

《心理学报》审稿意见与作者回应

题目：藏语阅读知觉广度的眼动研究

作者：白学军，高晓雷，高蕾，王永胜

第一轮

审稿人 1 意见：

本研究使用眼动的技术，探讨了藏语文本阅读过程中知觉范围的大小。阅读过程中的知觉范围这一问题在不同语言中已经有了大量的研究；与以往研究相比，本研究只是换了一种新的语言，创新性有欠缺。同时，本研究在诸多方面有所欠缺。

意见 1：试验设计方面，作者认为知觉广度的研究常选用 X 作为掩蔽刺激，因此本实验亦以 X 为掩蔽材料。这样的做法显然不合理。这是因为，一些知觉广度的研究之所以选用 X 作为掩蔽，是因为这些实验皆以英文或其他拼音文字阅读为研究对象。而在中文研究中，则采用了汉字为掩蔽。同理，由于本实验的研究对象为藏语文字，理应使用藏文作为掩蔽刺激。

回应：非常感谢审稿专家所提出的宝贵意见。

诚如审稿专家所言，在拼音文字的研究中，掩蔽刺激的选择较为统一，通常选用“X”符号作为掩蔽材料，因为“X”符号简洁而不具有任何意义，对阅读干扰较少。如 Rayner(1981, 1986)在对英文阅读知觉广度的研究中，采用了“X”符号作为掩蔽材料(Rayner, K., Inhoff, A. W., Morrison, R. E., Slowiaczek, M. L., & Bertera, J. H. (1981). Masking of foveal and parafoveal vision during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance*, 7(1), 167; Rayner, K. (1986). Eye movements and the perceptual span in beginning and skilled readers. *Journal of Experimental Child Psychology*, 41(2), 211-236.)。

对中文阅读知觉广度的研究中，不同的研究者使用了不同的掩蔽刺激。常用的掩蔽材料一般为三种类型，即繁体字、相似汉字和星号“※”。Inhoff(1998)的研究中曾使用繁体字和相似的汉字作为掩蔽材料(Inhoff, A. W., & Liu, W. (1998). The perceptual span and oculomotor activity during the reading of Chinese sentences. *Journal of Experimental Psychology: Human Perception and Performance*, 24(1), 20-34.)；熊建平等(2007)的研究则使用繁体字作为掩蔽材料(熊建萍, 闫国利, 白学军. (2007). 高中二年级学生中文阅读知觉广度的眼动研究. *心理与行为研究*, 5(1), 60-64.)；而闫国利等(2008)年的研究则使用“※”作为掩蔽材料(闫国利, 熊建萍, 白学军. (2008). 小学五年级学生汉语阅读知觉广度的眼动研究. *心理发展与教育*, 24(1), 72-77.)。

在本研究中，考虑到藏语是拼音文字(归桑拉姆, 2015; 高晓雷, 王永胜, 郭志英, 张慢慢, 白学军, 2015; 崔占玲, 张积家, 顾维忱, 2009; 崔占玲, 张积家, 2009b, 2009c; 张积家, 崔占玲, 2008)(归桑拉姆. (2015). *藏语文词汇通达中语音中介效应研究*. 拉萨: 西藏大学硕士学位论文论文; 高晓雷, 王永胜, 郭志英, 张慢慢, 白学军. (2015). 藏-汉双语者语义与词汇表征特点研究. *心理与行为研究*, 13(6), 737-743; 崔占玲, 张积家, 顾维忱. (2009a). 藏-汉-英三语者言语产生中的词汇选择机制. *现代外语*, 32(1), 51-58; 崔占玲, 张积家. (2009b). 藏-汉-英三语者语言联系模式探讨. *心理学报*, 41(3), 208-219; 崔占玲, 张积家. (2009c). 藏-汉-英三语者词汇与语义表征研究. *心理科学*, 32(3), 559-562. 298-302; 张积家, 崔占玲. (2008). 藏-汉-英双语者字词识别中的语码切换及其代价. *心理学报*, 40(2), 136-147.)，而在拼音文字的研究中，通常选用“X”符号作为掩蔽材料。因此，本研究沿袭以往拼音文字研究中掩蔽材料

的使用惯例，采用了“X”符号作为掩蔽材料。

文中掩蔽材料部分已经做了更详细的说明，详见文中掩蔽材料描述部分的红色字体，请审稿专家予以审查。

意见 2: 实验采用小学阅读材料，同时难度评定结果显示句子非常容易理解，却有 1/5 的大学生被试正确率偏低而被剔除，这种情况十分不正常，说明被试或材料或指导语可能存在问题。剩余的有效被试仅有 28 名，被试量偏少。

回应: 非常感谢审稿专家对本研究和稿件的肯定及所提出的宝贵意见。

(1)经过认真讨论和分析，我们认为，本研究中“有 1/5 的大学生被试正确率偏低”可能是因为：①被试的配合度不够导致的不认真作答致使错误率偏高；②被试初次参与实验所产生的紧张情绪导致按键操作失误致使错误率偏高；③被试虽然都是藏族大学生，藏语口语表达也基本不存在问题。但口语表达与书面阅读性质不同，能说不等于会读。且实际情况是，藏族大学生藏语水平存在明显的地区差异，学历并不是体现藏语书面阅读水平的指标。在选择被试之初，没有考虑到上述因素，可能导致部分藏语阅读水平极低者参与到实验当中致使错误率偏高；④本研究正确率标准设置过高及问题数量偏小也是正确率偏低的一个因素。

(2)在认真思考了审稿专家的意见后，我们努力对本研究的被试样本量进行了随机取样补充，现将研究的被试量扩充至 35 人，其中男生 13 人，女生 22 人。

具体修改和补充的内容如下：“*西藏大学 35 名藏族大学生(22 名女生，13 名男生)参与了本实验，平均年龄 $M=20.09(SD=1.12)$ 岁。所有被试视力或矫正视力正常，母语均为藏语。*”

