

《心理学报》审稿意见与作者回应

题目：基于手机 APP 的双维 n-back 训练的认知与情绪效益

作者：潘东旒 王道湍 李雪冰

第一轮

审稿人 1 意见：

修改意见：论文《基于手机 APP 的双维 n-back 训练的认知与情绪效益》设计基于手机应用的新型情绪性双维 n-back 训练，并证明了具有一定的效果，研究具有一定的创新性，写作较为规范，但是基于该研究的研究报告还存在如下问题，需要作者作出进一步的修改和说明：

大的方面：

意见 1：

本研究主要解决的问题是什么？是利用手机 APP 进行认知功能干预的有效性探索，还是情绪性工作记忆训练的情绪效益？希望作者明确，如果是第一种，即利用手机 APP 进行认知功能干预的有效性探索，那研究重点应该在于探讨此类认知训练载体的信度、效度以及与其他认知训练载体（如生态训练，电脑单机版训练）相比较的优势，如果是探讨情绪性工作记忆训练的情绪效益，那么研究的重点应该是探讨在工作记忆训练中加入情绪因素的重要性以及分析其理论意义。

回应：

首先非常感谢审稿人中肯的建议和意见，我们经仔细思考，做了相应的修订。本研究的最主要解决的问题还在于情绪性工作记忆训练的认知与情绪效益。但是由于训练是基于手机应用而非传统的实验室训练，APP 干预的有效性在本研究中也需被并行的考察。我们根据您的建议对引言部分进行了较大范围的改写，增加了对情绪性工作记忆训练的研究现状，以明确工作记忆训练中加入情绪因素的意义与价值。具体见引言中标红的文字。此外，我们弱化了利用 APP 进行训练的部分，并将其编排在引言偏后的部分仅做必要的说明，以明确研究的主要目标。

当然，由于基于 APP 的文字-空间 n-back 训练的适用性也是本研究考察的目标之一，并且它的适用性是进行工作记忆训练情绪效益考察的前提，因此对它们讨论在精简文字后进行了保留。

意见 2：

基于第一个问题，如从研究的理论价值上看，应该要解决的核心问题是将情绪因素引入到认知训练中去，作者在引言的第六段阐述了将正性情绪工作记忆训练引入到研究中的四点理由，建议在之后重点介绍本研究的研究假设、研究设计和研究逻辑；

回应：

根据审稿人的建议，我们对引言部分进行了较大幅度的改写：回顾了工作记忆训练产生情绪效益与情绪性工作记忆训练的相关背景，阐发了当前负性工作记忆训练存在的问题，对引入正性训练（即直接对比中性、负性、正性训练）的必要性进行了说明，并在之后列出研究设计，以及对比正性、负性、中性的训练可能出现几种假设情形。当然，由于将正性训练引入研究尚属首次，本研究仅对正负情绪工作记忆训练的迁移做探索性的考察。

意见 3:

将情绪引入工作记忆训练是本研究的创新,但是仅仅从普通被试身上很难得出有意义和价值的研究结论,建议在文章的展望部分,介绍关于特殊被试(如情绪障碍被试)进行情绪工作记忆训练的可行性和意义。

回应:

根据您的建议,我们已经进行了相应的补充,详见讨论 4.4 红色字体部分

小的方面:

意见 4:

建议“训练应用编写”只介绍采用什么软件编制,无需详细介绍;

回应:

已参考您的建议,将方法中的“训练应用于编写”一节删除,而将其纳入方法的 2.3 “训练 APP”内,只对应用的基本功能与开发环境做了简单的介绍。具体见第 6 页,已用红色文字标出。

意见 5:

被试如何招募?两组被试是否匹配了基本的认知能力?

回应:

被试来自北京市若干所高校的大学生。通过个人/社团微信公众号与线上论坛进行招募。控制组与三组工作记忆训练组被试均通过电子抽签的形式随机分配(即安装完 APP 后,每个参与者均有一个训练口令,这个口令是随机给出的),该内容已经补充到方法的 2.1 中,用红色文字标明。

由于采用了随机分组,我们并没有对组间被试具体的认知能力进行匹配。此外,由于被试均为同水准的高校大学生,被试内的同质性相对较高。

尽管如此,我们仍旧对组别之间各个任务的前测成绩进行了统计检验,均没有发现组别之间的差异。在文章的“结果”部分,我们对训练前的空间工作记忆广度任务成绩进行了单因素方差分析,结果发现,4 组被试训练前的成绩没有显著差异, $F_{(3, 62)}=0.55, p=0.653$; 对训练前的空间工作活动记忆任务成绩进行了单因素方差分析,结果发现,4 组被试训练前的成绩没有显著差异, $F_{(3, 62)}=0.38, p=0.766$; 对训练前的转换任务成绩进行了单因素方差分析,结果发现,4 组被试训练前的成绩没有显著差异, $F_{(3, 62)}=1.51, p=0.233$; 对训练前的空间工作记忆广度任务成绩进行了单因素方差分析,结果发现,4 组被试训练前的成绩没有显著差异, $F_{(3, 62)}=1.18, p=0.325$ 。

