

《心理学报》审稿意见与作者回应

题目：青少年手机压力与心理健康的关系：基于多元宇宙样分析和密集追踪方法

作者：黄顺森;来泉雄;张彩;赵心媚;代欣然;祁梦迪;王欢蕾;王文荣;王耘

第一轮

审稿人 1 意见：

本研究采用大样本调查和生态瞬时评估探究青少年数字化压力和心理健康的关系，大样本和生态瞬时评估都是比较难得的研究手段，但遗憾的是，研究关注的路径假设、研究设计、数据分析都存在关键缺陷。以下意见供参考：

回应：

a) 感谢您的宝贵意见，这对我们修改手稿具有很大帮助。

意见 1： 数字压力是使用数字化设备的压力，但本研究中其实只关注手机压力，数字化设备还有很多，包括教学中，现在提倡数字化教育，那些设备是不是也算数字化设备呢？因此，论文标题包括综述关注大概念“数字化压力”，研究内容关注小概念“手机压力”，二者可能是包含但不是等同的关系。建议更加具体化研究主体。

回应：

- a) 非常好的建议，感谢您的评论！
- b) 我们采纳了您的建议，使用手机压力作为具体化概念，在标题和文中相应的部分将“数字化压力”修改为“手机压力”。综述部分也有一些“数字化压力”相关研究作为背景予以保留。

意见 2： 背景综述也指出数字化压力和抑郁、焦虑、生活质量、幸福感等心理健康指标都有关系。为何本研究只关注抑郁和主观幸福感，这两个指标能够足够代表个体的心理健康吗？焦虑也很重要，为何不纳入呢？

回应：

- a) 非常好的观点，感谢您的评论！我们认为抑郁和主观幸福感不能代表个体心理健康的全部。未来研究还应该囊括更多积极和消极的指标，为此我们在讨论部分进行了说明（**见第 4 节，第四段，10~11 行**）。
- b) 目前关于心理健康的定义仍然存在争议，比如有过程性的或状态性的定义(Bhugra et al., 2013)，且并未有一个统一测量心理健康的工具。本文的抑郁和主观幸福更多的是从结果状态的角度定义心理健康。在一些心理健康的整合模型中，包括积极心理健康取向和消极心理健康取向(Mass é et al., 1998)。消极取向中，抑郁是一个重要的指标（如，截至 2000 年，关于抑郁的研究文献有 70856 篇，远高于焦虑的文章）；在积极取向中，幸福感是一个极其具有代表性的指标(李金珍 等, 2003)。因此，使用抑郁和主观幸福感代表个体心理健康水平具有一定的合理性。我们在综述部分进行了说明（**见 1.1 小节，第二段，14~18 行**）。考虑到大规模测试中，纳入过多的题目，学生作答疲劳，降低作答质量。因此，并未在设计的时候纳入更多变量（如焦虑）。

意见 3： 作者在总结已有研究不足的时候，提出的几点都值得再思考：（1）作者提到已有研

究过多聚焦于社交媒体使用和人际关系内容，没关注手机使用，然而手机是可以使用社交媒体的设备之一，二者虽不等同，但是也不互斥。(2) 作者提到已有研究忽视了积极心理健康维度，前面的数字媒体使用—数字化压力—心理健康理论模型针对的是消极维度。双因素心理健康模型并不能足够支撑为何要本研究要关注积极维度，更重要的是，数字化压力是会导致更多的主观幸福感还是更少的主观幸福感呢？如果是导致更少的主观幸福感，不也是消极的吗？(3) 作者提到“数字化压力与心理健康之间的关系在不同研究中不一致”，这一点很关键，是否有必要开展稳健性研究呢？但是却没有详细展开。前面论述的都是数字化压力会带来消极心理健康。(4) 作者提到不一致结果的解释，“这与研究者在数据分析过程中的主观操纵(如使用单一结果变量、用不同指标代表数字化压力、纳入不同的控制变量等)有关”，结论下的过于武断。难道不可能是因为被试特点不同(如文化)，测量工具，测量过程等其他因素影响吗？