详见文中被试描述部分的红色字体，请审稿专家予以审查。

意见 3: 实验数据分析不合理。作者指出，为了确定最大窗口的有效性而将其与整行条件相比较。结果表明在若干指标上存在显著差异(如，平均注视时间被试分析差异显著；阅读速度项目分析差异显著；向右眼跳幅度项目分析差异显著，等)，但是作者却无视以上的统计差异而指出：最大窗口条件下不影响阅读，所设定的最大窗口条件是有效的。此处逻辑不通。

回应: 非常感谢审稿专家所提出的宝贵意见。我们认真的讨论和思考了审稿专家的意见。对被试分析和项目分析有了进一步的认识，被试分析的目的是通过对实验处理效应进行检验，将实验结果推广到被试所代表的总体中。而项目分析的目的是通过对实验处理效应进行检验，将实验结果推广到项目所代表的材料总体中(丁国盛，李涛.(2006). *统计教程——从研究设计到数据分析*. 北京：机械工业出版社.)。通常情况下，两种分析的结果是一致的，也有存在分歧的情况。被试分析和项目分析结果的稳定说明所选被试和材料均能反应出一定的规律，能反映和代表各自的总体情况。实验的理想结果是被试和材料均具有代表性，但也有所偏重。本研究的目的是通过一定数量藏族大学生(被试)阅读不同窗口条件下藏语句子的实验操作，探测藏语阅读知觉广度的大小，从而将本研究结果推广到被试所代表的藏族大学生群体，而非本研究所选取阅读材料对于藏语代表性大小如何的问题。因此，在本研究中，被试分析的结果对于研究目的的支撑作用更为重要。另外，通过补充被试数量，实验结果表现出了更稳定的趋势，被试分析与项目分析的结果总体一致。

具体内容请看结果分析部分的红色字体，请审稿专家审查。

意见 4: 实验材料的介绍前后文不一致。作者先提到“改编形成 136 个句子”，但后文又指出“70 句实验材料分成 7 组，每组有 10 个句子”。那么，实验材料究竟是多少句？

回应: 非常感谢审稿专家所提出的宝贵意见。在采用眼动追踪技术开展阅读研究时，通常以句子和篇章为阅读材料，这些阅读材料不管是自编、改编或者是从已有文本中直接选取，为了符合被试所代表的群体的阅读水平，都要选取一定量的同质被试(不参加随后的正式实验)

对拟作为正式实验材料的这些阅读材料进行通顺性和难度的评定,最后根据实验要求从中选出通顺性和难度评定值在一定范围内的材料作为正式实验材料。因此,为了达到最终实验设计所需要的材料(句子、篇章)数量,在初选材料时,材料数量必须大于最终所需要数量,因为通顺性和难度要求会导致很多材料不符合实验要求。以往以眼动仪为工具所开展的阅读研究,如闫国利等(2013, 2014, 2011)、王丽红等(2014)(闫国利, 张巧明, 张兰兰, 白学军. (2013). 不同掩蔽材料对阅读知觉广度的影响. *心理科学*, 36(6), 1317-1322; 闫国利, 孙莎莎, 张巧明, 白学军. (2014). 自然阅读与校对阅读的知觉广度研究. *心理科学*, 37(2), 298-302; 王丽红, 白学军, 闫国利. (2014). 汉语阅读知觉广度的老化: 一项眼动研究. *心理与行为研究*, 12(6), 763-768; 闫国利, 王丽红, 巫金根, 白学军. (2011). 不同年级学生阅读知觉广度及预视效益的眼动研究. *心理学报*, 43(3), 249-263.), 皆采用了这种材料编制和选择方法。

本研究的主题是阅读知觉广度,阅读知觉广度研究要求最终正式实验材料是通顺性较高,而难度较低的句子。所以,本研究最初改编形成 136 个句子作为正式实验材料的备选材料,通过通顺性和难度评定后,从中选出 70 个句子作为正式实验材料,通顺性评定结果 $M=5.49(SD=0.36)$ (1 代表非常不通顺, 7 代表非常通顺), 难度评定结果 $M=1.15(SD=0.32)$ (1 代表非常容易理解, 5 代表非常难理解)。因为本研究采用单因素 7 水平(7 种窗口: 5、9、13、17、21、25 和整行)实验设计,实验材料的呈现采用拉丁方方法形成 7 种窗口条件的呈现顺序,每个被试随机接受其中一种刺激呈现顺序。70 句正式实验材料分成 7 组,每种顺序一组句子,每组有 10 个句子。

文中已对材料部分进行了更清晰的描述,详见文中实验材料部分的红色字体,请审稿专家予以审查。

意见 5: 请作者说明“校准的误差值控制在 0.25 以下”中的 0.25 指什么: 是正确率或其他指标?

回应: 非常感谢审稿专家所提出的宝贵意见。本文全体作者对此问题进行了认真的讨论,并做出如下解释说明: 校准的误差值控制在 0.25(李薇薇, 2013)(李薇薇. (2013). *维吾尔族大学生维吾尔语和英语阅读知觉广度的眼动研究*. 天津: 天津师范大学硕士学位论文.)以下”中的“0.25”指的是 Eyelink1000plus 型眼动仪使用过程中代表“注视准确度”的参数。在使用眼动追踪技术开展阅读研究时,按照仪器使用要求及以往研究的惯例,该误差值通常需控制在 0.25 以下,以保证最终数据分析时,尤其是以字词等局部区域为分析单位时,眼动仪自动追踪并记录的注视点数据信息与实际应注视位置的对应。请审稿专家予以审查。

审稿人 2 意见:

研究者采用眼动仪考查藏语的阅读知觉广度。研究设计科学,实验材料选择有中国语言特色,得到的研究结果对已有知觉广度的科学发现具有较高价值。评审人仔细阅读后,建议做以下修改。

意见 1: 最重要的要求: 请研究者用 linear mixed-effects model 处理数据,这样可以避免一系列由于被试和材料带来的潜在问题。研究者需要具体报告模型的结构和数据检验方法。

回应: 非常感谢审稿专家所提出的宝贵意见。针对审稿专家所提出的“用 linear mixed-effects model 处理数据”这一问题,全体作者进行了认真的讨论和深入的思考,深感审稿专家所提的问题意义重大。审稿专家所提的数据处理方法为本研究数据处理提供了新的思路,它背后的真知灼见令我们折服。审稿专家指出:“用 linear mixed-effects model 处理数据,这样可以避免一系列由于被试和材料带来的潜在问题”,这一观点令全体作者倍感兴奋。如审稿专家