这些结果表明,组间的被试训练前的行为水平上是类似的。由于工作记忆训练干预的施加,使得不同组后测行为任务的成绩产生了差异。

意见 6:

对于实验材料的评定,应详细介绍评定的过程和方式,以及评定被试的一些基本信息;

回应:

根据您的建议,该部分在 2.2.2 中进行了详细的呈现,均用红色字体标明。

意见 7:

在控制组任务介绍中,建议只需介绍采用了数字判断任务,无需在介绍为什么不采用 1-back 任务;

回应：

已根据建议进行了删改。

意见 8：

前、后测任务中，应明确介绍每项任务的出处；

回应：

前后测各个任务的出处已介绍。具体见 2.4 部分红色文字。

意见 9：

实验流程中，前、后测任务的测试顺序是如何安排的，是否有考虑？

回应：

前后测中，所有任务均是独立呈现的。即每一个任务都为 e-prime 文件，视空间工作记忆广度任务、活动记忆任务、数字转换任务、stroop 任务、情绪 stroop 任务分别命名为任务 1、任务 2、任务 3、任务 4、任务 5。被试要求按顺序进行任务，任务间可以稍作休息，以调整状态。所有被试完成任务的顺序是一致的，此外前测与后测的任务顺序是一致的。这部分内容已经补充到“实验流程”中。

这样操作的主要原因如下：

由于这些任务都与工作记忆能力紧密相关，任务之间顺序变化可能会影响单个任务完成的成绩。由于任务间成绩的差异并不是我们研究的主要问题，前后测任务成绩的变化才是我们关注的变量。因此，固定任务呈现的顺序成为控制不同任务之间相互影响带去前后测任务变化的一个操纵。即采用了一种恒定法的思路，在前后测中，由于任务的顺序是固定的，那么不同任务之间产生的影响也是一致的，并且可以通过，前测与后测成绩的减法消除。当然，这个操纵隐含了一个假设，就是训练并不会对前后测中不同任务的相互作用产生影响。我们认为这个假设基本还是可接受的。

此外，还有另一种平衡法的思路，即不同的被试完成不同任务的顺序不同，将这种任务间的相互影响平衡。但是由于本研究中前后测的任务较多，并且任务间的作用难以估计（且并不是我们主要的研究问题），不同被试任务间的顺序的完全平衡可能较难实现，可能还会影响主要问题的阐明，因此采用了顺序一致的编排。

尽管如此，对于任务顺序如何呈现我们也经过了一些考量。其基本的原则是趣味性强的，耗时较短的放在前面；需要大量重复试次的，耗时较长的放后面，以使得被试的注意状态在整体上保持最佳，提高数据的质量。

意见 10：

图一分辨率太低，建议修改；

回应：

已根据您的建议重新绘制。

意见 11：

结果部分训练效果的计算方法是否有依据；

回应：

训练效果量基本是后测的成绩-前测的成绩指标。不同任务的成绩由不同参数表征，均为衡量该任务完成水平主要的指标。空间工作记忆广度量为一个较为成熟的成绩指标；活动记忆任务采用正确率作为指标，前后测任务反应时的差异作为训练效益，也是较为普遍使用的。

意见 12:

活动记忆任务前测时被试的成绩已经很高，是否考虑存在天花板效应？

回应:

感谢您的仔细审阅。的确如此，实验中的活动记忆任务由于难度较低、试次较少而出现了天花板效应。由于天花板效应的存在，使得训练差异在理论上更难检测到，但我们还是发现在训练后，个体在该项任务上的成绩比原先已经很高的成绩更好了，即主要结论并没有受该天花板效应的影响。这可能是由于任务难度低，前测中高分被试后测的分数仍然很高；而前测中原本活动记忆能力与策略较差的个体，经过训练能够获得了较大的提升。尽管如此，我们仍需要在文中指明我们的这点局限，具体见结果 3.3 与讨论 4.1 红色字体部分。

.....

审稿人 2 意见:

工作记忆训练能够提升个体的工作记忆能力，但迄今尚未见有关利用手机 APP 进行工作记忆训练的相关研究。该研究利用自主研发的手机 APP，结合新型情绪性汉字-位置双维 n-back 任务对个体进行工作记忆训练，探究了其对个体的工作记忆的容量、以及刷新、转换、抑制功能的影响。文章整体逻辑思路较清晰，旨在将基础心理学中的工作记忆理论，通过手机 APP 应用到普通个体的生活中去，有一定的创新性和应用价值。假如该款手机 APP 能够做成游戏，也许能够让更多的人参与训练来达到提升工作记忆能力的效果。该文在实验设计的严密性、文字表达精炼性、文章结构的完整性等方面存在不少问题。具体分述如下：

引言部分

意见 1:

引言部分逻辑清晰，阐述得当，但是表达欠简洁，内容有些冗余。通篇来看，引言与讨论的有些内容重复。建议作者适度精简文字。

回应:

感谢您的仔细审阅与宝贵建议，综合您与另一位审稿专家的意见，我们已对引言中与讨论重复的部分进行了删改并明确了研究目标，具体见引言中标红的部分。

意见 2:

被试数量较少。采用组间设计，但是每组被试的数量不足 20 人（仅有 16 或 17 人）。建议作者依据前人研究来确定被试量，补充数据。

回应:

在以往涉及情绪性工作记忆训练的研究中，训练人数与当前的采用的被试数类似。例如 Schweizer & Dalglish (2011) 的研究中，控制组人数(n = 16); 中性 n-back 训练人数 (n = 14); 情绪性 n-back 训练人数(n = 15) ; Schweizer, Grahn, Hampshire, Mobbs, & Dalglish (2013) 的研究中，控制组(n=15) ; 情绪性工作记忆训练组 (n=17)。这些研究与我们的研究类似，均以健康大学生为被试。此外，当训练被试涉及临床个体时，例如对 PTSD 青少年进行情绪性工作记忆训练，其被试人数可能更少，情绪性工作记忆训练组 (n=15)，控制组训练组(n = 15) (Schweizer, Samimi, Hasani, Moradi , Mirdoraghi, & Khaleghi,2017)。这些研究有涉及脑影像与脑可塑性的研究，更多的则与我们的目标相似是测查行为任务改变的研究。

另外，我们根据这些研究中前后测行为任务的成绩变化计算效应值，并利用 G*Power 计算最少所需要的被试量，其计算结果亦小于本研究中所采纳的被试（计算得出的最少被试量仅需要每组 4-10 个）。

鉴于本研究开展前被试量选择即根据前人研究综合确定,并且结果显示训练在各认知任务中的迁移效应较为稳定一致,各训练组与控制组的效果差异也被稳定的观测到,因此就没有再进一步增加被试量。

- [1] Schweizer, S., & Dalgleish, T. (2011). Emotional working memory capacity in posttraumatic stress disorder (PTSD). *Behaviour Research & Therapy*, 49, 498–504.
- [2] Schweizer, S., Grahn, J., Hampshire, A., Mobbs, D., & Dalgleish, T. (2013). Training the emotional brain: improving affective control through emotional working memory training. *Journal of Neuroscience*, 33, 5301–5311.
- [3] Schweizer, S., Hampshire, A., & Dalgleish, T. (2011). Extending brain-training to the affective domain: increasing cognitive and affective executive control through emotional working memory training. *PLoS One*, 6, e24372.

意见 3:

文中未提及前后测任务,如空间工作记忆广度任务(广度)、活动记忆任务(刷新)、数字转换任务(转换)等的顺序是否进行了顺序平衡?假如没有,请给出理由。另外,在前后测时,被试须要一次性做多种任务,且均与工作记忆有关。请问如何排除不同任务之间的影响以及练习效应?

回应:

谢谢审稿人的仔细审阅。在前后测中,所有任务均是独立呈现的。即每一个任务都为 e-prime 文件,视空间工作记忆广度任务、活动记忆任务、数字转换任务、stroop 任务、情绪 stroop 任务分别命名为任务 1、任务 2、任务 3、任务 4、任务 5。被试要求按顺序进行任务,任务间可以稍作休息,以调整状态。所有被试完成任务的顺序是一致的,此外前测与后测的任务顺序是一致的。这部分内容已经补充到“实验流程”中。

这样操作的主要原因如下:

由于这些任务都与工作记忆能力紧密相关,任务之间顺序变化可能会影响单个任务完成的成绩。由于任务间成绩的差异并不是我们研究的主要问题,前后测任务成绩的变化才是我们关注的变量。因此,固定任务呈现的顺序成为控制不同任务之间相互影响带去前后测任务变化的一个操纵。即采用了一种恒定法的思路,在前后测中,由于任务的顺序是固定的,那么不同任务之间产生的影响也是一致的,并且可以通过,前测与后测成绩的减法消除。当然,这个操纵隐含了一个假设,就是训练并不会对前后测中不同任务的相互作用产生影响。我们认为这个假设基本还是可接受。

此外,还有另一种平衡法的思路,即不同的被试完成不同任务的顺序不同,将这种任务间的相互影响平衡。但是由于本研究中前后测的任务较多,并且任务间的作用难以估计(且并不是我们主要的研究问题),不同被试任务间的顺序的完全平衡可能难以实现,可能还会影响主要问题的阐发。

尽管如此,对于任务顺序如何呈现我们也经过了一些考量。其基本的原则是趣味性强的,耗时较短的放在前面;需要大量重复试次的,耗时较长的放后面,以使得被试的注意状态在整体上保持最佳,提高数据的质量。

意见 4:

即时的情绪状态以及持久的心境都会对影响实验结果。该研究使用 POMS 作为个体的情绪控制量表,但是仅仅在前后测中使用。在每天训练的过程中,是否使用了其他情绪量表测量个体的即时情绪状态,来排除情绪对训练的影响?假如没有,请给出理由。

回应：

感谢您的问题。即时的情绪状态与持久的心境的确会给实验结果造成影响。例如持续低迷的心境可能难以提升训练的动机。

然而我们并没有测量被试在训练过程中的情绪状态。这主要是考虑到，由于训练是在移动端中进行的，被试可以自主选择每日训练的时间与时长（这里的时长是指分散的完成训练还是集中式的完成训练），训练即时的情绪状态难以获取。但从另一方面而言，也正是由于被试可以自主选择训练的时间，似乎可以使训练时个体的情绪变异得到一些控制：从实验期间与实验后的反馈来看，大家都几乎都愿意选择在一天之中比较悠闲放松的时刻，打开 APP 进行训练。

