

基于工作记忆理论的汉语阅读广度测量方法讨论

吴熠丹 韩 晓

东南大学，南京

摘 要 | 工作记忆广度的大小因人而异，对工作记忆广度的测量在认知心理学、语言学等领域的研究中都有重要意义。以 Daneman & Carpenter 的测量方法为代表的英语阅读广度测量方法的确立为以后的研究提供了思路和借鉴，现有的汉语阅读广度测量方法多少都有着它们的影子。但也因此，现有的汉语阅读广度测量方法存在着生硬套用英语测量方法、缺乏对汉语自身特点的考虑等问题。本文通过对这些测量方法的分析与讨论，对今后的汉语阅读广度测量方法的研究与探索提供了一些新的思路。

关键词 | 工作记忆；工作记忆广度；英语阅读广度测量；汉语阅读广度测量

Copyright © 2021 by author (s) and SciScan Publishing Limited

This article is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/). <https://creativecommons.org/licenses/by-nc/4.0/>



1 引言

工作记忆（working memory）是个体认知过程中对信息进行暂时性储存和加工的一种具有有限容量的系统^[1, 2]，在学习、阅读理解、问题解决和

基金项目：本论文获得“江苏省大学生创新创业训练计划项目（S202010286198）”资助。

作者简介：吴熠丹，东南大学外国语学院日语系本科生。

韩晓，教育学博士，讲师，主要研究方向为日语教育学、认知心理学。

文章引用：吴熠丹，韩晓. 基于工作记忆理论的汉语阅读广度测量方法讨论[J]. 语言学, 2021, 3 (1): 324-334.

<https://doi.org/10.35534/lin.0301016>

智力活动等高级认知活动中发挥着重要作用^[3]，占据着人类认知活动的核心地位^[4]。

“工作记忆 (working memory)^[5]”这一概念最早是由 Miller 等人 (1960) 提出。Atkison & Shiffrin (1968) 使用工作记忆概念来解释心理学领域的研究，将其与短时记忆概念相对应^[6]。Baddeley & Hitch (1974) 从短时记忆的研究中提出了工作记忆模型，指出工作记忆区别于组成成分单一的短时记忆，是由中央执行系统、语音回路、视空间模板三个独立成分所构成的复杂系统，由此将“工作记忆”与“短时记忆”区分开来，同时正式提出了工作记忆的概念^[7]。Baddeley (1986) 将工作记忆明确定义为“个体在执行认知任务时对信息暂时储存与加工的能力”，为此后有关工作记忆的理论发展奠定了基础，也突出了工作记忆在认知活动中发挥的重要作用。

在工作记忆的定义中，有两点值得注意，其一是工作记忆的双重任务（即其既负责信息储存任务，又负责信息加工任务），其二是工作记忆的有限容量。为了解释储存、加工这项双重任务，许多学者展开了相关讨论与研究。其中较为著名的是 Daneman & Carpenter (1980) 提出的工作记忆的资源共享模型。他们认为，工作记忆的储存和信息处理共享着相同的有限资源，即一方任务容量的增加会导致另一方任务容量的减少，储存任务容量与加工任务容量互相制衡^[8]。

工作记忆是一个加工资源——即容量（又称广度）有限的记忆系统。对工作记忆这一容量有限的系统进行的测量被称为工作记忆广度任务 (working memory span)，其主要有 3 种方法，分别为听力广度测量 (listening span test)、阅读广度测量 (reading span test) 和运算广度测量 (operation span test)。而在本文所讨论的阅读广度测量方面，Daneman & Carpenter (1980) 基于资源共享模型和工作记忆的特点率先提出了最初的阅读广度测量方法^[8]。这一测量方法长期以来在心理语言学及二语习得领域的研究中广受借鉴，对后来的阅读广度测量发展有着奠定意义。许多学者在其基础上进行改进和创新，设计出了不同的阅读广度测量方法，促进了工作记忆研究的发展与进步。

而与英语阅读广度测量相比，现阶段，国内涉及汉语阅读广度的研究相

对有限,已有的汉语阅读广度测量(C-RST)的研究与开发缺乏具有说服力的成果。因此,本文将对国内外已有的较为典型的英、汉语阅读广度测量方法进行梳理和总结,分析其中的优点与不足,为汉语阅读广度测量方法的研究提供借鉴。