所言，我们采用 linear mixed-effects model 数据处理方法，势必会在本研究实验数据处理的精准性和灵敏性上具有极大的助益作用。因此，再次感谢审稿专家所提的宝贵意见。

针对上述问题，全体作者进行了认真的思考和深入的学习。我们首先对 linear mixed-effects model 数据处理方法进行了系统的学习(SPSS, I. (2005). *Linear mixed-effects modeling in SPSS: an introduction to the MIXED procedure.*), 并就此咨询了暨南大学和华南师范大学了解该方法的两名教师。与此同时，因为本研究采用 SPSS20.0 统计软件进行数据处理，因此，我们也查阅了可能涉及该方法的多本统计书籍(温忠麟. (2016). *心理与教育统计(第二版)*. 广州: 广东高等教育出版社; 王璐. (2010). *SPSS 统计分析基础、应用与实践*. 北京: 化学工业出版社; 薛薇. (2009). *SPSS 统计方法及应用(第2版)*. 北京: 电子工业出版社; 张厚粲, 徐建平. (2004). *现代心理与教育统计学*. 北京: 北京师范大学出版社; 骆方, 刘红云, 黄崑. (2011). *SPSS 数据统计与分析*. 北京: 清华大学出版社; 张文彤. (2004). *SPSS 统计分析高级教程*. 北京: 高等教育出版社.)。最终，我们在张文彤(2004)主编的《SPSS 统计分析高级教程》中找到了涉及本研究数据类型(重复测量数据)而与 linear mixed-effects model 有关的一些论述(张文彤. (2004). *SPSS 统计分析高级教程*. 北京: 高等教育出版社, Page77.)。

该教程指出：重复测量数据包括两种类型：(1)同一时间点，或者不考虑时间因素的重复测量。比如多个医生对同一批病人的症状分别进行评定；同一个样品分别用多台仪器检测得到的数值等。(2)按时间顺序的重复测量数据，又称为纵向数据(Longitudinal Data)，在社会学和经济中则被称为 Panel Data，指个体随着时间推移被重复观测得来的资料。例如在调查研究中的前瞻性研究，对出生婴儿的身高、体重进行动态监测，定时测量其身高、体重以观察婴儿生长发育情况。从上述两种重复测量数据的举例中，我们可以发现，linear mixed-effects model 通常用来处理医学研究中的重复测量数据。而即便是在医学研究领域内部，对于重复测量数据处理，除可以采用混合效应线性模型(linear mixed-effects model)(又被称为多水平模型，它是一般线性模型的扩展)外，重复测量方差分析(基于最小二乘原理进行配合的一般线性模型)(General Linear Model-Repeated Measures)，也是常用的方法(高萌, 张强, 邓红, 宋魏. (2008). 方差分析和混合线性模型在重复测量数据中的应用探讨. *现代预防医学*, 35(7), 1221-1223.)。且通过文献查阅发现，混合效应线性模型(linear mixed-effects model)更多应用于存在时间顺序的医学研究中的重复测量数据处理(张莉, 黄莉, 熊昌辉, 李洪艳, 谭珊, 李迪民. (2015). 混合线性模型在婴幼儿生长发育研究中的应用. *中国卫生统计*, 32(1), 10-13; 陈长生, 徐勇勇, 曹秀堂. (1996). 医学研究中重复观测数据的统计分析方法. *中国卫生统计*, 13(6), 55-58; 秦正积, 沈毅, 王燕南, 肖静, 何书. (2014). 三种重复测量资料的统计分析方法比较研究. *中国卫生统计*, 31(3), 542-545; 余松林. (2006). 混合线性模型的应用. *中国医院统计*, 13(1), 70-75; 姚宁, 曾庆, 李革, 张婷, 窦贵旺. (2010). 重庆地区儿童尿碘重复测量数据混合线性模型研究. *广东微量元素科学*, 17(5), 25-30.)

而医学研究中的重复测量与心理学研究中的重复测量从某种角度上来说并不完全对应。本研究实验设计为单因素 7 水平被试内设计，所有被试均接受同样的实验处理(阅读相同的实验材料)，不考虑时间因素。因此，本研究所得到的重复测量数据属于张文彤(2004)重复测量数据分类中的第一类，即为同一时间点，或者不考虑时间因素的重复测量。基于心理学与医学的学科差异，在一些问题和概念的认识和理解，以及实际操作中临床与非临床实验操作思路和结果处理与检验方面既存在共性，也存在一些细小差别。在心理学研究中，针对重复测量数据，通常采用 General Linear Model-Repeated Measures 方法处理该类数据。采用这一方法，符合医学和心理学研究对于重复测量数据处理的共同要求，同时也沿袭和继承了心理学在该类型数据处理上的方法规范。且通过查阅文献，本研究同类研究中均采用了 General Linear Model-Repeated Measures 方法进行数据处理(李韵静, 张红葵, 白柳, 王越, 程小云, 范宏伟, 闫国利. (2015). 小学五年级学生英语阅读知觉广度的眼动研究. *心理与行为研*

究, 13(2), 225-229; 张玉晶. (2015). 不同水平五年级维吾尔族学生母语阅读知觉广度. *心理发展与教育*, 31(6), 703-709; 闫国利, 孙莎莎, 张巧明, 白学军. (2014). 自然阅读与校对阅读的知觉广度研究. *心理科学*, 37(2), 298-302; 闫国利, 巫金根, 王丽红. (2013). 小学五年级语文学优生与学困生阅读知觉广度的眼动研究. *心理科学*, 36(3), 622-626; 闫国利, 张巧明, 张兰兰, 白学军. (2013). 不同掩蔽材料对阅读知觉广度的影响. *心理科学*, 36(6), 1317-1322; 伏干, 闫国利. (2013). 初中二年级学生字词阅读知觉广度的眼动研究. *心理科学*, 36(2), 284-289; 闫国利, 张巧明, 白学军. (2013). 中文阅读知觉广度的影响因素研究. *心理发展与教育*, 29(2), 121-130; 巫金根, 闫国利, 刘志方. (2014). 大小字号的文本对中文读者阅读知觉广度和眼动模式的影响. *心理科学*, 37(1), 10-15; 吴捷, 张兰兰, 巫金根. (2010). 高中生英语阅读知觉广度的眼动研究. *心理学探新*, 30(2), 29-34.)。

