)句中理解、句际理解和语篇外理解空格;2)功能和语义空格;3)一句多空和一句一空的空格,句中理解空格即借助空格所在句子的意思即可还原单词的空格,句际理解空格即借助于上下文的意思可还原单词的空格,语篇外理解空格即借助于背景知识可还原单词的空格(以下简称句中、句际和语篇外);语义空格即要求填写名词、动词、形容词、副词等实义词的空格,功能空格即要求填写连词、介词、冠词等虚词的空格(以下简称语义和功能);一句多空即一句中含一个以上需还原的单词的空格,一句一空即一个句子中仅含一个空格的空格(以下简称多空和一空)。

笔者的研究旨在对上述可能影响完形填空难度的几个变量进行实证调查,这些变量包括语篇类型、空格类型及答题方式。通过对实验数据的收集、分析来研究这些变量对完形填空测试难度的影响,并试图在命题难度的把握上找到一种更合理科学的测试方法。

本研究要解答的问题为:语篇类型对完形填

空的难度有无影响?有何影响?在语篇类型确定的情况下,空格类型是否对难度产生影响?如有,是何影响?在语篇和空格类型为控制变量的情况下,答题方式是否会对难度有影响?影响如何?以这些问题为中心,笔者选用中学英语阅读材料中常见的语篇类型,针对不同的空格类型及答题方式进行实验。

### 3. 研究方法

#### 3.1 实验对象

笔者以上海市某中学的98名高三学生作为实验对象,其中37名女生,61名男生,平均年龄17.4岁,所有学生均有大约十年学习英语的经历。

#### 3.2 实验材料和实验任务

本研究选用3个语篇,均为高中学生熟悉的语篇类型,分别为1篇故事(讲述比尔盖茨少年时的故事),1篇含文化背景的说明文(男女在社会中的角色)(以下简称文化)和1篇含科普知识的说明文(光的知识)(以下简称科普)。分别采用填空式和选择式答题方式,构成6个测试语篇。

一般来说,完形测试的语篇长度在275至500字左右较为妥当,一般以能留50个空格为宜。根据测验长度与信度的关系,测验题目越多信度越高,文章的首尾一、二句不留空格,空格间隔一般不少于4个单词,平均间隔字数不超过11个单词(舒运祥,2001)。笔者在研究了2003至2005年的上海市高考英语试卷后发现,借助于Microsoft Word中的Flesch-Kincaid Grade Level发现,高考中的完形测试语篇的可读性难度(Readability Index)介于6~12级。数字越大,可读性难度越大。为了保证实验的可信度,笔者所选测试语篇的长度、可读性难度相当(可读性难度6级左右)。

为了制作测试工具,用于测试的语篇在测试之前进行了处理,处理情况如表1:

语篇类型	语篇长度	难度等级	删词率	答题方式 空格类型	填空题(N=40)								选择题(N=40)							
					句中	句际	语篇外	语义	功能	多空	一空	句中	句际	语篇外	语义	功能	多空	一空		
故事	397	5.6	1: 8.4		20	18	2	30	10	23	17	20	18	2	30	10	23	17		
文化	377	5.7	1: 8.5		24	13	3	30	10	30	10	24	13	3	30	10	30	10		
科普	398	6.5	1: 8.6		19	17	4	30	10	23	17	19	17	4	30	10	23	17		

表1. 语篇加工处理结果

表1表明,本研究包括3个语篇——故事、文化和科普,选用的语篇长度为400字左右,其可读性难度等级6级左右;采用变化比率删词的方法,平均删词比率为1:8.5左右,即平均间隔字数为7.5字左右。每篇删词40个,对于删词类型,每一类的空格数基本相当。

本实验的任务包括以填词的方式和选择的方式完成6篇完形测试。分两周进行,第一周的周一、二和三完成填空式测试,时间长度25分钟;第二周同样的时间完成选择式测试,时间长度为20分钟。测试顺序依次为故事、文化和科普。测试前,施测者向学生明确说明测试目的、要求和时间长度。两次测试的语篇相同。

