

《心理学报》审稿意见与作者回应

题目：阅读水平调节儿童阅读眼动注视模式的发展：基于9~11岁儿童的证据

作者：梁菲菲;王永胜;杨文;白学军

第一轮

审稿人1意见：

总评：该篇文章探讨了阅读模式的发展变化，分别以年级和阅读水平作为两个变量，每个变量分为三个水平，来探讨阅读模式的发展即由不成熟到成熟，是什么原因导致的。作者区分了两个重要理论，基于阅读水平发展的熟悉型的假设，以及基于年龄变化的眼球运动协调假设。总体来看，作者探讨的问题具有较大理论价值，文笔较为流畅，但研究方法和控制手段存在一定瑕疵。一些问题需要澄清。

意见 1：3-5年级的划分，是否能够对应眼球的生理成熟度的等级划分，3年级的同学可能8岁，四年级的同学也可能8岁或者九岁，这些信息需要呈现出来。此外，希望有一个更客观的指标来衡量生理成熟度。

回应：3-5年级的划分是严格对应儿童生理年龄的。在选择被试时，我们首先做的是在各年级中选取入学年龄相同的儿童作为测试对象，根据其语文水平的高低，分别选取阅读水平高、中、低三组被试参加正式的眼动实验。例如，在三年级被试的选择上，我们首先选择2006年9月1日-2007年8月31日出生，并在2013年9月1日正式进入三年级的儿童作为测试对象。由于眼动实验是在三年级下半学期进行的，因此，统计出来的平均年龄是在9.1岁。非常抱歉在撰写论文的时候，未注意到用“年级”一词表达儿童的生理年龄会带来理解上的偏差。感谢专家提出并予以修改的机会，为了明确这一点，在文中，我们摒弃以“年级”作为区分生理成熟度的说法，而改用儿童的实际生理年龄作为生理成熟度的反映，即在文中用儿童的“生理年龄”来代替“年级”，具体修改详见文中蓝色字体部分。

意见 2：阅读材料选择评分都比较容易，是否因为比较容易，在平均注视时间等指标上达到了天花板效应，导致没有出现阅读水平的主效应。

回应：专家提出的解释有存在的可能性。当查阅并分析相关文献之后（Blythe & Joseph, 2011; Chen et al., 2003; Feng et al., 2009; Reichle et al., 2013; 韩玉昌等人, 2005），发现阅读材料简单似乎不是导致阅读水平主效应在平均注视时间上未达到显著的主要原因，证据如下：首先，Blythe 等人（2011）在综述儿童阅读眼动特征发展的研究报告中呈现出十篇关于不同年龄儿童平均注视时间的数据。研究结果一致发现，不论不同年龄被试阅读与年龄匹配的文本，阅读水平匹配的文本，还是阅读不同难度的文本，其平均注视时间随儿童生理年龄增长均呈现出逐步下降的发展趋势。因此，Blythe 等人总结道，儿童阅读眼动注视特征的发展趋势受文本难度的影响很小；第二，Feng 等人（2009）发现，即使儿童的平均注视时间随着生理年龄的增长而缩短，但平均注视时间的缩短也主要表现在长注视的减少上，其峰值均在180ms左右。这和本研究结果是一致的。如果是由于阅读材料简单，平均注视时间达到“天花板效应”的话，平均注视点的分布应该也达到“天花板”效应。因此，本研究结果似乎表明，当儿

童阅读水平差异较小时，在注视时间上的差异应该反映在其分布，而不是平均数上。

意见 3: 读材料呈现之后，一个关键控制缺少，即没有问题回答环节，作者如何确保读者认真阅读。

回应: 为了让儿童尽量在自然情境下阅读（阅读过程不受问题设置的影响），本实验没有在每一篇阅读材料后设置问题环节，而是当儿童在眼动仪上完成五篇阅读材料的阅读之后，由实验者口头提问与阅读材料相关的问题，如“你能简单说说“小白菜”这一篇短文讲述了一个什么故事吗？”，进入统计分析的所有被试均能叙述出每一篇阅读材料的主题内容，表明所有被试在实验过程中均认真阅读了实验材料。已在“实验程序部分”将如上内容进行了补充，详见 p4倒数第二段黄色字体部分。

意见 4: 者的假设有三种，在我看来应该有四种预测，只有阅读水平主效应，只有年级主效应，两个主效应都有却没有交互作用，两个主效应都有又有交互作用。这样讨论从理论预测上更加完整一些。因为作者在前期无法给出某一个偏向有没有交互作用。