在保证遵循(医学、心理学)重复测量数据分析处理规范和惯例的基础上, 为了增进本研究同类研究的可比性及在同类研究中的认同度, 本研究沿用了以往同类研究中共同采用的, 已被广泛认同的数据处理方法, 即采用 **General Linear Model-Repeated Measures** 进行了数据分析处理。请审稿专家考虑我们的思考及解释说明的合理性, 对我们所采用的数据处理方法进行再审核。

再次感谢审稿专家的辛勤工作。

意见 2: 由于这是一个单因素 7 水平的实验设计, 所以两两比较的 p 值需要校正。请研究者报告校正后的 p 值, 且报告校正的方法。

回应: 非常感谢审稿专家所提出的宝贵意见。我们对由于我们写作过程中的疏漏所带来的困惑致以歉意。在本研究中, 我们采用了 **SPSS20.0** 统计软件处理了实验数据, 在这一过程中, 采用了 **Bonferroni** 对两两比较的 p 值进行了校正。我们在文中所报告的两两比较的 p 值是经过 **Bonferroni** 方法校正过的。对于这一点, 我们已在文章修改过程中, 在两两比较部分对所采用的 p 值校正方法进行了补充说明。

详见文中两两比较结果部分的红色字体: “**两两比较 P 值采用 Bonferroni 方法进行校正**”, 请审稿专家予以审查。

意见 3: 第 8 页, “事后检验表明, R2 窗口下的平均注视时间显著长于 R4、R6、R8、R10、R12 和整行窗口($p < 0.05$)”。如果这里都显著, 那么请用 $ps. < 0.05$ 。论文其他部分存在相似问题, 请一并修改。

回应: 非常感谢审稿专家所提出的宝贵意见。审稿专家的细致和严谨令我们敬服, 我们在今后的写作中, 也将以此为契机, 不断提高写作要求, 精益求精。我们已根据审稿专家意见, 对文中事后检验部分($p < 0.05$)进行了修改。详见文中($ps. < 0.05$)红色字体, 请审稿专家予以审查。

意见 4: 在论证写作层面, 请结合习近平哲学社会科学讲话, 宏观概述该论文的学术价值。

回应: 非常感谢审稿专家所提出的宝贵意见。审稿专家的意见, 提升了我们所做研究的学术价值和社会高度, 高屋建瓴, 意义卓大。我们已根据审稿专家意见, 仔细研读了习近平主席哲学社会科学讲话的原文, 对其中的内容和精神主旨进行了认真的领悟与吸纳。并最大程度的将讲话精神与本研究相结合, 使之成为本研究的引领与归结。

详见引言部分红色字体: “然而, 目前有关藏语阅读的研究还很少(高晓雷, 王永胜, 郭志英, 张慢慢, 白学军, 2015; 崔占玲, 张积家, 2009; 张积家, 崔占玲, 2008), 而藏语阅读知觉广度这一基本阅读问题也尚不清楚。因此, 采用“移动窗口”范式开展藏语阅读研究, 不仅具有鲜明的民族特色, 对于藏语的传承和学习也具有重大意义。”

意见 5: 提高英文摘要表达水平。

回应: 作者再次对文中中英文摘要部分进行了认真的斟酌修改, 并请一名具有多年 SCI 英文文章发表经验的心理学专业副教授(博士后)对本研究英文摘要进行了进一步修改。

审稿人 3 意见:

几个具体问题:

意见 1: 摘要中 R2 等符号, 对于没有相关知识背景的不知道指代什么内容, 按照闫国利等人的研究, 一般写作 L2R2, 因为这里是左右两侧对称呈现, 不过这种呈现方式并不多见, 请在文章中写出为什么对称呈现, 是否是传统的探讨阅读知觉广度的方式。

回应: 非常感谢审稿专家所提出的宝贵意见。

(1)正如审稿专家所言, 在表述对称呈现的窗口时, 有研究者采用 L0R0、L1R1、L2R2、L3R3、L4R4 的形式, 如闫国利等(2014)(闫国利, 孙莎莎, 张巧明, 白学军. (2014). 自然阅读与校对阅读的知觉广度研究. *心理科学*, 37(2), 298-302)。也有研究者根据所研究文字的阅读方向, 只采用了代表文字向前阅读方向的字母(L 代表 left, R 代表 right)来标识窗口条件, 如维吾尔语从右向左书写和阅读, 李薇薇(2013)在采用对称窗口研究维吾尔语时, 分别在注视点两侧呈现 2、4、8、12、14、16 个字符, 分别产生 5、9、17、25、29、33 个字符空间的窗口, 将窗口条件表述为 L2、L4、L8、L12、L14、L16 (李薇薇. (2013). *维吾尔族大学生维吾尔语和英语阅读知觉广度的眼动研究*. 天津: 天津师范大学硕士学位论文.)。经过我们的讨论, 认为闫国利等(2014)的窗口条件表述方式可能相对更容易理解一些。因此, 根据审稿专家的意见, 我们已经将文中 R2、R4、R6、R8、R10、R12 等符号改为了 L2R2、L4R4、L6R6、L8R8、L10R10、L12R12。具体见文中 L2R2、L4R4、L6R6、L8R8、L10R10、L12R12 红色字体, 请审稿专家审查。

