

《心理学报》审稿意见与作者回应

题目：脑电信号的重测信度：在多项静息态和任务态实验中的对比

作者：覃慧怡，丁丽洪，段威，雷旭

第一轮

审稿人 1 意见：

本文从不同角度评估了 EEG 测量的重测信度。

意见 1：但需要从个体差异的角度就其结果的生物学意义进行解释和讨论，请补充，具体参考资料请参考如下的个体差异测量理论框架：
<https://www.nature.com/articles/s41562-019-0655-x>

回应：非常感谢专家提出的宝贵意见，专家给出的这篇参考文献也强调了在认知神经科学领域重测的重要性以及个体间的差异导致的信度和效度之间的联系与区别，我们认真研读后也在本文合适的位置进行了引用。

虽然反映个体差异的效度难以量化，但是信度反映了效度的上限，从实践的角度考虑，我们能做的就是提高感兴趣的特征 (V_t)、减小随机噪声 (V_r)，这也是本研究的目的。我们研究的对象是电生理测量中的脑电信号，相比于 fMRI 有更高的时间分辨率。静息态脑电能提供个体有关认知和人格特征差异的相关信息(Deco, Jirsa, & McIntosh, 2011)，任务态脑电用于研究个体的认知功能的特点及动态过程 (Lugo et al., 2020)。正如 Zuo 等人(2019)所言，不同的实验范式提供了不同程度的效度。为了寻找更合适的重测实验范式和指标，我们选择了两种静息态范式和学习、练习效应较小且在脑电及其重测研究领域十分常用精神运动警觉任务 (psychomotor vigilance task, PVT) 和 Oddball 任务态范式，在测量个体的人格特征的同时也方便在不同研究间进行对比。我们的研究发现，执行任务过程中脑电的重测信度比反应被试稳定精神状态的静息态和 PVT 范式更低。同时，考虑到样本量与信度间密切的双向关系，在保证质量和研究可行性的前提下，我们最终保留了 42 名被试的数据，经过统计功效检验符合最小样本量要求。

根据您的意见，我们在修改稿的引言和讨论中分别补充了下述关于重测信度的内容。

“.....理论计算表明，信度决定了最大可检测到的效度大小，而在心理研究领域中，信度、样本量和效应量之间的相互作用经常被低估(Zuo et al., 2019)。.....”（第 2 页“1 引言”第 3 段蓝色部分）

“.....同时，可靠性也取决于预处理步骤和脑功能指标选择，通过横向比较，有利于找到最适合探讨个体差异的指标(Zuo et al., 2019)。.....”（第 3 页“1 引言”第 5 段蓝色部分）

“.....从实践的角度来看，研究人员所能做的就是尽最大努力测量和消除非感兴趣信号（如运动伪迹等）来提高信噪比，并使用更优的方法提高信效度(Zuo et al., 2019)。”（第 14 页“4.4 多种任务态和静息态的对比”第 2 段蓝色部分）

新增参考文献：

Zuo, X. N., Xu, T., & Milham, M. P. (2019). Harnessing reliability for neuroscience research. *Nat Hum Behav*, 3(8), 768-771. doi:10.1038/s41562-019-0655-x

审稿人 2 意见:

该论文采用多种任务态和静息态实验, 计算并对比了五种 EEG 信号在时域、频域、空间位置的重测信度。该论文是脑电研究中少有的专注于不同类型脑电信号重测信度的文章, 对脑电相关研究的开展和应用具有重要的参考价值。

意见 1: 预处理参数的选择对重测信度结果是否有重要的影响? 比如说滤波参数的选择和重参考的选择对结果是否有显著的影响? 另外, 作者没有说明重参考到什么电极了?