回应: 感谢专家提出了更为严谨的研究假设！已根据专家建议对原有假设进行了修改，详见 p3第一段黄色字体部分。

意见 5: 为什么四年级的期末考试低分组这么低，与其它组别没有显著差异吗？对结果会有哪些影响。

回应: 四年级低分组儿童期末考试成绩低的原因可能有两个：第一，期末考试试卷相对较难；第二，四年级低分组儿童阅读水平的确较低。如果是第二个原因的话，我们预期，四年级低分组儿童在期末考试和语文阅读水平测试中的成绩均显著低于其他年级低水平组儿童，然而检验结果发现，四年级低分组儿童只有在期末考试成绩中显著低于三、五年级低分组儿童， $t_s > 2.8, p_s < 0.05$ ；在语文阅读水平的测试成绩上，四、五年级低分组儿童不存在显著差异， $t(28) = 1.12, p > 0.05$ 。这似乎表明，造成四年级低分组儿童期考考试成绩较低的原因主要和期末考试试卷的难度有关。需要说明的是，由于不同年级进行测试时用到的期末考试试卷和语文阅读水平测试题目均适合各年级的阅读水平，即各个年级用到的测试题目均不一样，因此，各年级间低分组儿童的可比性意义不大。此外，我们在选择被试时，虽然是基于两次考试成绩的总和进行的大排名，同时也考虑到所选被试两次考试排名的一致性，即两次考试排名之间的差异小于7。按照上述解释，我们认为四年级低分组儿童期末考试较低存在一定的偶然性，对实验结果可能不会造成系统性的影响。

意见 6: 作者呈现了注视点时间分布，但只给出了描述统计，无法做出推论性信息，是否可以用别的统计方法，或拟合手段，将不同组的差异统计出来。

回应: 感谢专家提出的建设性完善建议。我们增加了注视点时间分布图的统计检验，即对其分布图的偏态系数做了相应的差异检验。结果详见 p5倒数第一段黄色字体部分以及 p7黄色字体部分。

审稿人 2 意见:

该研究对3-5年级儿童阅读中眼动模式与阅读水平的关系进行了系统研究，结果表明，

不同阅读水平者间眼动模式的特点在不同年级之间存在差异。研究在材料准备、检验、被试、结果呈现等方面清晰、扎实，表达清楚。主要问题：

意见 1：用学生年级作为年龄的测量指标，并进一步作为生理成熟的指标，以不同年级表现出阅读水平与眼动模式关系的差异作为探讨成熟和语言经验作用的证据，极不恰当。因为年级虽然可能与年龄高相关，但是年级本身反映着受教育的情况，反映着系统阅读（教育和学习）经验的多少。因此，要么对于研究问题、研究的假设、逻辑做重大修改，要么引入能够真正代表生理成熟的变量重新进行分析。

回应：根据专家的建议，摒弃以“年级”作为区分生理成熟度的说法，而改用儿童的实际生理年龄作为成熟度的反映。事实上，在选择被试时，我们首先做的是在各年级中选取入学年龄相同的儿童作为测试对象，根据其语文水平的高低，分别选取阅读水平高、中、低三组被试参加正式的眼动实验。例如，在三年级被试的选择上，我们首先选择2006年9月1日-2007年8月31日出生，并在2013年9月1日正式进入三年级的儿童作为测试对象。但由于眼动实验是在三年级下半学期进行的，因此，统计出来的平均年龄是在9.1岁。非常抱歉在撰写论文的时候，未注意到用“年级”一词表达儿童的生理年龄会带来理解上的偏差。感谢专家提出并予以修改的机会，为了明确这一点，在文中用儿童的“生理年龄”来代替“年级”，具体修改详见文中蓝色字体部分。

意见 2：问题提出第四段“语言熟悉性假说”（linguistic proficiency hypothesis）应为“语言熟练度假说”；“眼球运动协调性假说”（oculomotor-tuning hypothesis）是不是改为“眼球运动调试假说”更合适？

回应：根据专家提出的建设性修改意见，已将“语言熟悉性假说”修改为“语言熟练度假说”；“眼球运动协调性假说”修改为“眼球运动调试假说”，详见文中黄色字体部分。

意见 3：方法部分“为了排除排名最后的被试存在阅读障碍的可能性，选择倒数第6至20名作为低阅读水平组被试”，为什么？理论的必要性和方法的合理性何在？

回应：从倒数第6名开始选择低阅读水平组儿童，是因为在向各班级语文任课教师访谈时发现，排名倒数第一和倒数第五的两名9岁三年级儿童是阅读困难儿童，他们在平时的学习中存在很多困难，不适合参加实验。虽然其他两个年级倒数后五名的儿童没有该现象，但为了在三个年级中实施统一的被试筛选标准，因此，均从倒数第6名开始选择。

意见 4：者强调该研究对生理年龄（其实只是日历年龄吧）进行了控制，请提供有关生理年龄的控制或匹配证据，即除了描述年级特征外，应当具体说明在3个年级中，不同阅读水平组的生理年龄均是匹配的。

回应：对各年龄段不同阅读水平儿童的生理年龄进行了差异检验。单因素方差分析结果发现，各个年龄段不同阅读水平组被试的生理年龄不存在显著差异（ $F_s > 4.65, p_s < 0.05$ ），表明个阅读水平组儿童的生理年龄是匹配的，具体补充详见 p4 第一段黄色字体部分。