虽然就整体而言，即时状态对训练成绩的影响并不是我们重点研究的问题，但由于没能获得有效的数据，训练即时的情绪状态（每日情绪的动态变化）与训练水平之间的关系无法在本研究中被精确的考察，对于这一点局限，我们在讨论中进行了说明与展望，具体见 4.3。

意见 5：

为什么只训练了 5 天？图 1 的内容表明，随着训练时间的增加，三种情绪任务的训练效果一直在增加。商榷：5 天训练时间是否适宜？或许考虑增加训练时间，会有新发现？

回应：

感谢您的建议。本研究的主要目标之一是考察新型双维 n-back 训练的适用性。压缩训练时长出于验证新型任务有效性的考量。短期（4-7 天）的认知训练在以往研究中已有先例，无论是情绪注意偏向训练(赵鑫, 张鹏, 陈玲, 2014)，或者是其他类型考察情绪效益的工作记忆训练(Wanmaker, Geraerts, & Franken, 2015)。

当然，正如您所提到，结果显示在 5 天内，训练组的训练效果一直在增加，而没有到达平台期，如果继续进行追踪考察可能会有一些新的发现，但考虑到实验结束后被试的流失，增加训练时间需要重新征集被试完成实验，在实践上存在一些困难。对于这点局限，我们已在讨论 4.3 中进行了说明。

不过尽管如此，我们的结果显示训练在各认知任务中的迁移效应较为一致，各训练组与控制组的效果差异也被稳定的观测到，短期的新型工作记忆训练仍旧能产生一定的训练效益，说明该基于 APP 的新型任务的有效性，较为理想的实现了本研究的目标。根据您的建议，下一步我们或在该任务的基础上，进行较长时间的训练与追踪。

[1] Wanmaker, S., Geraerts, E., & Franken, I. H. A. (2015). A working memory training to decrease rumination in depressed and anxious Individuals: A Double-Blind Randomized Controlled Trial. *Journal of Affective Disorders*, 175C, 310–319.

[2] 赵鑫, 张鹏, & 陈玲. (2014). 注意偏向训练对社交焦虑的干预:方法、效果与机制. *心理科学进展*, 22(8), 1246-1257.

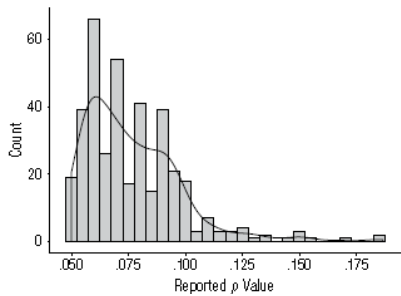
意见 6：

多处细节表述不当、不符合统计学理论。比如“3.2 空间工作记忆广度任务”中的“控制组与负性训练组差异边缘显著($p=0.162$)，与正性训练组训练差异边缘显著 ($p=0.180$)”。何为边缘显著？边缘显著的话， p 值一般要低于 0.1。

回应：

感谢您的细心审阅。边缘显著的表达确实不够严谨。尽管有研究对心理学研究中关于边缘显著的报告与表述进行了统计，发现在高级别的心理学期刊中，边缘显著的表达越来越普遍，而其范围甚至扩展到了 0.18，如图横坐标为报告为边缘显著下实际的 p 值，纵坐标为文

章数^[1]。但是，将 p 值高于 1 报告为边缘显著并不应该是我们效仿的对象。



那么当前研究中，我之所以使用了“边缘显著”的说法，是由于尽管对工作记忆广度任务前后测增量进行单因素方差分析时的 p 值大于 0.1，却不能直接拒绝训练效益的存在，因为通过对不同组别进行前后测匹配样本 t 检验，控制组前后测没能达到显著性水平，但三组训练组前后测成绩均有显著差异。更重要的是，通过分别计算控制组与训练组训练前后的 cohen'd，在空间工作记忆广度任务中，我们发现控制组的 d 值为 0.27；负性训练组的 d 值为 0.79；正性训练组的 d 值为 0.55，均已经达到了中-大的效应值^[2]。

为了严谨的表述，根据您的建议，并且考虑到与其他任务的统计方法相统一。我们删除将训练效果量（后测-前测的空间工作记忆广度的增加值）的单因素方差分析。而是分别对不同组别的前后测进行匹配样本 t 检验。结果如下，控制组的训练前后活动记忆任务成绩差异不显著， $t_{(16)}=-1.07, p=0.3$ ；负性训练组训练后的空间工作记忆广度成绩显著高于训练前成绩， $t_{(16)}=-2.54, p=0.022$ ；中性训练组训练后的活动记忆任务成绩高于训练前成绩 $t_{(15)}=-6.47, p<0.001$ ；正性训练组训练后的活动记忆任务成绩显著高于训练前成绩 $t_{(15)}=-4.16, p=0.001$ 。

[1] Pritschet, L., Powell, D., & Horne, Z. (2016). Marginally significant effects as evidence for hypotheses: changing attitudes over four decades. *Psychological Science*, 27(7), 1036.