2 英语阅读广度的测量

以 Daneman & Carpenter (1980) 的阅读广度测量为代表,在英语阅读广度测量方面有以下三种比较经典的阅读广度测量方法。这几种方法有所不同但都被广为接受,因此对于汉语阅读广度测量研究而言也颇具借鉴意义。

2.1 Daneman & Carpenter (1980) 的阅读广度测量方法

2.1.1 测量方法

测量共由 60 个句子组成,其中阅读广度由 2 递升至 6,每个广度的测量由 3 组句子组成,每组句子数量与其阅读广度水平相对应(如:阅读广度水平 2 下有 3 组句子,每组又分别有 2 个句子,以此类推)。

测量过程中,被试被要求大声朗读卡片上的句子并记住句尾单词,朗读完毕即换至下一张卡片。在每组句子展示完毕后,被试需要按照顺序回忆出句末单词。阅读广度水平与句子数量随着测量的进行递增。当被试无法回忆出句末单词或回忆错误时,测量终止。测量成绩根据该广度下成功的组数而定,以阅读广度水平 3 为例,若是 3 组中成功了 2 组则记该被试的阅读广度为 3^[8]。

2.1.2 讨论

Daneman & Carpenter (1980) 的阅读广度测量作为原始版本,为后世提供了研究的范本,颇具借鉴价值,至今仍被许多学者所接受采用。然而,随着工作记忆容量与阅读广度相关研究理论的进行和发展,其中存在的问题也日益凸显出来。

首先,测量本身具有不适用性。作为测量的对象,每一位被试都是不可忽视的变量。而尽管测量对被试做出了明确的要求,但被试个人的被控和自觉的

过程必然是不相同的，这也就导致了同样的测量方法不一定适用于所有被试这一不适用性。

其次，测量过程存在不可控性。其中，最大的不可控体现在被试是否能够按照要求同时对信息进行储存和加工。比如，要求被试大声朗读句子并不能保证被试一定对句子进行了理解，同时被试有可能使用一些技巧方法来帮助记忆句末单词。

最后，测量结果存在不普遍性和不充分性。不普遍性即在固定的测量环境之下得出的测量结果难以用来解释自然情况下的普遍能力，不充分性则在于测量结果只能反映信息储存而未体现被试的信息加工能力。

2.2 Lesser (2010) 及 Water & Caplan (1987) 的阅读广度测量方法

2.2.1 测量方法

Lesser (2010) 的阅读广度测量相较于原始版本要更加先进、更加准确。测量中，句子的呈现方式由纸面呈现换成了由计算机控制、屏幕呈现，操作过程更加简便化、程序化。其对于被试的要求也有所不同，在 Lesser 的阅读广度测量中，被试不仅需要朗读句子，还被要求对句子的正误进行判断。理想情况下，这种方法在一定程度上有利于促使被试对句意进行理解。而最终测量结果的呈现也有所不同，其分数的计算包括了判断句子正误的反应时间、正确判断的句子的数量、准确记忆的句尾单词数量三方面^[9]，相较于原始版本则更加全面可信。

同为 Daneman & Carpenter (1980) 阅读广度测量的改进版本，Water & Caplan (1987) 的阅读广度测量又与 Lesser (2010) 的测量存在着不同之处。Water & Caplan (1987) 的阅读广度测量过程也是由计算机控制的，但在对被试的要求和分数的计算方面有所不同。在他们的测量中，被试无需朗读句子，但要判断句子在语义、句法上的正误。而在分数计算方面，则是将判断正误的平均反应时间、判断正误的平均精度、正确顺序下准确记忆的单词数量这三项作为衡量数值^[10]。

2.2.2 讨论

以上两种对于改进阅读广度测量的探索与尝试,虽然采取的具体方法是不完全一致的,但其目的都是被试能够更加自觉地按照指令进行对信息的储存以及加工工作,并且在结果的呈现上同时衡量被试对信息的储存能力与加工能力,以更好地反映被试的阅读广度和工作记忆容量。