意见 5：在“实验设计”描述“年龄（包括三、四、五年级）”，不准确，此处并没有提供年龄信息而是年级，而年级不适合作为生理成熟水平的指标，实际上包括了阅读经验因素。

回应：由于全文已摒弃了“年级”的说法，实验设计也相应改成了年龄（包括9、10、11岁三

个年龄组)。具体修改详见 p4 实验设计部分。

意见 6: 3.1”对注视点持续时间分布进行分析, 结果发现三个年级儿童表现出了不一致(见图1右)”请提供具体的统计量(包括效应量)。

回应: 感谢专家提出的建设性修改建议。我们增加了注视点时间分布图的统计检验, 即对其分布图的偏态系数做了相应的差异检验。结果补充详见 p5 倒数第一段黄色字体部分以及 p7 黄色字体部分。

意见 7: 在各项有关不同年级中阅读水平调节作用差异的分析中, 请补充年级和阅读水平交互作用的检验。有关不同年级中阅读水平的不同调节作用分析应首先提供年级和阅读水平交互作用的检验; 当交互作用检验显著, 进一步分年级开展有关阅读水平作用的分析和比较。目前没有交互作用的分析, 只有分年级的观察, 请进一步完善。

回应: 参照关于儿童阅读发展文献中的大部分做法(如 Blythe 等人, 2012), 当儿童阅读的是与阅读水平相匹配的实验材料时, 虽然阅读材料在难度、通顺性、易读性等方面不具有显著差异, 但由于不同年龄儿童阅读的实验材料不同这一事实, 不适合将年级作为一个自变量进行分析。鉴于此, 在本研究中进行交互作用分析的意义可能不太大。在 p4 实验设计部分有详细说明, 请见黄色字体部分。

第二轮

审稿人 1 意见:

该篇是我审过的文章, 探讨了阅读模式的发展变化, 在第一轮审稿中提出的统计以及具体变量的澄清问题, 已经得到很好的解决, 建议发表。

审稿人 2 意见:

意见 1: 作者根据前一轮审稿意见进行了修改, 并说明了修改的考虑。关于问题提出第四段”语言熟悉性假说”(linguistic proficiency hypothesis) 应为”语言熟练度假说”。作者虽然在回信中表示修改, 但是并没有体现在正文中, 请注意修改。另外, 作者在修改中注意不再使用年级而使用年龄概念, 但是日历年龄仍然不是说明生理成熟的良好指标, 因为在比较统一的学制下年龄和年级高度相关, 因此日历年龄依然是生理成熟和环境教育的共同外部标示, 而绝非生理成熟本身。因此建议在文章的立意上不要过于强调生理成熟的问题。再有, 研究排除了阅读最差的学生, 理由是这些孩子可能不适合完成实验, 理由不足。请提供证据, 或者作为研究局限在讨论中说明。

回应: 根据专家提出的建设性修改建议, 做了如下三方面修改: 首先, 将全文中”语言熟悉性假说”全部修改为”语言熟练度假说”, 详见文中”粉色字体部分”; 其次, 在文章立意上弱化了”生理成熟”的说法, 修改主要体现在”问题提出”部分, 详见文中”黄色字体”部分; 最后, 针对选择阅读能力差这一组被试时, 从倒数第六名开始选择的理由相对不充分这一事实, 将其作为研究局限在讨论中进行了说明, 详见 p9 第四段黄色字体部分。

第三轮

主编终审意见：

作者根据前两轮审稿意见，已经非常有效地修改了文章中存在的问题。但仍有两个问题需要进一步说明和完善：(1) P4 页“2.4 实验仪器”部分第二行“刷新率为 1024×768 像素”。注意：“1024×768 像素”应该为“屏幕分辨率”，而非“刷新率”，请作者核实修改。(2) 在 P4 页“2.6 眼动分析指标”部分，如果作者能再简要补充一下各个分析指标的具体计算方式（或“操作定义”），将会更有利于读者理解该文章，以及与已有或将来同类研究进行比较。例如，本文在选择眼动指标时，引用了三篇文献（Feng et al., 2009; Jincho et al., 2014; Vorstius et al., 2014）。然而，对于“注视点持续时间分布”该指标，Feng et al. (2009) 在分析时，只分析了“向前注视点”（“Only forward fixations were plotted”），而没有分析“回视注视点”，因为“向前注视点”和“回视注视点”反映的认知加工并不相同。这个问题在本文“4 讨论 / 4.1 阅读水平对儿童阅读眼动注视模式发展的影响”部分的倒数第二段也有论述，而且“3 结果 / 3.3 平均向前眼跳次数和平均回视次数分析”部分也只发现了“向前眼跳次数”存在阅读水平差异。

回应：根据主编的建设性建议，做了如下两方面修改：第一，将“刷新率”修改为“屏幕分辨率”；第二，简要修改了每个分析指标的操作性定义。见 p4“2.6 眼动分析指标”部分。