(2)对于审稿专家所提的“为什么对称呈现”的问题。全体作者进行了认真的讨论, 并在文中相应部分做了进一步的补充和完善。已有研究显示, 阅读知觉广度在向前的阅读方向上表现出更大的范围(Rayner, 1998; 闫国利等, 2014)(Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372-422; 闫国利, 孙莎莎, 张巧明, 白学军. (2014). 自然阅读与校对阅读的知觉广度研究. *心理科学*, 37(2), 298-302.), 知觉广度更大, 每次注视所能获取的信息更多, 对理解文本和提高阅读效率的贡献率更大。因此, 对知觉广度较大一侧进行探测更具意义。对称窗口能够最大限度的保证对知觉广度较大一侧范围的准确、完整探测(张巧明等, 2013; 闫国利等, 2014; 李韵静等, 2015)(张巧明, 王爱云, 闫国利. (2013). 大学生阅读知觉广度影响因素的回归分析. *心理与行为研究*, 11(2), 190-194; 闫国利, 孙莎莎, 张巧明, 白学军. (2014). 自然阅读与校对阅读的知觉广度研究. *心理科学*, 37(2), 298-302; 李韵静, 张红葵, 白柳, 王越, 程小云, 范宏玮, 闫国利. (2015). 小学五年级学生英语阅读知觉广度的眼动研究. *心理与行为研究*, 13(2), 225-229.), 而藏语是从左至右横向书写和阅读的, 在阅读这一语言时, 注视点右侧可能存在更大的知觉范围。因此, 本研究中, 根据藏语自左向右书写的特点和对称窗口能够最大限度的保证对知觉广度较大一侧范围的准确、完整探测这一优势, 选用了窗口对称呈现。

在实验设计部分已经对上述问题进行了进一步的修改和说明, 具体见实验设计部分红色字体, 请审稿专家予以审查。

(3)对于审稿专家所提的“对称呈现, 是否是传统的探讨阅读知觉广度的方式”的问题。全体作者进行了认真的讨论, 并做出如下解释。在采用移动窗口范式探测阅读知觉广度时, 可视窗口有注视点左右对称呈现(对称窗口)和注视点左右不对称呈现(非对称窗口)两种类

型。采用上述两种窗口设置方式开展阅读知觉广度的研究都很普遍。如闫国利等(2011, 2013)阅读知觉广度的研究中采用了不对称窗口(闫国利, 王丽红, 巫金根, 白学军. (2011). 不同年级学生阅读知觉广度及预视效益的眼动研究. *心理学报*, 43(3), 249-263; 闫国利, 张巧明, 张兰兰, 白学军. (2013). 不同掩蔽材料对阅读知觉广度的影响. *心理科学*, 36(6), 1317-1322.); 而张巧明等(2013), 闫国利等(2014), 李韵静等(2015)阅读知觉广度的研究中则采用了对称窗口。

请审稿专家予以审查。

意见 2: 第四页笔误, “在阅读阿拉伯语时, 读者在向左不对称的窗口条件下阅读表现更好”。
回应: 非常感谢审稿专家所提出的宝贵意见。根据审稿专家的意见, 我们再次查阅了相关文献, 进行了确认。结果发现: “在阅读阿拉伯语时, 读者在向左不对称的窗口条件下阅读表现更好”的表述无误。

“Arabic, which is one of the world’s most widely read languages and is read from right to left.....When English was read, performance across window conditions was superior when windows extended rightward. However, when Arabic was read, performance was superior when windows extended leftward and was essentially the reverse of that observed for English.”(Jordan, T. R., Almabruk, A. A., Gadalla, E. A., McGowan, V. A., White, S. J., Abedipour, L., & Paterson, K. B. (2014). Reading direction and the central perceptual span: Evidence from Arabic and English. *Psychonomic Bulletin & Review*, 21(2), 505-511.)

请审稿专家予以审查。

意见 3: 第四页, 介绍藏语时, 藏语的字和词是怎么区分的, 词具有心理现实性, 而这里提到了藏语中的字的符号切分, 读者会想知道藏语中的构词规则?

回应: 非常感谢审稿专家所提出的宝贵意见。根据审稿专家意见, 我们进行了认真的思考, 咨询了西藏大学藏文研究专家王志敬教授和拉巴泽仁教授, 并深入查阅了相关文献, 做出如下解释, 并在文中藏文介绍部分补充了关于藏语构词规则的详细说明。

(1)关于审稿专家提出的“藏语的字和词是怎么区分的”这一问题。通过咨询藏文研究专家和查阅文献, 我们得出如下认识。藏语字和词的区分与现代汉语类似, 又有所不同。要厘清藏语字和词是如何区分的, 首先要弄清楚现代汉语中几个相互联系的概念: 音素、音节、字、语素和词。音素是根据语音的自然属性划分出来的最小语音单位; 音节是听觉能感受到的最自然的语音单位, 音节由一个或几个音素按一定规律组合而成。语素是语言中最小的音义结合体, 词是句中最小的能够独立运用的语言单位, 字是记录语言的书写符号。语素是构词单位, 词是造句单位, 记录语素和词的书写单位是字。在汉语中, 汉字基本上是和语言中的语素相对应的, 一般情况下一个汉字记录一个音节。因此, 单音节中的语素或只有一个单音节语素构成的词, 都用一个字表示, 此时, 词、语素、字三者一致。比如: “地、牛、火”。多音节语素或多音节词, 其每个音节需一个字表示, 这时的词、语素和字便不完全一致, 比如: “仿佛”是一个语素组成的词, 用两个字表示; “葡萄干”是用两个语素组成的词, 用三个字表示。

藏语字和词的区分与现代汉语类似, 又有所不同。藏文是参照梵文字母体系创制的辅音字母式的音素拼音文字, 在拼音文字里, 表音的字母是单词的重要组成部分, 音素和字母对应。藏语以音节为拼写单位(张积家, 崔占玲. (2008). 藏-汉-英双语者字词识别中的语码切换及其代价. *心理学报*, 40(2), 136-147.), 多个字母按一定规则组合构成一个音节, 即一个藏文字。藏语字形结构均以辅音字母为核心, 称为“基字”, 一个基字可以单独构成一个字, 也存在多个辅音字母围绕一个“基字”前后附加和上下叠写构成字的情况, 元音符号加在辅音

字母的上、下、正中(归桑拉姆. (2015). *藏语文词汇通达中语音中介效应研究*. 拉萨: 西藏大学硕士学位论文.)。与现代汉语类似, 藏语字基本上是和语言中的语素相对应的, 一般情况下一个藏语字记录一个音节。因此, 只有一个单音节语素构成的词, 都用一个字表示, 此时, 词、语素、字三者一致。多音节语素或多音节词, 其每个音节需一个字表示, 这时的词、语素和字便不完全一致, 可能用两个字表示一个语素、一个词; 也可能用三个字表示两个语素、一个词。