### 3.3 评分方法

完形测试评分的方法有很多种,最常用的有两种:原词评分法(the exact word method)和可接

受词评分法(the acceptable word method)。Miller and Coleman(1967)发现原词评分法与其他方法的相关度高达0.99, Klare, Sinaiko and Stolorow(1972)的研究也表明它们的相关度在0.90以上(转引自舒运祥,2001)。鉴于这两者高度相关,本研究采用原词评分法。

### 4. 实验结果与讨论

实验结束后的各类数据统计分析结果如下:

#### 4.1 语篇类型和难度

经统计,三个填空式测试的平均得分分别为:故事21.847,文化19.745,科普18.520;三个选择式测试的平均分分别为:故事33.622,文化31.622,科普31.653。

为了比较三个不同语篇测试成绩的均值差异,笔者对测试结果的原始数据进行了单向方差分析(One-way ANOVA),结果见表2。

因变量	变异来源	自由度	均方	F 值	P 值
填空式语篇	组间	2	277.401	16.455	.000
	组内	291	16.858		
选择式语篇	组间	2	128.697	13.021	.000
	组内	291	9.884		

表2. 不同语篇类型的难度比较和方差检验

表2表明,不管是填空式还是选择式完形测试,不同类型的语篇之间存在极为显著的差异( $F = 16.455, 13.021, P < .01$ )。

Post-hoc SLD(方差齐性)检验(见表3)进一步表明,对于填空式试题,语篇类型对完形测试的难度有影响( $P < .05$ );而对于选择式试题,故事与文化、科普有显著性差异( $P < .01$ ),而文化与科普之间没有显著性差异,见表3。

I) 组	J) 组	均值差 I - J	标准误差	P 值
填空故事	填空文化	2.1020	.58656	.000
	填空科普	3.3265	.58656	.000
填空文化	填空科普	1.2245	.58656	.038
选择故事	选择文化	2.0000	.44913	.000
	选择科普	1.9694	.44913	.000
选择文化	选择科普	0.0306	.44913	.949

表3. 不同类型语篇之间难度的 LSD 检验

统计分析的结果显示,语篇类型对完形测试的难度有影响。故事显然比文化和科普容易,而文化与科普只在填空式测试中有差异,文化比科普要容易些;在选择式测试中,完形的难度不受影响,即文化与科普没有显著性差异。

#### 4.2 空格类型和难度

对于空格类型与测试难度的研究,笔者从三个不同角度进行研究:第一组是根据帮助我们完成完形填空的线索为依据划分的,即句中、句际和语篇外;第二组是以空格中所用词的类别来划分,即语义和功能;第三组是以空格在句子中的数量为划分依据,即一句多空和一句一空。这三个小组分别称为“线索空格”、“词类空格”和“数量空格”。

就线索空格情况而言,数据分析得到的三类空格的平均分分别为:填空句中0.5156,填空句际0.4951和填空语篇外0.2920;选择句中0.8189,选择句际0.7964和选择语篇外0.7540。

为了比较三个不同的空格类型所得测试成绩的均值差异,笔者对测试结果的原始数据进行了单向方差分析(One-way ANOVA)。结果显示(如表4),不管是填空式还是选择式完形测试,不同

因变量	变异来源	自由度	均方	F 值	P 值
填空式语篇	组间	2	.425	12.629	.000
	组内	1173	.34		
选择式语篇	组间	2	6.78	165.511	.000
	组内	879	.041		

表4. 句中、句际和语篇外空格的难度比较和方差检验

类型空格之间的难度存在极为显著的差异(  $F = 12.629, 165.511, P < .01$  )。这说明不同类型的空格对完形测试的难度有影响。

Post-hoc Tamhane's T2 ( 方差不齐性,  $P = .000$  ) 检验( 见表 5) 进一步表明, 无论是填空式还是选择式测试, 句中理解与句际理解空格之间无显著性差异(  $P = .195, .236 > .05$  ), 语篇外理解空格与其他两种空格类型有显著性差异(  $P = .000 < .01$  )。语篇外理解空格在三者中最难。