[2] Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods Jmasm*, 8(2), 597-599.

意见 7:

书写错误。如“3.6 情绪工作记忆训练对情绪偏向的效果”中的“中性训练组 $t_{(15)}=1.088, p=0.294$ ，正性训练组 $t_{(15)}=-1.326, p=0.205$ 训练后的负性 Stroop 效应均无差异”。p 值不应大于 1!

回应:

感谢您的细心审阅，此处的 p 分别为 0.294 与 0.205。我们已再次仔细阅读全文，将书写错误进行了改正。

意见 8:

“参见 3.7 情绪工作记忆的一般情绪效果”。对 POMS 前后测量表，使用一般情绪效益作为因变量指标，结果显示“多重比较表明，正性训练组的紧张减少显著高于控制组； $p=0.029$ ；负性训练组的慌乱情绪增加显著高于中性训练组， $p=0.04$ ”。问题：如何排除被试的紧张和慌乱情绪对总体实验结果的影响？

回应:

感谢您的仔细审阅与问题。首先需要明确的是 POMS 测查的并非即时的情绪状态，而是在本研究中为训练期间总体心境水平。因此，POMS 得出的紧张与慌乱分数并不能说明其对前后测各任务水平的影响。但被试在进行自陈量表填写的过程中，还是有很大可能依据当前的情绪状态做出勾选，这也是您所关切的方面。

在本研究中，我们假设工作记忆训练对个体情绪状态与认知任务作业水平的影响是独立

的，这也遵循前人研究的一般性思路。但是正如您所提出的，情绪与认知之间存在复杂的交互，这在我们的研究中可能体现在前后测任务中，情绪状态本身对认知任务完成的影响；也可能体现在工作记忆训练对情绪状态与认知能力的改变并不独立。这是我们之前并未仔细考量的。

根据您的提问，我们首先将前后测情绪分量的变化与各个任务的完成度变化进行了皮尔逊相关。发现的唯一有价值的相关体现在 Δ 慌乱（即前后测增加的慌乱）与 Δ 空间工作记忆广度（空间工作记忆广度增益）成绩成一个弱的相关，而未在其他任务作业与情绪变化上发现直接的关联。

	1	2	3	4	5	6	7	8	9	10	11	12
Δ 刷新	1											
Δ 转换	0.096	1										
Δ 负性反应时	-0.086	0.013	1									
Δ 正性反应时	-0.174	-0.033	.358**	1								
Δ 空间广度	0.151	0.028	-0.219	-0.044	1							
Δ 紧张	-0.193	-0.072	0.026	-0.098	-0.18	1						
Δ 愤怒	-0.163	-0.059	-0.002	0.014	0.025	.750**	1					
Δ 疲劳	0.058	0.103	-0.211	-0.244	-0.114	.600**	.534**	1				
Δ 抑郁	-0.192	-0.052	0.038	0.019	-0.126	.821**	.794**	.621**	1			
Δ 精力	0.115	-0.06	-0.1	0.013	-0.074	-.392**	-.443**	-.266*	-.461**	1		
Δ 慌乱	-0.134	-0.204	-0.089	-0.078	-.273*	.742**	.406**	.512**	.618**	-0.21	1	
Δ 自尊	0.067	0.074	-0.044	-0.02	0.002	-.521**	-.508**	-.416**	-.612**	.713**	-.466**	1

接下来，为考察个体空间工作记忆广度是否独立的获得了提升（而非由于情绪变量的影响而产生了变化），我们将 Δ 慌乱分数作为协变量，进行协方差分析。

以下为进行协方差分析的中间过程【组别*增加的慌乱交互作用不显著， $p=0.403$ ，不能认为各组协变量与观测指标之间的回归直线斜率不等，满足回归直线平行的条件（这是协方差分析的假设检验之一）；四组方差齐性检验结果显示， $F=1.01$ ， $p=0.392$ ，不能认为两组方差相等，（这是协方差分析的假设检验二）； Δ 慌乱对 Δ 空间工作记忆广度有影响， $F=3.33$ ， $p=0.073$ ；组别之间也有独立的差异， $F=2.25$ ， $p=0.91$ ，边缘显著。】

对 Δ 慌乱进行调整后，控制组 Δ 空间工作记忆广度为 0.329，（95%CI: -0.163—0.822），负性训练组控制组 Δ 空间工作记忆广度为 0.871，（95%CI: 0.373—1.369），中性训练组 Δ 空间工作记忆广度为 1.249，（95%CI: 0.733—1.766），正性训练组 Δ 空间工作记忆广度为 0.759，（95%CI: 0.250—1.263）。

这些分析表明，增加的慌乱水平确实影响了个体空间工作记忆广度成绩的提升，但在对 Δ 慌乱进行调整之后，组间的独立效应仍然存在，因此，我们的主要结论仍然成立。

此外，更需要注意的是，在以上分析中，我们进行协方差分析的各个变量均为前后测的增益 Δ 。当我们单纯使用后测的慌乱成绩进行分析时，慌乱分数并没有影响个体的空间工作记忆广度增益， $F=0.056$ ， $p=0.813$ ；或者我们分析后测中慌乱分数对空间工作记忆分数（而非前后测增益）的影响，也没有发现显著的效应， $F=2.50$ ， $p=0.118$ 。这也就是说，心境量表下慌乱分数本身对作业成绩上的影响并不凸显，但是工作记忆训练过程既影响了个体的情绪状态，又影响了个体的空间工作记忆成绩，又存在对两者的交互影响。