在原始版本中,被试被要求朗读句子,这实际上默认了被试朗读句子的过程等同于理解句子的过程。而以上两种改进版本,前者多了判断句子正误的要求,后者则将朗读条件去除、只需判断正误,实际上都提高了对被试的要求,使其在明确要求下不得不理解句子。然而,正误的判断是一种单纯的判断,答案非正即误,被试也有可能利用这一点胡乱作答或在某些句子上混淆过去。若是基于此情况,即便是分数的计算更加科学全面,最终成果也会同样欠佳。如何在这方面进一步完善测量方法,便是需要进一步思考的问题。

3 汉语阅读广度的测量

汉语阅读广度测量(C-RST)是一种以汉语母语者为对象的用来测量被试工作记忆容量水平的工具。与英语阅读广度测量一样,由工作记忆的特点所决定,该测量同样采用双任务研究方法,包括储存、加工两项任务。在国外先行研究的基础上,汉语阅读广度测量有所开发与实践,但仍然相对有限,且取得的研究成果在信度和效度方面的效果也有待进一步的验证。笔者将以下面3种较为典型的汉语阅读广度测量为例,对其各自的特点进行介绍并探讨其利弊所在。

3.1 杨奇伟(2007)的汉语阅读广度测量方法

3.1.1 测量方法

在该实验的汉语阅读广度测量中,句子由主试通过操作键盘在计算机屏幕上——呈现。对被试的要求则与原始版本一致,需要被试将句子朗读出声,同时按照正确顺序记忆每组句子的结尾词(句末双音节汉字词语),句子朗读完毕即切换至下一页。当一组句子朗读完毕,则要求被试按正确顺序回忆出该组

结尾词。当被试回答错误或回忆不起结尾词时,测量终止。分数的计算采用 Daneman & Carpenter (1980) 的计分方法。但与原始版本有所不同,该测量调整了结尾词的呈现方式。在 Daneman & Carpenter (1980) 的测量中句子是完整地呈现给被试的,而该测量为防止被试提前看结尾词,首先呈现的是不带结尾词的句子,一秒后再补上结尾词^[11]。

3.1.2 讨论

该汉语阅读广度测量在原始版本的基础上延迟了结尾词的出现,防止了被试提前关注结尾词。这无疑有利于被试更好地进行句子的朗读步骤,而不是将注意力过分集中在结尾词的记忆任务上。

然而,与 Daneman & Carpenter (1980) 测量中存在的缺憾相类似,对于是汉语母语者的被试来说,汉语是他们有着极高熟练度的第一语言,因此在进行句子朗读的时候被试并不一定对句意进行了理解。这也就意味着,即使这种方法有助于被试不提前看结尾词,信息储存任务与加工任务的并行(即被试在测量中做到既记忆结尾词又理解句意)仍然是难以保证的。

而除上述不同点带来的优缺点之外,该测量几乎与原始版本一致,也就不可避免地存在着与原始版本一样的问题。即测量本身的不适用性、测量过程中的不可控性以及测量结果中的不普遍性和不充分性仍然存在,其所带来的缺憾也仍然没有得到改善。

3.2 鲁忠义、范宁(2006)的汉语阅读广度测量方法

3.2.1 测量方法

该汉语阅读广度测量由计算机控制测量进程。测量中,每个句子完整呈现,呈现时间为6秒,被试需要在这期间阅读句子并记住结尾词。6秒后切至下一页,屏幕上会出现一个与刚才阅读过的句子有关的短句,被试需要点击鼠标选择“正确”或“错误”来判断该短句是否与测量句意思一致。判断完成则页面自动跳转至下一页。当每组句子结束后,被试需按顺序回忆结尾词,回忆不出或错误则测量终止。最终结果的呈现采用 Daneman & Carpenter (1980) 的计分方法^[12]。

3.2.2 讨论

该测量与上一个测量相比,采用了更加灵活的方法。首先,在进度的控制上,采用了机器操作。因为没有要求被试阅读出声,由机器控制的句子呈现的时限安排在一定程度上防止了被试反复看结尾词。而在信息处理即理解句意的要求上,该测量通过每句结束后附带一个用于判断句意的短句的办法,更加约束了被试在阅读时按照要求去理解句意。