I) 组	J) 组	均值差 I - J	标准误差	P 值
填空句中	填空句际	-.0216	.01192	.195
	填空语篇外	.2515	.01878	.000
填空句际	填空语篇外	.2732	.01847	.000
选择句中	选择句际	.0149	.00865	.236
	选择语篇外	.1308	.01785	.000
选择句际	选择语篇外	.1159	.01814	.000

表 5. 句中、句际和语篇外空格难度之间的 Post-hoc Tamhane's T2 检验

就词类空格和数量空格而言, 测试得到的平均分如下: 填空语义 0.49 和填空功能 0.4977, 填空多空 0.4935 和一空 0.5123; 选择语义 0.8248 和选择功能 0.7579, 选择多空 0.8126 和选择一空 0.8045。笔者对原始数据进行 t 检验( t 检验是测试成绩的两组均数的比较) 结果( 见表 6) 显示, 就词类空格而言, 对于填空式完形试题, 语义和功能空格的难度无显著性差异(  $P = .663 > .05$  ), 而对于选择式试题, 两者有极其显著性差异(  $P = .000 < .01$  )。就数量空格而言, 填空式测试的多空和一空之间有显著性差异(  $P = .038 < .05$  ), 而对于选择式试题, 两者无显著性差异(  $P = .227 > .05$  )。

T 检验	填空式		选择式	
	语义 - 功能	多空 - 一空	语义 - 功能	多空 - 一空
T 值	-.435	-2.084	9.236	1.210
自由度	391	391	391	391
P 值	.663	.038	.000	.227

表 6. 语义、功能和多空、一空空格难度之间的 t 检验

### 4.3 答题方式与难度

从以上讨论中可以看出, 无论是对于不同语篇类型, 还是不同空格类型, 填空式的完形测试都比选择式的要难。对两种不同的答题方式所得到的测试成绩进行 t 检验分析, 结果显示  $P = .000 < .01$ , 两者的难度呈显著性差异。

## 5. 结果讨论

本研究的目的是试图探究影响完形填空难度的因素, 笔者以语篇类型、空格类型、答题方式作为变量, 研究各因素与难度的关系。

语篇类型对完形测试的难度有影响。在语篇的可读性难度相似或相同的情况下, 就不同类型语篇而言, 故事比文化和科普容易, 数据分析表明, 故事与文化和科普的难度差异达到了显著水平。笔者认为差异的原因可能是因为读者对文本的认知方式差异造成的。在相同阅读难度的情况下, 故事容易读是因为故事有情节, 而情节的线索是以生活经验为基础的, 读者在解读文本时可以大量动用生活常识来建构文本的篇章关系, 从而建构起故事; 而在文化和科普文本中, 大量的篇章关系是依据学术逻辑来建立的, 读者必须依据文本中的线索来建立篇章联系, 一旦某个线索中断, 可能整个篇章的关系就会中断。而在选择式完形填空中, 这种差异消失了。笔者认为原因可能是选择式完形填空是一种有线索的测试, 所以对学生来说, 相对认知负担轻, 因而, 故事、文化、科普之间的语言难度就不起作用了。而对于文化与科普, 只有在填空式测试中, 两者有难度差异, 即科普比文化难; 而在选择式测试中, 两者虽然在测试所得的平均分上有差异, 但其差异未达到显著水平, 即两者难度相当。因此, 我们可以在教学或命题中, 依据不同的考试目的, 采用不同类型的语篇, 以期达到较好的区分效果。

空格类型对难度有影响。根据 Bachman (1985) 对空格类型的划分, 本研究把空格类型分为句中理解、句际理解和语篇外理解三类, 语篇外理解最难, 且与其它两者有显著差异, 句中理解比句际理解容易, 但就难度而言, 未达到显著差异的水平。"语篇"( 即上下文) 是理解的一个重要因素, 上下文可以为语言理解提供线索, 一旦篇章理解程度提高了, 空格填空就变得容易。

依据 Brown (1988) 的划分, 空格可分为语义和功能两种类型。本研究发现, 对于填空式完形测试, 语义空格较之功能空格更难, 但其难度差异未达到显著水平; 对于选择测试, 功能空格更难。功能词语比实词难以学习和掌握, 功能词的意义更抽象, 理解深度要求更高, 对二语学习者来说是一个挑战。