回应：

- a) 感谢您的评论，非常具有建设性。
- b) 手机使用与社交媒体的确并不互斥，但是前者比后者有着更加广泛的使用情境，即手机使用不仅包含社交媒体，还包括游戏、网络课程、多任务操作等等，所以聚焦手机使用对青少年群体更具代表性(Huang et al., 2022)。我们在手稿中增加相关的论述来反应这一点(见 1.1 小节的第二段，4~6 行)。
- c) 关于双因素模型，原手稿中的论述可能还需要澄清。本研究纳入主观幸福感并不是假设压力会导致更多的幸福感(即压力产生积极的结果)，我们仍然假设压力产生消极的结果。双因素模型积极维度是指心理健康的指标是积极的，而不是前因变量产生的效果是积极的。我们在文中相应部分增加了相关论述来体现您的建议(见 1.1 小节第二段，13 行)。
- d) 前文论述并未过多探讨数字化压力与心理健康关系的不一致。主要目的是为了引入数字化压力相关概念和理论模型，所以在后续以往研究不足进行总结时才讨论不一致结果。对于后者，我们在文中(见 1.1 小节的第三段，1-2 行)部分详细展开进行了说明来体现您的建议。
- e) 关于“主观操纵的评论过于武断”，您的评论绝对是有道理的，我们对此做出了修改，改为“是不可重复的重要原因之一”(见 1.1 小节，第三段，8~9 行)。另外，我们做出这种判断主要是针对本研究，提出研究思路：在同一个数据库，也需要进行稳健性检验，因为不同的操纵同样也会导致不同的结果。Aarts 等(2015)在 Science 发文发现，使用同样的数据库，60%左右的心理学的研究不可重复。数字媒体领域内具有重要影响力的研究争议中(Orben & Przybylski, 2019a, 2019b; Twenge et al., 2022)，Orben 等(2019)和 Twenge 等(2022)在同一个数据库中，对屏幕时间影响心理健康得出了相反的结论，其原因在于二者的主观操纵(分析策略)不同。此外，对于已经成熟的领域，元分析可以帮助确定效应是否存在，但是对与新兴领域，十分有必要进行稳健性检验，以合理优化研究的资源和投入(黄顺森 等, 2023)。文化、测量工具、测量过程可能也是影响不一致结果的原因，为此我们在文献综述部分再次突出了数字媒体研究中的主观操纵性导致的不可重复性(见 1.1 小节，第三段，3~8 行)，同时也在讨论与局限部分进行了说明来体现您的建议(见第 4 节，第四段，11~12 行)。

意见 4： 研究者提出反刍是数字化压力和心理健康的中介变量，是否有证据佐证压力导致反刍。为何不可能是反刍作为一个调节变量，因为它可能是个体的特质(后续也说了反刍也具有特质性)，进而调节数字化压力和心理健康的关系？

回应：

- a) 感谢您的评论！我们增加了压力导致反刍的实证证据来体现您的建议（见 1.2.2 小节，11-12 行）。同时，反刍作为特质时可能作为一种稳定的调节变量，但是本研究从理论出发，更加关注反刍作为中介变量和其波动性特征。我们认为未来研究可以考虑反刍的特质性在数字化压力与心理健康中的调节作用，并在局限部分进行了说明（见第 4 节，第四段，12~13 行）。

意见 5： 研究者提出的压力-情绪视角中，前面综述以及后续测量都把压力定义为“消极情绪反应”，而不是指代压力源和压力强度，那么这个路径就显得尤为有问题，变成消极情绪反应-情绪-心理健康的路径。