(2)关于审稿专家提出的“藏语中的构词规则”这一问题。通过查阅文献, 我们找到了早期对这一问题的经典论述。藏语的构词规则如下: ①由一个单音节构成一个词, 如“**ལྲོ་** (水)”、“**ང་** (我)”, 也存在由一个有意义的单音节加一个成词的附加成分(前加的为数很少, 主要是后加的)构成一个完整的词的情况, 藏语中名词和形容词都可以作为附加成分, 且以形容词居多, 如“**མ་མ་** (妈妈)”、“**ལ་ལ་** (伯父)”、“**འང་པ་** (房子)”; ②由单音节词素构成复合词, 如“**ང་ཉིད་** (自己)”、“**ཚོ་སྲོག་** (生命)”; ③在单音节词后加一个构词的附加成分产生新词, 如“**ཚོང་པ་** (商人)”、“**ཉ་པ་** (渔夫)”; ④由多个有意义的单音节词构成多音节词, “**ཕྱི་ཚོགས་འིང་ལྷགས་** (社会主义)”、“**ཐོས་ཅད་ཚོས་བཏང་** (不顾一切)”(金鹏. (1986). *汉语和藏语的词汇结构以及形态的比较*. *民族语文*, 3, 1-13.)。

关于藏语构词规则已在文中藏语介绍部分进行了补充完善, 修改内容详见文中藏语介绍部分的红色字体, 请审稿专家予以审查。

意见 4: 第五页, 请简单解释为什么阅读材料选择小学读本?

回应: 非常感谢审稿专家所提出的宝贵意见。针对审稿专家的问题, 我们经过认真思考, 做出如下解释说明:

在选择小学读本为阅读材料时, 我们首先考虑到材料难度对阅读知觉广度的影响。

研究表明, 材料难度影响阅读知觉广度。当材料难度增加时, 知觉广度会减小。特别是注视点的内容的难度, 即中央凹负荷会影响我们对旁视野信息的提取(闫国利, 2010)(闫国利, 巫金根, 臧传丽, 白学军. (2010). 阅读知觉广度眼动研究述评. *心理学探新*, 30(2), 23-28.)。在 Rayner(1986)的研究中, 让不同的被试阅读与其年龄水平相当的材料, 小学四年级和大学生阅读两种句子, 其中一种是符合小学生阅读水平的句子, 另一种是适合大学生阅读的句子。研究结果发现, 句子难度影响知觉广度的大小。当句子难度较大时, 读者需要更多的注意加工落在中央凹区域的单词, 使其知觉广度变小。伏干(2012)采用呈现随眼动变化技术, 考察材料难度和窗口大小对大学生中文句子阅读的影响。结果发现: (1)难材料阅读知觉广度为被注视汉字左侧 1 个汉字到右侧 2 个汉字之间; (2)易材料阅读知觉广度为被注视汉字左侧 1~2 个汉字到右侧 3~4 个汉字之间; (3)知觉广度的左右不对称性受阅读材料难易度的影响, 较易材料的阅读知觉广度的不对称性大于较难材料的阅读知觉广度(伏干. (2012). 材料难易度对汉语阅读活动影响的眼动研究. *心理研究*, 5(5), 39-45.)。

本研究采用小学五年级水平藏文为阅读材料, 首先基于藏语阅读知觉广度的具体范围尚不清楚, 且阅读知觉广度研究要求实验材料要比较容易。同时, 考虑到材料难度对阅读知觉广度存在影响, 大学生作为熟练阅读者, 理论上具有更大的知觉广度(这一点已在以往研究中得到证实), 而小学五年级水平藏文相对于大学生的阅读水平, 又相对较容易。那么, 让大学生阅读相对容易的小学五年级水平藏文材料, 有利于最大限度的探测出藏语阅读知觉广度的具体范围, 这将成为后续藏语阅读相关研究的参考, 也是后续采用移动窗口范式开展藏语阅读知觉广度研究的基石。

请审稿专家予以审查。

意见 5: 第五页, 整句条件和其它条件的长度不同, 是否选用任何字体和字号都不能保证等长, 国外一般都用 courier 字体, 来确保每一个字符等长, 这里是否存在藏文符号的特殊性?

回应: 非常感谢审稿专家所提出的宝贵意见。

(1)审稿专家提出“整句条件和其它条件的长度不同, 是否选用任何字体和字号都不能保证等长”。通过研读, 我们考虑到此问题背后可能蕴含了句子不等长是否会影响实验处理效果这一问题, 我们首先对这一问题做出回答。

首先, 本研究采用的是项目内设计, 共有 7 种窗口条件。同时, 采用拉丁方的方法形成了 7 种窗口呈现顺序(7 个 block), 保证每个句子在每种窗口条件下都会呈现 1 次。因此, 每个句子自身的特征, 如字长、句长不会影响最终的实验处理的效果(Inhoff, A. W., & Liu, W. (1998). The perceptual span and oculomotor activity during the reading of Chinese sentences. *Journal of Experimental Psychology: Human Perception and Performance*, 24(1), 20-34.)。

其次, 本研究关注所选取的 5 个眼动指标: 平均注视时间、注视次数、总注视时间、阅读速度和向右眼跳幅度的值均是以字符为单位求得值, 而不是以句子为单位求得值, 且均为平均值。因此, 单个句子的句长不会影响到最终所求得均值的大小。

再次, 采用眼动仪研究阅读过程, 包括阅读知觉广度的研究, 由于要保证实验材料(句子)的代表性, 不可能做到每个句子长度都一样。以往同类研究也往往采用的是将句长控制在一定区间内的做法。如李韵静(2015)的研究中将每个句子长度控制在 7 到 8 个单词之间(李韵静, 张红葵, 白柳, 王越, 程小云, 范宏伟, 闫国利. (2015). 小学五年级学生英语阅读知觉广度的眼动研究. *心理与行为研究*, 13(2), 225-229.); 闫国利等(2014)的研究中将实验句句长控制在 17-20 个字(闫国利, 孙莎莎, 张巧明, 白学军. (2014). 自然阅读与校对阅读的知觉广度研究. *心理科学*, 37(2), 298-302.)。伏干(2013)的研究中的句长为 14-16 个字(伏干, 闫国利. (2013). 初中二年级学生字词阅读知觉广度的眼动研究. *心理科学*, 36(2), 284-289.)。本研究中, 采用了将句长控制在 17~23 个字符的办法。

(2)审稿专家指出“国外一般都用 courier 字体, 来确保每一个字符等长, 这里是否存在藏文符号的特殊性?”