对于空格所在位置, 即位于同一句还是不同句中, 对填空式完形测试会产生影响, 一句一空的空格比一句多空的空格容易, 而对选择式测试无

显著影响。这也说明空格密度高,句子的意义破碎,读者难以完成句子的阅读,测试题给学生的线索就小,线索越少,测试难度就越高。

就难度适中而言,选择完形更受青睐,尤其对于中等或以下水平的学生,由故事编制的选择完形更适合他们。而要区分高水平的学生,笔者建议用填空完形测试。就空格类型而言,根据本研究结果,笔者建议,对于中等偏下的学习者,可采用句中理解和语义空格组合为主的选择完形测试;对于水平较高的学习者,可采用句中或句际、语义空格和一句多空的组合为主的填空完形测试。

## 6. 结语

难度的控制在测试中其实是一个很重要的方面,可以影响学生的情感,适合的难度可以保持学生的兴趣,促使他们学得更好。对于任何测试而言,难度的把握是保证信度和效度的重要手段之一。如果我们遵循科学命题的原则,就必须非常清楚测试目标及影响测试结果的因素。本研究通过实验,得到了一些结果,了解了哪些因素对完形填空测试的难度产生影响并是如何影响的。但是,本研究也并非完美,因为影响测试难度的因素还可能的情感因素,本研究中并没有涉及。另外,实验对象来自一个学校,可能造成样本不够全面,笔者希望在以后的研究中对此加以改进。

命题是一门艺术,命好题更是一项艰巨的工作。对完形填空测试来说,经验对于语篇的选择和空格的选择是非常重要的,然而,经验加研究才是上策。笔者希望本研究能使外语教师从中受益,也希望更多的外语教育工作者参与外语测试方面的具体研究。

## 参考书目

- Alderson, J. C. 1979. The cloze procedure and proficiency in English as a foreign language (J). *TESOL Quarterly*, 13, 219 - 27.
- Alderson, J. C. 1980. Native and nonnative speaker performance on cloze tests (J). *Language Learning* 30, 59 - 76.
- Alderson, J. C. 1983. The cloze procedure and proficiency in English as a Foreign Language (C). In Oller, J. W. Jr. (ed.) *Issues in Language Testing Research*. Rowley, MA: Newbury House.
- Alderson, J. C. & A. H. Urquhart. 1985. The effect of students' academic discipline on their performance on ESP reading test (J). *Language Testing* 2, 192 - 204.
- Alderson, J. C. 2000, *Assessing Reading* (M). Cambridge: Cambridge University Press.
- Bachman, L. F. 1985. Performance on cloze test with fixed-ratio and rational deletions (J). *TESOL Quarterly*, 19, 535 - 556.
- Bachman, L. F. 1990. *Fundamental Considerations in Language Testing* (M). Oxford: Oxford University Press.
- Bachman, L. F. and Palmer A. S. 1996. *Language Testing in Practice* (M). Oxford: Oxford University Press.
- Brown H. D. 1987. *Principles of Language Learning and Teaching* (M), second edition. Englewood Cliffs, NJ: Prentice-Hall.
- Brown J. D. 1983. A close look at cloze: validity and reliability (C). In Oller, J. W., (ed.). *Issues in Language Testing Research*. Rowley, Mass. Newbury House: 237 - 50.
- Brown J. D. 1988. What makes a cloze item difficult? (C) *University of Hawaii Working Papers in ESL*, 72, 17 - 39.
- Brown J. D. 1989. Cloze item difficulty (J). *JALT Journal* 11, 46 - 67.
- Brown, J. D. 1993. What are the characteristics of natural cloze tests (J)? *Language Testing* 10: 93 - 116.
- Carroll, J. B. 1961. Fundamental considerations in testing for English language proficiency of foreign students (A). *Testing the English Proficiency of Foreign Students*. Washington, DC: Center for Applied Linguistics.
- Cohen, A. D. 1980. *Testing language ability in the classroom* (M). Rowley, Massachusetts: Newbury House
- Davis, A., A. Brown, C. Elder, K. Hill, T. Lumley & T. McNamra 1999. *Dictionary of Language Testing* (M). Cambridge: Cambridge University Press.
- Farhady, H. 1996. A text-driven method for the deletion procedure in cloze passages (J). *Language Testing* 13: 191 - 207.
- Halliday & Ruqaiya Hasan. 1976. *Cohesion in English* (M). London: Longman.
- Harrison, A. 1983. *A Language Testing Handbook* (M). London: Macmillom Press.
- Henning, G. 2001. *A Guide to Language Testing: Development, Evaluation and Research* (M). Heinle & Heinle Publishers, a division of Wadsworth, Inc.
- Hinofotis, F. B. 1980. Cloze as an alternative method of ESL placement and proficiency testing (J). *Research in Language Testing*. Ed. John W. Oller, Jr. & Kyle Perkins. Rowley, MA: Newbury House: 121 - 28
- Hughes, Arthur. 1989. *Testing for Language Teachers* (M). Cambridge: Cambridge University Press.