回应：

- a) 感谢您的评论，这是一个好问题！本文对数字化压力的定义为“个体在每日与各种数字化情境的互动中，感知和体验到的消极情绪反应(Hefner & Vorderer, 2017; Huang et al., 2022; Steele et al., 2020)”，但这个定义并不意味着抛弃了压力源和压力强度。定义中提到“各种数字化情境”，各种情境包括聊天、游戏、学习、搜索和浏览信息等，这些情境其实就是压力源的一种体现（即 Lazarus 等人提到的环境刺激）。另外，本文使用的量表在压力强度上也有所体现(Huang et al., 2022)：量表评估的方式采用的 4 点计分（1=从来没有，4=总是），得分越高表明压力感知强度越高。为了避免可能的误解，我们在前言（见第 1 节，第一段，第 6 行）和方法部分（见 2.1.2 小节，第一段，第 7 行）增加了相关的论述来体现您的评论和意见。

意见 6： 关于认知和情绪这两个中介变量的位置，为何一定是认知在前，情绪在后。消极情绪不也可能增加反刍吗？

回应：

- a) 好问题！感谢您评论。本研究提出认知-情绪的链式中介是基于 Selby 等 (2008, 2013) 提出的情绪级联模型，该模型认为反刍过程是行为失调的根本原因，反刍思维导致负性情绪循环叠加，迫使个体行为异常。这并不意味着反刍一定在前，只是在情绪级联模型中，反刍更加占据核心和重要位置(李莹娜, 李丹, 2022)，为此理论上我们认为反刍更加靠前。未来研究应该深入探讨反刍和情绪的协同和相互作用，及对心理健康的影响。我们在文献综述增加了理论论述（见 1.2.4 小节，4-5 行），同时讨论部分的未来展望（第 4 节，第四段，7~10 行）也对此进行了说明。

意见 7： 研究 1 假设提到，“数字化压力与心理健康之间存在稳健联系，且效应量较大”有什么依据获得这个假设呢？综述似乎没有提及效应量大小的讨论。如果已有研究效应都挺大的，为何还要进行稳健性分析呢？

回应：

- a) 感谢您细致的评论！
- b) “效应量较大”的假设可能存在表述歧义，我们在修改稿中删除了这个表述。
- c) 为了增加稳健性分析的合理性，我们在文献综述部分简明地对效应量大小进行了讨论。同时从经验角度出发，强调稳健性分析的重要性。一是主观操作性问题引发的效应差异和效应矛盾现象。二是考虑到，即使是一些成熟的领域，也存在矛盾性的结果，通常需要进行元分析进行效应检验的大小。对于新兴领域，避免不必要的研究资源的浪费，也有必要进行稳健性分析，而多元宇宙样分析非常适合新兴领域进行效应的稳健性分析（黄顺森 等, 2023）。我们在文献综述部分对此进行了说明（见 1.1 小节，第三段，9~15 行）。

意见 8: 研究 1 中, 为何只选择这两个年级, 导致年龄范围很狭窄, 是否能够有效代表“青少年”群体?

回应:

- a) 感谢您的评论! 研究 1 是基于“区域教育质量健康体检”项目, 该项目旨在评估学生的教育质量, 在样本选择上具有其自身原因和属性。根据《义务教育法》, 四年级学生年龄在 9-10 岁左右, 八年级学生在 13-14 岁左右, 属于青少年群体, 但是无法完全代表青少年群体。为此, 我们在讨论的局限性部分对这点进行了说明(见第 4 节, 第四段, 2~3 行)。

意见 9: 数据缺失类型应该是三种: 完全随机缺失, 随机缺失, 非随机缺失。这里的“非完全随机缺失”是什么意思?

回应:

- a) 感谢您细心的评论! “非完全随机缺失”属于笔误, 已进行修正(见 2.1.1 小节)!

意见 10: 手机使用程度(时间)应该是一个非常重要的协变量或者调节变量, 为何没有考虑? 尽管作者剔除了没有手机使用经历的被试, 但是仍需要考虑使用程度的问题。

回应:

- a) 感谢您的评论。我们采纳了您的建议, 纳入了手机使用时间作为协变量(见 2.1.2 部分, 第四段, 第四行), 并更新了图 2 以及相应的结果描述(见 2.2 小节, 第二段)

意见 11: 研究者是如何界定手机使用经历的呢?