该汉语阅读广度测量更加完善,但不能否认的是这些改进点并不能够保证被试在理想状态下进行测量。一方面,前文也有指出,正误的判断只有两种选择,因此在这样的单纯的要求下,被试仍可能采取投机方法、忽视句意理解。另一方面,为完善测量而设计的附加要求(即短句判断正误),又会反过来给测量带来影响,比如判断句子正误难度过高反而会扰乱被试对结尾词的正常记忆。

3.3 Kim 等(2016)的汉语阅读广度测量方法

3.3.1 测量方法

该测量由计算机控制。过程中要求被试朗读句子并判断句子语义的通达度,同时记住句尾的英文字母。每组句子结束,被试需要按照正确顺序依次说出相应字母。最终结果采用 all-or-nothing 单位计分法,即按照正确顺序正确回忆出字母则计1分,反之则计0分,总分为所有分数累计之和^[13]。

3.3.2 讨论

此项汉语阅读广度测量要求被试在朗读出声的同时判断句子语义通达度,这无疑有利于被试更加自觉地进行句意的理解工作,对于确保被试进行信息加工任务来说是一份双重保险。然而,该测量在信息储存上的要求以及分数计算方法的选择上则稍显不足。

首先,该测量要求被试按顺序记忆的是句末英语字母,而非常规汉语阅读广度测量中的双音节汉字词语。与后者相比,前者显然较为简单好记,可能会降低被试记忆任务的难度。因此,有必要讨论其是否适用于针对汉语母语者的阅读广度测量中。其次,作为测量结果的呈现,该测量采用的分数累加的方法

也较为简单, 仅仅体现了正确的组数。同时由其计分方式可见, 该测量实际上并未设置梯度, 这便导致分数的计算与呈现也难以体现梯度, 测量结果的效度值得商榷。

4 关于汉语阅读广度测量的进一步思考与讨论

目前, 国内围绕或涉及汉语阅读广度展开的研究相对有限, 汉语阅读广度测量方法还有较大的提升空间。基于前文的介绍与分析可知, 国内各类汉语阅读广度测量虽不完全相同, 但也存在着类似的问题。比如, 汉语阅读广度测量存在着对英语阅读广度测量进行简单模仿、测量中缺乏对信度和效度的验证、测量结果呈现单一化等问题。如何解决这些问题, 提高汉语阅读广度测量的信效度, 是今后研究的重点。基于上文, 笔者认为还需对以下几点进行重点关照:

(1) 汉语阅读广度测量的特殊性。汉语阅读广度测量的特殊性主要体现在被试对汉语的熟练程度上。作为汉语母语者, 被试对汉语有着较高的熟练度。因此, 即便被要求朗读句子, 被试也极有可能仅仅读出声音、却不分辨句意。同时, 汉语本身存在特殊性, 比如在声调韵律、呈现形式等方面区别于其他语言。而这或许会在一定程度上干扰被试对信息的储存与加工工作, 比如特殊的结尾词可能会帮助被试更好地记忆等。因此, 如何设计出符合汉语和汉语母语者特殊性的测量方法是一大难点。

(2) 被试的投机与技巧使用。许多先行研究考虑到了朗读要求的局限性, 因而设计了促使被试按要求理解句意的附加任务, 且多为正误判断。然而, 在仅有两个选项的条件下, 被试可能为了更集中于记忆任务而存在投机心理, 进行随意选择。另一方面, 在结尾词的记忆上, 被试可能会采用一些特殊的技巧方法来更好地记忆结尾词。比如, 慢速或大声读结尾词和联想记忆法等等。因此, 如何设计方法才能有效规避被试的投机和使用技巧也是需要继续探索的问题。

(3) 测量结果的合理性、全面性。即汉语阅读广度测量采用什么样的计分方法, 才能更加合理更加全面地呈现出测量结果。作为最初的测量版本, Daneman & Carpenter (1980) 的阅读广度测量计分方法体现了梯度, 但却只反映

了被试的信息储存能力。而前文的介绍中也有汉语阅读广度测量采用的是分数累加的方法,这一方法则不仅没体现被试信息加工能力,也没展现出信息储存能力的梯度。汉语阅读广度测量结果呈现的全面性,需要考虑到信息储存与信息加工能力的同时展现,同时还要体现其梯度和准确性。如何设计计分方法才能达成理想效果,需要进一步的讨论。