回应: 感谢专家提出的问题。包括滤波、独立成分分析 (independent component analysis, ICA)、重参考以及人工去除伪迹等在内的常用脑电预处理步骤都会影响信号质量, 并进一步影响重测结果。例如, 如果存在异常通道, 进行平均参考可能会污染整个数据集 (Bigdely-Shamlo, Mullen, Kothe, Su, & Robbins, 2015); ICA 对伪迹的抑制可能也会导致大脑信号部分丢失 (Castellanos & Makarov, 2006)。也有研究评估了不同的预处理方法对脑电信号的可靠性的影响, Suarez-Revelo 等人对照了包括 EEGLAB (Delorme & Makeig, 2004) 和标准化的预处理 (preprocessing, PREP) 在内的 4 种不同的预处理方法 (Bigdely-Shamlo, Mullen, Kothe, Su, & Robbins, 2015), 发现在不同的频段和预处理方法中, 重测信度差异很大。重参考和 ICA 可以提高某些频段 (如 β 和 γ) 的重测信度, 方差分析显示预处理方法与提高所有频段的重测信度无关 (Suarez-Revelo, Ochoa-Gomez, & Duque-Grajales, 2016)。近期的一项研究使用了多种预处理方法, 证明了完整的预处理步骤可以提高重测信度, 尤其可以显著减少睁眼静息态 EEG 中由于眼电、肌电等造成的伪迹 (Duan et al., 2021)。关于主试手工去除伪迹的可靠性以及预处理中的差异, 是如何影响 ERP 的重测信度的研究发现: 主试手工去除 EEG 伪迹的差异并没有改变主要的 ERP 结果 (Shirk et al., 2017)。

在滤波参数范围的选择上, 绝大多数研究会选择在 1-50Hz 的范围内 (本研究选择的 0.3-45Hz) 进行滤波, 以囊括脑电常用的 6 个频段: delta (1~4 Hz)、theta (4~8 Hz)、alpha1 (8~10 Hz)、alpha2 (10~13 Hz)、beta (13~30 Hz)、gamma (30~50 Hz)。一般来说, 预处理对于提高脑电信号质量从而提高脑电功率的重测信度很重要, 所以需要采用统一的预处理流程 (Cohen, 2017)。

专家在此提到的预处理对重测信度的影响, 对本文结论的推广有参考意义。本研究采用的是已有的传统预处理方法, 目的是保持不同模态数据处理间的一致性, 增加可对比性。但是也如本文在“4.5 局限性”中提到的“选择了传统的数据处理方法, 这虽然使得结果在更多的研究中具有可比性, 但对于许多新方法 (如功能连接, 脑网络等) 的参考意义有限。”根据专家的建议, 我们在修改稿中补充了下述关于本研究预处理的局限性的内容:

“第一, 选择了传统的数据处理方法, 这虽然使得结果在更多的研究中具有可比性, 但考虑到目前对于预处理方法和参数影响脑电信号重测信度的结论尚不一致 (Duan et al., 2021; Suarez-Revelo et al., 2016)。本研究的结论对于区别较大的预处理策略和许多新方法 (如功能连接, 脑网络等) 的参考意义有限, 未来的研究也应更多着眼于新策略新方法, 推动脑电数据处理和分析流程的标准化 (Cohen, 2017)。” (第 14 页“4.5 局限性”第 1 段蓝色的部分)

关于重参考, 非常抱歉在之前版本中没有写出。我们使用的重参考方法是全脑平均参考, 因此, 在“2.3 rsEEG 预处理与频谱分析”中, 我们增加了一句: “第三步, 对数据进行全脑平均重参考, 并……”。

新增参考文献:

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K. M., & Robbins, K. A. (2015). The PREP pipeline: standardized preprocessing for large-scale EEG analysis. *Front Neuroinform*, 9, 16. doi:10.3389/fninf.2015.00016

Castellanos, N. P., & Makarov, V. A. (2006). Recovering EEG brain signals: artifact suppression with wavelet enhanced independent component analysis. *J Neurosci Methods*, 158(2), 300-312.
doi:10.1016/j.jneumeth.2006.05.033

Cohen, M. X. (2017). Rigor and replication in time-frequency analyses of cognitive electrophysiology data. *International Journal of Psychophysiology*, 111, 80-87. Retrieved from <Go to ISI>://WOS:000392554600008. doi:10.1016/j.ijpsycho.2016.02.001

意见 2: 坏道的数目为 0~14, 坏道的数量偏多, 坏道的数量对重测信度结果是否有重要的影响?

回应: 去除坏道也属于预处理的一部分, 对结果会有影响(详细内容见上), 但是本研究的数据整体坏道数目不多, 42 个被试的 rsEEG 中只有 3 个被试的坏道数量大于 5 个(其中 2 个被试有 7 个坏道, 1 个被试有 14 个坏道, 平均坏道数目为 1.61 ± 1.96 个), ERP 中有 5 个被试的坏道数量大于 5 个(其中 3 个被试有 6 个坏道, 1 个被试有 8 个坏道, 1 个被试有 9 个坏道, 平均坏道数目为 1.82 ± 1.73 个), 对整体数据质量的影响不大。坏道的位置主要位于前额叶和颞叶的位置, 属于周边区域, 受眼动、肌电影响较大, 与我们在空间域得出的重测信度结论一致。

意见 3: 数据的长度和 trial 的数量对重测信度结果是否有重要的影响?