回应:

- a) 感谢您仔细的评论。我们采用题目“你平时使用手机吗, 包括用你自己或者家人的手机?”评估手机使用经历, 采用 4 点计分(1=从不, 2=有时, 3=经常, 4=总是)。选择“从不”使用手机的被试未被纳入分析。我们在手稿方法部分进行了说明(见 2.1.3 小节, 第一段, 5~7 行)。

意见 12: 多元宇宙样分析是一个比较新的数据处理策略。总体上目前采用这个策略的研究还是比较有限。关键在于, 是否有必要采用这个方法, 或者说这个方法能否带来一些有意义的结果。从方法上看, 这个策略是尽量在减少选择性分析的偏差。具体从这个研究来说, 实际上分析策略也就只针对手机压力这个测量是采用条目处理, 协变量是否进行不同组合, 那为何不考虑结果变量 y 的测查方式, 变量类型等呢? 另外, 本研究所选择的自变量、因变量和协变量是否在最开始的选择中也存在选择偏差呢? 例如研究者并非一开始有具体的研究问题而去设计调查或者实验, 而是从一个已有的大数据集中挑出其中的几个可能关系比较相关的变量或者特定测量工具进行分析和报告。总之, 虽然方法看起来很复杂, 但是实际上带来的意义感觉并不是很大, 能控制的偏差也很有限, 不能因此直接得出稳健的结论。也许更值得考虑的是: 学校、省市、地区水平的变异, 例如采用多水平模型。

回应:

- a) 感谢您具有建设性的评论。手机压力测量条目的处理方式、协变量的组合方式以及因变量指标是本研究减少选择性分析偏差的主要方面。诚如您所言, 变量类型和变量测查方式的确可能是造成效应差异的原因, 但是同一个数据库选择性报告和选择性分析也是心理科学结果不可重复的核心原因之一(Aarts et al., 2015; Orben & Przybylski, 2019b; Twenge et al., 2022)。针对同一个数据库进行多元宇宙样分析, 有利于检验效应的稳健性, 主要因为新兴领域的研究积累较少, 不像成熟领域已经具备不同测查方式和变量类

型的研究(黄顺森 等, 2023)。为此, 我们在文献综述部分对此进行了论述(见 1.1 小节, 第三段, 9~15 行)。

- b) 关于变量的选择, 无论是自变量、因变量还是控制变量, 为了避免选择性偏差, 我们的选择均具有理论和实证研究的支持(如见 1.1 小节, 第二段, 14~18 行; 2.1.3 小节, 12~16 行; 2.1.2 小节, 第四段)。但是, 本研究确实不可能测量一个构念的所有方面(比如心理健康), 我们在讨论局限部分进行了说明(见第 4 节, 第四段, 10~11 行)。
- c) 对于多水平模型, 考虑到多水平模型的前提是足够大的 ICC 和足够大的样本量。一般来讲 ICC 多介于.05~.20, ICC 的中位数落在.12, 所以.12 被认为是判定 ICC 足够大的标准(邱皓政, 2017, p78); 对于样本量(群组)而言, 足够的样本量才能使多水平模型才能获得稳健和无偏的估计, 如果样本量(群组)太少, 则会导致 Level-1 和 Level-2 固定效果和随机效果的错误估计(Maas & Hox, 2005)。经验性研究指出, 多水平稳健估计的样本量标准为 50 个群组及以上(Ali et al., 2019; Maas & Hox, 2005)。本研究中, 仅学校水平和班级水平符合该标准。同时, 本研究中手机压力(总分)、抑郁、幸福感的三水平(个体-班级-学校)嵌套模型的 ICC 在水平 2(班级-学校)为 0.169、0.171、0.187, 在水平 3(学校)为 0.100、0.098、0.101。为此, 我们重新使用多水平嵌套的多元宇宙样分析模型进行估计(见 2.1.3 小节, 26~29 行; 2.2 小节第二段及图 2), 结果并未发现与非多水平模型之间存在很大的差异。

意见 13: “首先, 手机压力与主观幸福感策略组合(图 1 (A、B))有 80 种”, 这里是图 2 吧? 回应:

- a) 感谢您细致的评论, 已进行修正。

意见 14: 关于预测效应, 如果是回归分析, 是采用什么指标作为效应量呢? 这部分没有太明确。

回应:

- a) 感谢您的提醒!
- b) 我们使用的效应量标准是偏相关系数的平方, 我们在 2.1.3 小节的结尾部分提到了效应量指标, 并进行了更加详细的说明(见 2.1.3 小节, 24~25 行)。

意见 15: 图 1 对于手机压力的描述性统计其实并不是很直观, 建议呈现均值标准差, 让读者衡量被试的手机压力整体情况。

回应:

- a) 感谢您的好建议, 我们非常乐意呈现这些内容。我们在附件中将均值和标准差作为补充材料(见附件 2 中的图 S1)

意见 16: 作者好像进行了手机压力的流行率分析, 如“52.6%四年级学生和 78.2%八年级学生有过手机压力的体验”, 如果这里不是说手机经历的体验, 而是手机压力的体验, 既然手机压力有流行率, 就应该有划界分(cut-off), 问卷描述部分没有看见描述?

回应:

- a) 感谢您的评论, 研究并未进行严格的流行率分析, 仅仅是根据被试对题目选项(从来没有、偶尔、经常、总是)的回答进行判定。我们在文中使用“流行情况”这个词可能容易引起误解, 为此我们修改了表述(见 2.1.3 小节, 第三行)。

意见 17: 生态瞬时评估中, 很奇怪的一点是, 为何前 18 天报告反刍, 后 18 天报告情绪,

而不是同时报告？

回应：

- a) 感谢您的评论，根据我们的认知-情绪模型的假设，我们前 17 天和后 18 天分别报告反刍和情绪，并未同时报反刍和情绪。

意见 18：“Little’s MCAR 检查发现数据属于完全随机缺失”，能判断“完全随机缺失”？

回应：

- a) 感谢您仔细的评论。Little’s MCAR 检验可以实现，如果结果不显著，表明数据属于完全随机缺失(刘红云, 2019, p8)。

意见 19： 虽然研究使用了生态瞬时评估，但是设计有一些不足，也没有很好地采用合适的统计手段突出生态瞬时评估的优势。一方面，研究的因变量只测了两次，而不是生态瞬时评估，导致因果模型信息不完全。而且，两个中介变量分为前后两个阶段测量让人费解。另一方面，均值中介模型和波动中介模型不能很好勾勒变量的相互关系和发展变化，潜变量结构方程模型，潜变量增长曲线模型也许是更好的考虑。

回应：

- a) 感谢您的评论！本研究在探讨手机压力对心理健康的作用机制的时候聚焦于认知-情绪的特质性（均值）和波动性（标准差），并不过多聚焦变量的交互关系和变化（如相互作用）。对于情绪和反刍这两个变量而言，回溯法和单次测查可能因为记忆错误导致偏差，通过日记法等密集追踪手段可以获得较为稳定的认知-情绪强度的特质性指标和基于标准差的波动性指标（见 1.3 小节，第一段）。实践研究中，也有研究者将强度和波动性等指标纳入中介模型（如 Keng & Tong (2016)）。因此，本研究从特质性和波动性的角度出发探索中介机制。我们认为未来研究或许可以进一步探讨中介机制的相互作用和发展变化，为此讨论部分也进行了说明（见第 4 节，第四段，7~10 行）。

意见 20： 研究存在一些细节错误，例如多处引用格式错误：温忠麟 et al., 2016、周浩 & 龙立荣, 2004、黄顺森等, 2023.....

回应：

- a) 感谢您细致的评论！我们在手稿中已经进行修改，修改内容包括中文的 et al 改为“等人”，&符号不再使用于中文参考文献，中文作者与“等人”隔开。

.....