藏文字符确实存在一定的特殊性。藏语字形结构均以辅音字母为核心, 称为“基字”, 一个基字可以单独构成一个字, 也存在多个辅音字母围绕一个“基字”前后附加和上下叠写构成字的情况, 元音符号加在辅音字母的上、下、正中(归桑拉姆, 2015)。

与英语、汉语相比, 藏语特点非常独特: (1)作为拼音文字, 由视觉感官简单的字母构成字单元, 字的书写表现出一定的线性发展特征, 虽然存在字间标记, 但并没有英语的词间空格那么明显; (2)前后附加和上下叠写构成字(如“འཇམ་མགོན་པོ་འཕགས་པ་”)的书写特点, 又使其与汉字的左右、上下等构字结构类似, 具有一定的立体性; (3)藏语构字虽然内含立体结构, 但文字透明性(透明性指字形单元与语音单元对应的一致性程度, 即由形知音的程度)(张玉晶等, 2015)与汉字相比, 又相对很强(如汉字“我”, 无法由字形直接读出字音; 而其藏语对译字“ཨོ་ཨོ་”则可由藏文字母读音, 直接读出该字读音)。

文字结构不同带来的字词加工复杂性的差异, 字/词间信息是否明显及文字透明性等因素都会影响阅读的速度和信息提取的效率, 进而影响阅读知觉广度的大小。藏语在上述方面表现出与英语和汉语相同或不同的一些特点(见表 1)。

由于藏文构字既可以单辅音字母为一个字, 也存在上下叠写构字的情况。因此, 作为构字成分的藏文字符的宽度(长度)基本是相同的(约 15 像素), 但高度确是不一致的。在研究之初, 我们曾尝试采用 BatangChe(闫国利, 张兰兰, 郎瑞, 白学军. (2008). 大学生英语阅读知

觉广度的眼动研究. *心理研究*, 1(2), 80-85.)来保证每个字符大小(长×高)一致, 但改变字体并未改变藏文字符的大小。根据审稿专家的意见, 我们尝试了 *courier new* 字体, 同样无法改变藏文字符的大小。可见, 藏文字符确实具有特殊性。

表 1 藏语与汉语、英语的特点比较

	语言类型		结构		字/词间标记		透明性	
	拼音文字	表意文字	线性结构	立体结构	隔字符	空格	发音透明	发音不透明
藏语	是	否	有	有	有	无	是	否
汉语	否	是	无	有	无	无	否	是
英语	是	否	有	无	无	有	是	否

请审稿专家予以审查。

意见 6: 第六页显示器型号? 什么显示器能设置 140 刷新率。

回应: 非常感谢审稿专家所提出的宝贵意见。我们经过认真讨论和思考, 做出如下解释:

闫国利等(2010)指出: 移动窗口范式、移动掩蔽范式、边界范式、快速启动范式和消失文本范式这五个范式均要求被试机的显示器屏幕上呈现的内容要根据读者眼睛注视位置来做相应的瞬间改变, 这就对显示器的刷新频率有较高的要求。如果显示器的刷新频率低, 则会出现刺激呈现变化不及时, 直接影响实验结果。因此, 如果使用上述范式进行阅读研究时, 建议采用刷新频率大于 100Hz 的 CRT 显示器(闫国利, 巫金根, 胡晏雯, 白学军. (2010). 当前阅读的眼动研究范式述评. *心理科学进展*, 18(12), 1966-1976.)。本研究采用了移动窗口范式来探测藏语阅读知觉广度的大小, 据上述所言, 需要使用刷新频率大于 100Hz 的 CRT 显示器, 这是以往研究的实验要求和惯例。而目前日常生活中常见的多是液晶显示器, 且刷新频率多为 60 Hz 或 75Hz。因此, 这类显示器不能满足本研究实验要求。为此, 我们购买了最大刷新率更高的 CRT 显示器, 以满足移动窗口范式对被试机显示器刷新频率的要求。以下是我们所使用的被试机显示器详细信息。

产品名称: 索尼CPD-G520	显像管品牌: 索尼超微细点距特丽珑...	垂直扫描频率: 48Hz-170Hz
水平扫描频率: 30KHz-130KHz	最大分辨率: 2048x1536	尺寸: 21英寸
是否短管: 否	偏转角度: 108°	最佳分辨率: 1600x1200
刷新率: 341MHz	接口类型: D-SUB15	显像管类型: 纯平
是否支持高亮: 不支持	品牌: SONY	索尼型号: CPD-G520

请审稿专家予以审查。

意见 7: 第九页与前面数据分析有重合, 建议简化分析, 避免累赘。

回应: 非常感谢审稿专家所提出的宝贵意见。我们通过查阅同类研究相关文献, 经过认真讨论后, 在保证实验数据完整、规范的基础上, 对数据分析的呈现与表述进行了简化。

详见文中数据分析部分的红色字体, 请审稿专家予以审查。

第二轮

审稿人 1 意见:

意见 1: 论文作者较好地回答了审稿人提出的问题, 对论文进行了必要修改。审稿人基本满意。同时, 基于 *generalized linear mixed model* (不是 *general*) 的 *linear mixed-effects model* 或者 *logic mixed-effects model*, 当前国内研究者还不是十分了解, 特别是针对不同实验设计的

模型建构和数据处理。所以，论文研究者仍然采用了 F1+F2 的方式进行被试分析和项目分析。由于报告的数据比较一致，所以审稿人对这种数据分析方式表示理解。我们正在结合使用实践和国外最新发展，着手写作一篇具有操作性的综述性论文，包括问题起源、统计学概念和操作、基于实验设计的模型建构、在语言研究数据分析中的实践应用(含眼动和 EEG)，期待能够推动国内语言认知研究者准确或者加深认识。希望论文作者关注和学习，在后续语言认知研究中尝试使用，从而提升数据分析能力。