事实上，当我们考察一项干预如何作用于认知与情感状态两个方面时，需要一个假设对训练是否独立的影响情绪状态与认知能力进行规定。十分感谢您的宝贵的建议，我们会在后续研究中，改进相关的实验设计，对训练如何既影响情绪状态又影响认知能力进行更为精细

的考察。而在本研究中，我们也就该局限性在讨论 4.3 中进行了指明。

此外，尽管在统计上未显著，在后测中，慌乱分数对空间工作记忆分数还是存在一些影响 $p=0.118$ 。这可能是因为空间工作记忆广度任务是完成情绪问卷测量后，最开始完成的任务，与个体开始时完成任务的情绪状态存在一些关联。但随着任务的进行，个体在完成任务过程中，受原本情绪状态的影响降低甚至消失。当然这仅仅是出于推测，由于这并非本研究关注的问题，且未达到显著性水平，我们仅将此点在回应中进行说明，而不再文本中进行呈现。

意见 9:

讨论部分逻辑表述较为清晰，但是叙述赘余。建议精炼表达，提高可读性。

回应:

根据您的建议，我们已经对讨论部分进行了删改以提高可读性。

意见 10:

结论部分为何缺少结论？

回应:

根据您的建议，我们已经补充了结论部分，请见 5.结论。

第二轮

审稿人 1 意见:

论文有了较大的改进，研究目的近一步明确，研究结果近一步清晰，还有如下问题，需

意见 1:

摘要部分：“结果表明，短期双维 n-back 训练能够使个体的视空间工作记忆任务、活动记忆任务、数字转换任务、经典 Stroop 任务产生相较控制组更大的进步。这些任务分别测验了工作记忆的容量、以及刷新、转换、抑制功能，较为完整的表征了工作记忆的全貌。”建议修改为：“结果表明：短期双维 n-back 训练能够使个体在视空间工作记忆任务、活动记忆任务、数字转换任务、经典 Stroop 任务上的成绩产生相较控制组更大的进步，表明训练可以提高个体工作记忆的容量和中央执行功能。”

回应:

已根据您的建议进行了已修改。此外，针对另一名审稿专家提出的结论与摘要部分重复的问题，也对中文摘要在内容与措辞上进行了小修，已用绿色文字标明。

意见 2:

问题的提出部分 文中提出：“具体而言，正性、负性、中性的训练在情绪注意任务的迁移上可能出现几种情形：（1）训练材料的效价一致地提升了对应效价的注意偏向，即负性训练增加负性注意偏向，正性训练增加正性注意偏向。这种结果提示训练将提升对训练材料的直接注意，导致情绪一致性效应。提示负性训练的使用须更谨慎的考察，而正性训练或更有改善注意偏向的潜能；（2）只有负性材料的训练减少了个体的负性注意偏向，而正性材料的训练没能影响（无论是增加或减少）个体的正性注意偏向。这种结果提示负性训练的确可以矫正个体负性偏向，提升自上而下的控制能力，增益工作记忆训练的情绪效用；（3）负性材料训练没能影响（无论是增加或减少）个体的负性偏向，正性材料对应影响了个体的正性注

意偏差。这种结果提示正性工作记忆训练或具有一定情绪注意偏向干预的效果；(4)无论是正性还是负性训练都没能对情绪偏向造成对应的影响,这种结果提示情绪注意偏向对训练中的材料性质并不敏感,工作记忆训练的情绪效益可能与特定的情绪材料关系较弱。”这意味着研究并没有提出明确的研究假设?作者是如何考虑?

回应:

审稿人的这个问题,我们也仔细考虑过,由于情绪工作记忆训练是一个较新的课题,相关研究较少,以往利用负性材料进行训练的研究,考察其迁移效果也局限在负性材料的认知任务中,这些研究结果无法推演到正性材料的训练上。由于目前并没有直接对比正性与负性材料的工作记忆训练效果迁移的研究;而对负性材料工作记忆训练是否存在特异效果(或者说工作记忆是否需要负性材料才能产生情绪效益)也存在诸多的不一致,因此我们对不同效价材料的工作记忆训练在情绪偏向任务中的表现抱有开放的态度,进而提出了上述的四种假设,而我们当前的研究就是验证这四种竞争性假设。

本研究中,我们发现各类情绪材料的工作记忆训练无论在执行功能任务还是情绪偏向任务上的表现都是类似的,反应了情绪性文字在工作记忆训练中产生的影响不大(至少在健康大学生中)。但这仅仅是一个探索性的结果,要明确回答伴随情绪的工作记忆训练是否更能提升个体的情绪认知控制功能,还需要改善研究设计,将训练过程中的情绪唤起进一步提升,此外对比多个情绪性任务成绩,才能明确情绪材料在工作记忆训练中的价值与意义。