审稿人 2 意见：

该文章研究采用多元宇宙样分析和密集追踪方法探讨数字化压力与青少年心理健康之间的关系及作用机制，研究扩展了数字化时代背景下“压力-认知/情绪”理论和“媒体使用-数字化压力-心理健康”模型，具有一定的现实意义和创新性，有助于理解数字化背景下青少年心理健康形成和发展的内在原因。总体而言，问题的提出和文献综述较好，研究思路明确，论文写作较规范，结论合理恰当。有几个具体意见和问题如下：

回应：

- a) 感谢您的评论，我们获益良多！

意见 1： 作者提到数字化压力指使用数字化设备（如智能手机、社交媒体和电脑）的过程中感受到的压力，但在研究工具的使用上主要青少年手机压力量表来测量感知到的数字化压力，手机压力属于数字化压力的一部分，能否代表数字化压力？而且作者在研究方法、结果、讨

论等部分均表述为“手机压力”，为什么不直接把题目改为“智能手机使用压力与青少年心理健康的关系”？须进一步阐述理由。另外，关键词有点多，3-5个关键词为宜。

回应：

- a) 我们采纳了您的意见，在研究主题上进一步聚焦到智能手机使用压力上。我们将题目改成了“智能手机使用压力与青少年心理健康的关系”，在正文中也修改了相应表述。此外，我们将关键词从7个缩减为5个。

意见 2：在摘要中，作者提到“反刍和消极情绪的强度”，而文中主要表述为“均值”，均值能否代表强度？建议前后统一表述。文中核心变量手机压力测验，研究1仅采用9个手机压力的题目，其图1中的描述统计只有每个题项的百分比，没有统计整体或者六维度的平均分+标准差，而研究2则使用了手机压力6维度的结果进行分析，为什么？

回应：

- a) 感谢您非常细致的意见和评论！根据您的意见，我们将手稿中的表述改成强度，并在文中对均值和强度进行了说明（见1.2.3小节，第六行；1.3小节，第一段；1.3小节，第二段，第2/7行）。
- b) 关于研究1和研究2手机压力的测量，图1没有整体和六维度的平均分是因为研究1中使用的是手机压力简版量表，是单一维度，所有没有六维度的均分和标准差，只有整体的，使用简版是为了避免大规模调查中的作答疲劳，从而有利于施测和提高作答质量（见2.1.2小节，第一段，第五行）。此外，我们增加了每个题目和总分的平均分与标准差（见附件2中的图S1）。而研究2采用了手机压力完整版量表，因此可以报告各维度的平均分与标准差。

意见 3：青少年心理健康的指标包括多方面（如焦虑、人格、社会适应等），作者为什么选择抑郁和主观幸福感作为研究变量指标？而不选择焦虑等作为研究指标，需阐述清楚。

回应：

- a) 感谢您的评论！我们在修改稿中，根据理论和经验的角度（如心理健康双因素模型、积极心理学取向），进一步说明消极维度选择抑郁，积极维度选择主观幸福感的原因（见1.1小节，第二段，9~18行）。诚然，两个指标也不能完全代表青少年心理健康，为此，我们在局限性部分进行了说明（第4节，第四段，10~11行）。

意见 4：在研究1的被试选择上，作者进行了大规模的调查，但为何只选择了四年级和八年级的学生进行研究？为什么没有选择10年级？四年级和八年级是否可以代表青少年？

回应：

- a) 感谢您这方面的考量！研究1是基于“区域教育质量健康体检”项目，该项目旨在评估学生的教育质量，在样本选择上具有其自身原因和属性。根据《义务教育法》，四年级学生年龄在9-10岁左右，八年级学生在13-14岁左右，属于青少年群体，但是无法完全代表青少年群体。为此，我们在局限性部分对这点进行了说明（见第4节，第四段，第三行）。

意见 5：研究2进行了为期35天的密集追踪调查，学生需要每天进行报告。这45天是开学之初，还是什么其它时间？学生在这个期间受到各种环境因素的影响存在很大差异（如学业压力、同伴交往、家庭因素等），作者如何考虑这些控制这些影响因素的？