回应: 我们的结论是长度和 trial 的数量可提高重测信度。时长和试次数对信号的信噪比的影响在先前的研究中也提及(Cassidy, Robertson, & O'Connell, 2012; Ding, Duan, Wang, & Lei, 2022; Duan et al., 2021), 更多的试次数可能有更好的重测信度。本研究中也存在两种状态试次数不统一的问题, 但与此同时我们也发现了试次数与重测信度的相关性: 闭眼 rsEEG 试次数最多, 对应了最大的 ICC, Oddball 偏差条件的 EEG 试次数最少对应了最小的 ICC。我们猜测试次数可能是通过影响信噪比间接影响重测信度的重要指标, 因此, 在任务态和静息态实验中增加试次数也是提高重测信度的有效方式。但需注意的是, 试次的增加也会延长实验时间, 对于被试的配合度、设备的长期工作性能(如电极更易脱落)等都提出更高的要求。根据专家的建议, 我们也在修改稿的讨论部分, 补充了下述关于试次数对结果影响的内容:

“.....与先前的研究一致(Cassidy et al., 2012; Ding et al., 2022; Duan et al., 2021), 我们推测试次数可能通过影响信噪比间接影响重测信度。.....”(第 14 页“4.4 多种任务态和静息态的对比”第 3 段蓝色的部分)

意见 4: ERP 分析中, 建议作者增加对 ERP 成分参数(如幅值)重测信度的考察。

回应: 我们目前的结果是对 ERP 在每个时间点的幅值的重测信度的考察, 出发点不是具体的成分而是时间点, 但在解释结果时也从 ERP 成分的角度进行了讨论。现在我们补充对每个被试在 Oddball 任务中的 P2 成分区间(150~275ms)和 PVT 的 N1 成分区间(130~200ms)的幅值求了 ICC 值, 结果分别为 $ICC_{PVT}=0.6703$ 、 $ICC_{Oddball-D}=0.4005$ 、 $ICC_{Oddball-S}=0.2833$ 。可见, 对成分幅值峰值的 ICC 计算结果与本研究中 200ms 左右的结果基本一致。根据专家的建议, 我们也在修改稿中补充了这部分计算结果:

“.....考虑到在计算 ERP 成分时, 通常按成分的幅值来进行分析, 因而我们又对 PVT 中的 N1 成分(130~200ms)(Hoedlmoser et al., 2011)和 Oddball 任务中的 P2 成分(150~275ms)(赵文瑞, 李陈渝, 陈军君, & 雷旭, 2020)分别计算了 ICC 值。从结果可见, 对成分的 ICC 计算结果与本研究中 200ms 左右的结果基本一致。PVT、Oddball-D 和 Oddball-S 分别得到的 ICC 为: 0.6703、0.4005 和 0.2833。.....”(第 8 页“3.3 时频域下的重测信度”第 1 段蓝色的

部分)

因为原文的“4.2 任务态 ERP 成分的重测性”中已指出了“三种信号都在 200 ms 左右出现了任务主要成分 (PVT 范式诱发的 N1 成分, Oddball 范式诱发的 P200 成分)”且结合 ERP 的成分结果做出了解释, 所以未再补充成分相关的讨论内容。

意见 5: 从数据采集 (三次采集) 来看, 作者可以计算短期和长期的重测信度? 分别评估和计算?

回应: 我们可以采用 3 次采集中两两之间的相关系数来分别计算短期和长期的重测信度。分析本研究的数据发现, Cz 电极处五种条件下间隔一周的两次测量结果的皮尔逊相关系数 (评估两次测量结果间的重测信度常用指标) 分别为 $r_{12_EO}=0.9999$ 、 $r_{12_EC}=1$ 、 $r_{12_PVT}=0.9928$ 、 $r_{12_Oddball-D}=0.9292$ 、 $r_{12_Oddball-S}=0.9565$; 间隔一个月的两次测量结果的皮尔逊相关系数分别为 $r_{13_EO}=1$ 、 $r_{13_EC}=0.9999$ 、 $r_{13_PVT}=0.9810$ 、 $r_{13_Oddball-D}=0.8084$ 、 $r_{13_Oddball-S}=0.9332$ 。结果如我们在文中 3.1 节提到的: 短期的重测信度高, 长期的重测信度略有降低。可见单独对短期或者长期的重测信度进行考察时参考意义有限, 所以我们计算 ICC 值比较短期和长期的多次 (本研究中是三次) 测量的重测信度, 对于临床应用的参考意义更大。