意见 3:

建议先介绍任务再介绍材料;

回应:

为进一步的简化与清晰研究方法的逻辑,参考您的建议,我们将任务搭载的应用部分作为 2.2 训练任务的头段进行说明,将介绍训练任务本身与任务中情绪材料筛选的部分合并归于 2.2.1。

意见 4

3.2 前、后侧任务中无需在任务后加括号说明该任务是测量什么能力。

回应:

已根据您的建议删除。

意见 5:

研究结果部分 4.1 图一中建议横轴中的图标统一为中文 4.2 表六的表头格式有问题

回应:

已根据您的建议修改。具体见绿色文字。

意见 6:

讨论部分 建议将以下内容统一整合至研究的局限与展望中:“然而值得关注的是,本研究中采用数字活动记忆任务作为测查刷新能力的指标,前测中任务的完成水平已经较高,天花板效应可能阻碍了训练效益的进一步显现。未来可使用其他刷新任务来进一步阐明训练对刷新能力的独立影响。”;“在我们的研究中,一个隐含的假设是训练对情绪状态与认知功能的影响是独立的,但是由于情绪与认知作业存在相互作用,训练增益对认知与情绪功能的改善也可能存在交互。未来研究可以采用更为精细的设计,追踪记录个体训练期间的情绪改变与认知作业水平,从而完整的考察工作记忆训练对个体情绪与认知功能的综合影响。”;“当然,值得关注的是,我们的实验人群为健康大学生被试,这局限了结果的可推广性。个体差

异会很大程度的影响训练效果的迁移(Jaeggi, Buschkuhl, Shah, & Jonides, 2014)。事实上, 健康人群的认知提升空间有限, 训练进步可能更难实现、更少维持; 而在具有情绪性认知缺损的个体中, 训练可能表现出某种补偿效应, 使获益更为显著(von Bastian & Oberauer, 2014)。因此, 情绪障碍人群可能才是情绪性工作记忆训练的真正受益者。例如抑郁患者被证明存在情绪工作记忆的缺损, 其对情绪材料的抑制与更新存在困难, 使之成为抑郁发展与维持的重要认知机制(Joormann, Levens, & Gotlib, 2011)。未来的研究还需要进一步拓展到各类情绪障碍人群中, 考察情绪性工作记忆训练对临床个体的效用。”

回应:

已根据您的建议进行了修改, 详见绿色文字。

意见 7:

参考文献格式存在问题: Ansari, S. (2015). The therapeutic potential of working memory training for treating mental disorders. *Front Hum Neurosci*, 9, 481. Gao, X., Zhou, R. L., & Dong, Y. Y. (2013) A Research of the Selective Effects of Positive Emotion on Verbal and Spatial Working Memory. *Chinese Journal of Clinical Psychology*, 2, 175–179. Jaeggi, S. M., Studer-Luethi, B., Buschkuhl, M., Su, Y. F., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning — implications for training and transfer ☆. *Intelligence*, 38, 625–635. Kim, N. Y., Wittenberg, E., & Nam, C. S. (2017). Behavioral and Neural Correlates of Executive Function: Interplay between Inhibition and Updating Processes. *Front Neurosci*, 11, 378. Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *J Clin Exp Neuropsychol*, 24, 781–791. Liu, X. P., Wang, Z. X., Chen, X. C., & Zhang, D. R. (2003). subsystems in visual workingmemory. *Acta Psychologica Sinica*, 35, 598–603. [齐冰, 白学军, & 沈德立. (2007). 材料类型和任务可预测性在任务转换中的作用. *心理科学*, 30, 41-44.] Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., . . . Engle, R. W. (2013). No Evidence of Intelligence Improvement After Working Memory Training: A Randomized, Placebo-Controlled Study. *Journal of Experimental Psychology-General*, 142, 359–379. Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychol Bull*, 138, 628–654. Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology General*, 121, 15-23. Szmalec, A., Verbruggen, F., Vandierendonck, A., & Kemps, E. (2011). Control of interference during working memory updating. *J Exp Psychol Hum Percept Perform*, 37, 137–151. von Bastian, C. C., & Eschen, A. (2016). Does working memory training have to be adaptive? *Psychol Res*, 80, 181–194. [王一牛, ZHOU Li-Ming, & 罗跃嘉. (2008). 汉语情感词系统的初步编制及评定. *中国心理卫生杂志*, 22, 608-612.] [祝蓓里. (1995). Poms 量表及简式中国常模简介. *天津体育学院学报*, 10, 35-37] [详细意见]

回应:

感谢您的耐心审阅, 已对参考文献检查修改。

审稿人 2 意见:

作者对“审稿专家 1”、“审稿专家 2”的意见进行了较认真的、合理的修改、回应, 总的来说, 修改后的论文在实验设计的严密性、论证的逻辑性等方面有了一定提升, 但在文字表达的正确性、精炼性方面, 以及研究方法呈现形式的清晰性方面, 摘要和结论的概括性、逻辑性方面仍存在不少错误和不足。具体分述如下:

意见 1:

全文存在不少语病和错漏，详见文中用绿色标出的文字，如“更好的验证”，“研究发现”，等等。本人认为，凡是用绿颜色标注的文字部分，均有疑问请作者斟酌。

回应:

感谢您的仔细审阅，我们对这些表述进行了修改，详见绿色文字。

意见 2:

虽然补充了“结论部分”，但该结论似乎不够有力、具体、清晰，拟进一步提炼。

回应:

我们对结论部分进行了修改，具体见绿色文字。

意见 3:

摘要部分的问题类似于结论部分。

回应:

我们对中文摘要部分进行了改写，具体见绿色文字

意见 4:

“2 方法”中，缺少实验流程图，是否应该考虑增加若干实验流程图？

回应:

根据您的建议，我们补充了实验流程图，详见图 1

意见 5:

“2.1 被试”中对被试的描述是：共 66 名来自北京市的健康大学生被试（男 34 人，女 32 人）参加实验，其年龄分布 18~29 周岁之间。问题：健康大学生被试中有年龄为 29 岁的吗？妥否？这或许涉及更多的问题。

回应:

我们的被试年龄主要集中在 18-24 岁，经原始数据查询，仅一名被试年龄为 29 岁，该被试为 11 号被试，姓名为：景 X，所属组别为负性工作记忆训练组，男性。

我们的被试均招募自北京某 211 学校的在校大学生，且均在问卷中“无精神疾病”题项中进行了勾选。年龄性别等个人信息是被试通过扫描二维码线上填写的。查询线上问卷原始答案，发现该被试填写确实为 29，数据录入未出现失误。由于已经无法再次联系到该名被试（所填手机号已停机，微信已被对方删除），我们暂无法确认该名被试的真实年龄（可能是被试的填写失误，但也可能确实是其真实年龄）。

对该名被试的数据进行检查，各项任务成绩均非极端值（前后测各个指标均在 3 个标准差以内），因此，在本研究中，我们仍旧对该名被试的数据进行了保留。

十分感谢您的仔细审阅，我们在日后的研究与实验中，应更加强调主试对实验的观察与控制。

第三轮

审稿人意见:

“xb17-386”的审稿（三审）意见：修后发表。注：凡是用浅蓝颜色标注的文字，为有疑问之处，请作者斟酌。

意见 1:

“审稿专家一”第二个问题是：这意味着研究并没有提出明确的研究假设？作者是如何考虑？

作者对“审稿专家一”第二个问题的回复是：

验证一种新型工作记忆训练的适用性与一般性迁移，对其在这些方面进行考察是必要的。此外，通过对比情绪注意偏向任务在训练前后的表现差异，我们可以对工作记忆训练中情绪成分的效果与作用进行澄清。具体而言，正性、负性、中性的训练在情绪注意任务的迁移上可能出现几种情形：（1）训练材料的效价一致地提升了对应效价的注意偏向，即负性训练增加负性注意偏向，正性训练增加正性注意偏向。这种结果提示训练将提升对训练材料的直接注意，导致情绪一致性效应；（2）只有负性材料的训练减少了个体的负性注意偏向，而正性材料的训练没能影响（无论是增加或减少）个体的正性注意偏向。这种结果提示负性训练的确可以矫正个体负性偏向，提升自上而下的控制能力，增益工作记忆训练的情绪效用；（3）负性材料训练没能影响（无论是增加或减少）个体的负性偏向，正性材料对应影响了个体的正性注意偏差。这种结果提示正性工作记忆训练或具有一定情绪注意偏向干预的效果；（4）无论是正性还是负性训练都没能对情绪偏向造成对应的影响，这种结果提示情绪注意偏向对训练中的材料性质并不敏感，工作记忆训练的情绪效益可能与特定的情绪材料关系较弱。总的来说，由于尚无先例直接对比正性与负性工作记忆训练在情绪注意偏向上的效果，我们本研究将对上述四种竞争性假设进行验证。

作者认为该研究有四种竞争性假设，易使人误以为该研究有四个假设。故本人认为，若将这四种竞争性假设凝练为一种假设是否更加切题？供商榷。

回应：

感谢您的建议。根据您的提议，且结合本研究重点呈现结果，我们对假设再次进行了凝练，具体如下：

总体而言，正性、负性、中性的训练在各类任务的迁移上可能出现两种情形：（1）训练材料的效价一致地影响了各类基础任务的迁移以及情绪注意控制任务的迁移；（2）训练材料效价的影响在基础认知任务以及情绪注意控制任务上出现了分离：正性或者负性效价的训练特异地影响了情绪注意偏向。由于尚无先例直接对比正性与负性工作记忆训练在情绪注意偏向上的效果，本研究将对上述两种竞争性假设进行验证，以评估情绪材料在工作记忆训练中的价值。

见文中对应位置浅蓝色文字。

意见 2：

全文拟存在一些错漏，详见文中用浅蓝色的文字，如引言部分的“可能使训练具有一定动机成分。”请作者斟酌。

回应：

已经更正，见修改稿浅蓝色文字；

意见 3：

文中表 1、表 2、表 3、表 4、表 5 和表 6 中的测试指标均未标注出单位。

回应：

已经更正，见修改稿浅蓝色文字。