回应：非常感谢审稿专家对我们的论文所做的工作。感谢审稿专家对我们的数据处理方法的认可和理解。我们期待审稿专家关于 linear mixed-effects model 综述性论文的早日面世，本文作者将密切关注，认真学习，并在后续研究中不断尝试使用，努力提升我们的数据分析能力。

再次感谢审稿专家的理解，以及提出的宝贵建议和期许。

第三轮

主编意见：

研究通过采用眼动的方式探讨了藏语阅读的知觉广度，对于理解藏语阅读特性以及知觉广度的语言加工的特异性具有理论意义和实际意义，文章的逻辑和设计符合要求，进一步完善论文写作的建议如下：

意见 1：P20 方法部分，2.1 被试 第一句，“平均年龄 20.09(SD=1.12)岁”，此处的括号为英文符号，请作者再次审查全文进一步规范写作。

回应：非常感谢主编在本文书写细节上提出的宝贵意见，本文全体作者接受主编意见，对文中有括号使用处进行了全面审查，并对误用处进行了一一修改，具体修改处已在文中用红色字体标出，请主编审查。

意见 2：P23，第 1 段第 2 行，“平均注视时间， $F(6,198)=19.63$ ， $\eta^2=0.37$ ， $p<0.001$ ； $F(6,330)=32.94$ ， $\eta^2=0.38$ ， $p<0.001$ ”。ANOVA 不能得到 η^2 ，而应该是偏 η^2 ，而且需要斜写，请作者确认并注意表述。

回应：非常感谢主编提出的宝贵意见，本文全体作者接受主编意见，并做出了如下修改和说明：文中 ANOVA 部分报告的确实是 Partial Eta Squared(η_p^2)值，我们已对符号的书写方式进行了修改，并用红色字体标出，请主编审查。

意见 3：P24 最后一段最后一句话与 P25 第二段最后一句话重复，显得表述冗余。如果两段内容都是针对同一个结果在进行讨论，那么建议作者是否可以考虑一下合并两段内容。

回应：非常感谢主编提出的细致入微的意见，本文全体作者接受主编意见，并做出了如下修改和说明：正如主编所言，这两段内容都是针对同一个结果在进行讨论，因此，我们遵循主编建议，将两段内容做了合并处理，请主编审查。

意见 4：P26 文献引用存在格式问题，例如“Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The quarterly journal of experimental psychology*, 62(8), 1457-1506.”，杂志的实意单词应该首字母大写，请作者再次通查全文。

回应：非常感谢主编提出的细致意见，本文全体作者接受主编意见，并已通查全文，尤

其是参考文献部分，并按主编意见进行了逐一修改，改动部分详见参考文献部分的红色字体，请主编审查。

意见 5: P20 2.1 被试 关于被试数量问题，之前也有审稿人提到，建议作者解释，为什么选取这个数量大小的被试，是出于什么考虑？又为什么认为将被试增加到了 35 个人就可以满足之前审稿人提到的被试数量问题？

回应: 非常感谢主编提出的宝贵意见，本文全体作者经过认真的讨论和思考，针对主编提出的有关被试量的两个问题做出如下解释和说明：

(1)针对第一个问题，也就是本研究最初为什么选取了 28 名被试。我们出于如下两个方面的考虑：①效应量大小。根据 J. Cohen(1969, 1988)的计算方法，在单因素方差分析中，效应量为 0.10 属于小的效果，效应量为 0.25 属于中等效果，效应量为 0.40 属于大的效果；使用 G-Power 软件，可以通过预期效应量计算出所需被试量的大小；在使用该软件计算被试量大小时，根据 J. Cohen(1969, 1988)的计算方法，通常将预期效应量设为 0.25。经过计算，本研究的合适被试量为 21 人(Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.)。②本研究为被试内设计，被试内设计需要的被试量少，操作方便、有效，但也存在可能产生顺序效应的缺陷。作为单因素 7 水平被试内实验设计，为了平衡可能产生的顺序效应，本研究采用拉丁方设计形成 7 种窗口条件的呈现顺序，也就是 7 个 block，每个被试随机接受其中一种刺激呈现顺序(完成一个 block)。因此，所需要的被试量应为 7 的倍数，28 是 7 的 4 倍。在满足上述两个条件的前提下，为了获得更稳定的实验效果，本研究在预期具有中等实验效果的 21 名被试的基础上，增加了 7 名，最终选取了 28 名被试。

(2)针对第二个问题，也就是“为什么认为将被试增加到了 35 个人就可以满足之前审稿人提到的被试数量问题？”。在专家审稿(退修一)中，有专家提出“有效被试仅有 28 名，被试量偏少”，因此，我们对本研究的被试量进行了随机取样补充，将被试量扩充至 35 人。之所以我们认为 35 个人可以满足之前审稿人提到的被试数量的要求，解释如下：①在统计学中，样本数为 30 是大样本和小样本的分界线，35 个人的样本量已属于大样本，从样本量对实验效果影响的角度来讲，35 个人的大样本产生的实验效果较之前更稳定；②本研究为单因素 7 水平被试内实验设计，为了平衡可能产生的顺序效应，采用拉丁方设计形成 7 种窗口条件的呈现顺序，也就是 7 个 block，每个被试随机接受其中一种刺激呈现顺序(完成一个 block)。因此，所需要的被试量应为 7 的倍数，35 是 7 的倍数，保证了每个 block 下都有 5 个被试；③被试增加到 35 人证明了实验效应是稳定的。我们比较了在 28 名被试和 35 名被试两个样本量的情况下，五个眼动指标上的平均效应量：28 名被试时为 0.395；35 名被试时为 0.40，实验效果相似，这也再次说明了无论是 28 名被试，还是 35 名被试都可以得到很稳定的实验效果。因此，基于上述三个方面的考虑，我们认为，将被试增加到了 35 个人就可以满足之前审稿人提到的被试数量的要求了。

请主编审查。