意见 6: 由于重测信度的影响因素众多, 建议作者慎重比较不同类型数据的重测信度并作出明确的结论。

回应: 感谢专家提出的宝贵建议, 我们已对文中讨论部分的用语进行了规范并在适当的地方补充了参考文献依据, 具体如下:

“我们根据影响数据质量的信噪比理论分析(Zuo et al., 2019).....” (第 12 页“4.1 静息态功率谱的重测性”第 1 段蓝色的部分)

“这部分结果符合基于信噪比理论提出的有效信息占比更大重测信度更高的假设(Zuo et al., 2019)。” (第 13 页“4.2 任务态 ERP 成分的重测性”第 2 段蓝色的部分)

“我们推测这可能基于某些频段代表了某种恒定的认知神经活动(Feyissa & Tatum, 2019).....” (第 14 页“4.4 多种任务态和静息态的对比”第 1 段蓝色的部分)

“.....从实践的角度来看, 研究人员所能做的就是尽最大努力测量和消除非感兴趣信号 (如运动伪迹等) 来提高信噪比, 并使用更优的方法提高信效度(Zuo et al., 2019)。” (第 14 页“4.4 多种任务态和静息态的对比”第 2 段蓝色的部分)

“.....与先前的研究一致(Cassidy et al., 2012; Ding et al., 2022; Duan et al., 2021), 我们推测测试次数可能通过影响信噪比间接影响重测信度。.....” (第 14 页“4.4 多种任务态和静息态的对比”第 3 段蓝色的部分)

“第一, 选择了传统的数据处理方法, 这虽然使得结果在更多的研究中具有可比性, 但考虑到目前对于预处理方法和参数影响脑电信号重测信度的结论尚不一致(Duan et al., 2021; Suarez-Revelo et al., 2016)。本研究的结论对于区别较大的预处理策略和许多新方法 (如功能连接, 脑网络等) 的参考意义有限, 未来的研究也应更多着眼于新策略新方法, 推动脑电数据处理和分析流程的标准化(Cohen, 2017)。” (第 14 页“4.5 局限性”第 1 段蓝色的部分)

“.....脑电重测性在多个实验条件、时频域指标、空间位置上都具有一定的可比性, 平均 ICC 重测信度维持在 0.33 到 0.55 之间。静息态比任务态具有更高的重测信度, 其中闭眼 rsEEG 的重测信度最好, PVT 任务所反应的精神警觉性也有较高重测稳定性, 这些都可以用试次数的增加和稳定的认知状态带来的高信噪比来解释。在频域上, 静息态 alpha 频段的重测信度最高; 在时域上, 任务态 200 ms 左右的成分最稳定; 在空间上, 中央区域的重测信度普遍优于外周区域, 时频空三个维度的证据初步支持了信噪比对重测信度的决定作用。总之, 我们的研究结果在更广泛的因素上讨论了重测信度。.....” (第 15 页“5 结论”第 1

段蓝色的部分)

新增参考文献:

- Cohen, M. X. (2017). Rigor and replication in time-frequency analyses of cognitive electrophysiology data. *International Journal of Psychophysiology*, 111, 80-87. Retrieved from <Go to ISI>://WOS:000392554600008. doi:10.1016/j.ijpsycho.2016.02.001
- Feyissa, A. M., & Tatum, W. O. (2019). Adult EEG. *Handb Clin Neurol*, 160, 103-124. doi:10.1016/b978-0-444-64032-1.00007-2
- Zuo, X. N., Xu, T., & Milham, M. P. (2019). Harnessing reliability for neuroscience research. *Nat Hum Behav*, 3(8), 768-771. doi:10.1038/s41562-019-0655-x
-

第二轮

审稿人 1 意见:

作者已经针对我的意见作出了相应修改, 本项基于 EEG 心理测量学信度研究可以为该领域提供实验设计和研究测量等方面指导性参考, 建议发表。

审稿人 2 意见:

作者已经很好地回答了我的意见, 没有更多的修改建议。

编委意见:

看了作者的回复, 同意接受。

主编意见:

本论文系统比较了睁眼、闭眼两种静息态脑电, 及精神运动警觉任务和 Oddball 任务中事件相关电位的重测信度。本论文的选题具有一定新颖性, 研究方法选用恰当, 数据处理过程科学规范。