5 总结

阅读广度测量存在着测量本身的不适用性、测量过程中的不可控性以及测量结果中的不普遍性和不充分性等问题。针对这些问题,研究者们基于 Daneman & Carpenter (1980) 的阅读广度测量,不断改进与探索理想的方法,使得阅读广度测量方法变得更加完善。而在汉语阅读广度测量方面,基于阅读广度测量原始版本而加以适当调整或较大改进的研究也在发展之中。但其中存在着的简单模仿英语测量、尚缺乏信度和效度的验证、结果呈现单一化等问题,又要求对汉语阅读广度测量的特殊性、被试的投机与技巧使用与测量结果的全面性进行进一步的思考与讨论。

虽然目前汉语阅读广度测量的最佳方法尚未可知,但在借鉴国内外相关先行研究的同时,还需要充分考虑汉语自身的特点,把握好汉语阅读广度测量的特殊性、注意点。通过不断进行更加合理的改进与设计,相信汉语阅读广度测量会向着更完善、更理想的方向发展。

参考文献

- [1] Baddeley A D. Working memory [J]. Science, 1992, 255 (5044): 556-559.
- [2] Baddeley A D. Is working memory still working? [J]. European Psychologist, 2002, 7 (2): 85-97.
- [3] Barrouillet P, Mignon M, Thevenot C. Strategies in subtraction problem solving in children [J]. Journal of Experimental Child Psychology, 2008, 99 (4): 233-251.

- [4] Caggiano D M, Jiang Y, Parasuraman R. Aging and Repetition Priming for Targets and Distracters in a Working Memory Task [J] . Neuropsychology Development & Cognition, 2006, 13 (3-4) : 552-573.
- [5] Miller G A, Galanter E, Pribram K H. Plans and the structure of behavior [M] . New York: Henry Holt and company, 1960: 226.
- [6] Atkinson R C, Shiffrin R M. Human memory: a proposed system and its control processes— science direct [J] . Human Memory, 1977: 7-113.
- [7] Baddeley A D, Hitch G J. The Psychology of Learning and Motivation: Advances in Research and Theory [M] . New York: Academic Press, 1974: 47-89.
- [8] Daneman M, Carpenter P A. Individual differences in working memory and reading [J] . Journal of Verbal Learning & Verbal Behavior, 1980, 19 (4) : 450-466.
- [9] Leiser M J. Learner-Based Factors in L2 Reading Comprehension and Processing Grammatical Form: Topic Familiarity and Working Memory [J] . Language Learning, 2010, 57 (2) : 229-270.
- [10] Waters G, Caplan D, Hildebrandt N. Working memory and written sentence comprehension [J] . The Psychology of Reading, 1987, 9: 531-555.
- [11] 杨奇伟. 英, 汉语阅读广度的影响因素: 工作记忆广度理论的潜变量分析 [D] . 西北大学, 2007.
- [12] 鲁忠义, 范宁. 工作记忆广度与汉语句子语境效应的关系 [J] . 心理学报, 2006, 38 (1) : 22-29.
- [13] Kim S A, Packard J, Christianson K, et al. Orthographic consistency and individual learner differences in second language literacy acquisition [J] . Reading & Writing, 2016, 29 (7) : 1409-1434.

Discussion on the Measurement Method of Chinese Reading Span Based on Working Memory Theory

Wu Yidan Han Xiao

Southeast University, Nanjing

Abstract: The size of working memory span varies from person to person, and the measurement of working memory span is of great significance in the research of cognitive psychology, linguistics and other fields. The establishment of English reading span test, which is represented by Daneman & Carpenter's method, provides ideas and references, and the existing Chinese reading span tests are influenced by them in some degree. But also because of this, the existing methods of Chinese reading span tests have problems, such as rigid application of English reading span tests and lack of consideration of the characteristics of Chinese itself. Through the analysis and discussion of these methods, this article provides some new ideas for the future research and exploration of the methods of Chinese reading span test.

Key words: Working memory; Working memory span; English reading span test; Chinese reading span test