- Jonz, J. 1987. Textual cohesion and second language comprehension (J). *Language Learning* 37, 409-38.
- Klein Braley, C. 1981. *Empirical Investigation of Cloze Tests* (M). *Dixx. Univ. of Duisbury*.
- Klein Braley, C. 1994. *Language testing with the C-Test*. A linguistic and statistical investigation into the strategies used by C-Test takers, and the prediction of C-Test difficulty (M). Higher doctoral thesis Habilitationsschrift) University of Disburg.
- Mauranen, A. 1988. The semantic cloze: an approach to testing EFL reading (R). Paper presented at TESOL convention, Chicago, March.
- Mullen, K. 1980. Rate reliability and oral proficiency evaluation (J). *Research in language testing*. Ed. John W. Oller, Jr. & Kyle Perkins. Rowley, MA: Newbury House.
- Oller, J. W. 1972. Scoring methods and difficulty levels for cloze tests of proficiency in English as a second language (J). *Modern Language Journal* 56, 151-58.
- Oller, J. W. 1973. Pragmatic language testing. *Language Sciences* 28, 7-12.
- Shohamy, E. 1978. Investigation of concurrent validity of oral interview with cloze procedure for measuring proficiency in Hebrew as a second language (D). Unpublished. University of Minnesota.
- Taylor, W. L. 1953. Cloze Procedure', a new tool for measuring readability (J). *Journalism Quarterly* 30: 415-33.
- 陈晓扣. 论英语测试完形填空试题 (J). 解放军外国语学院学报(第3期), 2001。
- 韩宝成. 外语教学研究中的统计方法 (M). 北京: 外语教学与研究出版社 2000。
- 李筱菊. 语言测试的科学与艺术 (M). 湖南: 湖南教育出版社, 1997。
- 刘润清. 语言测试和它的方法(修订版) (M). 北京: 外语教学与研究出版社, 1991。
- 柳亚杰. 定距离完形填空再研究 (J). 浙江树人大学学报(第6期), 2005。
- 毛延生. 完形填空测试难度的文体研究(硕士论文) (D). 2006。
- 毛延生. 国内完形填空研究20年述评 (J). 外语教学理论与实践(第2期) 2008。
- 舒运祥. 外语测试的理论与方法 (M). 上海: 世界图书出版社 2001。
- 徐强. 英语测试的理论与命题实践 (M). 安徽: 安徽教育出版社, 1992。
- 喻爱菊. 多项选择答案对 Cloze Test 的影响 (J). 外语教学与研究(第1期), 1990。
- 邹申. 英语语言测试——理论与操作 (M). 上海: 上海外语教育出版社, 1998。
- (通讯地址: 200241 上海市东川路 500 号 华东师范大学外语学院)
- (文字编校: 王亚民)

## A study on the factors that influence cloze test difficulty

by XI Cuihua

**Abstract:** The cloze test is considered to be a quick, economic method of measuring overall language proficiency as well as reading comprehension. This paper reports on the experiments carried out on the cloze procedure where the variables of text type, text topic, deletion type and response format are involved. Carefully designed cloze tests were administered to 98 middle school students. The scores for the cloze tests were computerized and the SPSS 11.5 was applied to process the data.

**Key words:** cloze test; text type; deletion type