回应：

- a) 感谢您的评论。这45天是从10月中旬开始到11月底结束（详见预注册地址）；为了减少

各种环境因素的影响，我们在施测的时候，招募了志愿老师协助记录和检查每个班学生的情况，以最大化保证环境的稳定性。我们在方法部分对此进行了说明（见 3.1.1 小节，第一段，第 7~8 行）。

意见 6: 作者在研究 2 发现反刍、日常情绪的链式中介作用，但并没有较好地解释这其中的原因，建议增加这部分讨论。

回应:

- a) 感谢您的建议，非常具有建设性！
- b) 我们在讨论部分进行了修改，更加深入讨论了这个问题（见 3.3.3 小节）。

意见 7: 文中的引用和参考文献格式部分有误，需修改。

回应:

- a) 感谢您细心的意见！我们在手稿中已经进行修改，修改内容包括中文的 et al 改为“等人”，& 符号不再使用于中文参考文献，中文作者与“等人”隔开。
-

审稿人 3 意见:

这篇文章利用新的方法探索了青少年数字化压力与心理健康的关系。研究方法具有创新性，结果可靠。我的评论和建议如下，希望对作者有用。

回应:

- a) 感谢您的建设性评论，对提高我们的稿件帮助很大。

意见 1: 前言写的很清楚，但是前言并未明确提及积极心理学视角，以及数字化压力影响幸福感的可能机制。建议单独增加一个部分讨论积极心理学视角。

回应:

- a) 感谢您的建设性意见，我们在综述部分增加了这些内容（见 1.1 小节，第二段，9~12 行）。

意见 2: 另外，并非所有作者都会熟悉多元宇宙抽样以及密集追踪，建议增加相关描述，以及对使用这些方法的必要性进行更多阐述和解释。

回应:

- a) 感谢您的建议！
- b) 我们在文献综述和方法部分增加了对多元宇宙样方法（见 1.1 小节，第二段，22~28 行）和密集追踪方法（见 1.3 小节，第一段，8~9 行；3.1.3 小节，第一段，7~9 行）的必要性的阐述或描述。

意见 3: 研究 1 中，在如此庞大的数据结构下，而且存在明显数据嵌套的情况下，我希望作者能考虑数据的嵌套结构，班级嵌套学校等等，至少要汇报 ICC，看是否需要进行多层线性模型分析，如果不考虑数据的嵌套结构，可能会导致结果也不一定可靠。

回应:

- a) 感谢您的建设性意见。考虑到数据的嵌套结构（三水平的多层模型），我们在修改过程中专门计算了 ICC，并把 ICC 大于等于 .12 作为是否进行多层线性模型的标准(邱皓政, 2017, p78)。Level 2（班级-学校）水平的 ICC 在手机压力、抑郁、幸福感中分别为：0.169、0.171、0.187，Level 3（学校）水平的 ICC 在手机压力、抑郁、幸福感中分别

为 0.100、0.098、0.101。为此我们在 *specr* R 包中进行“三水平嵌套的多水平模型”的多元宇宙样分析(Masur & Scharnow, 2020) (见 2.1.3 小节, 26~29 行)。在主观幸福感的模型中, 回归系数从-0.214 ~ -0.113 变成-0.175 ~ -0.074; 在抑郁的模型中, 回归系数从 0.401 ~ 0.580 变成 0.282 ~ 0.473。总体而言, 多层线性模型未使研究 1 的结论明显改变。我们根据新的结果, 对手稿进行了修改(见 2.2 小节第二段及图 2)。

意见 4: 研究 2 中, 最新的第七版 apa, 对于大于 .05 的 p 值, 应该用 $p = .XXX$ 的报告格式(保留三位小数)。

回应:

a) 感谢您细致的建议。我们在文稿中修正了这个问题, 同时参考期刊的要求, 对于大于 0.05 和 0.001 的 p 值, 均在图中进行了呈现(见图 3 和图 4)。

意见 5: 研究 2 中, 根据我的经验, 密集追踪采样非常难, 这点非常赞赏本文作者能做到 35 天的密集追踪调查。但是作者并未对具体的采样进行描述, 我建议详细描述密集追踪采用的采样过程, 包括伦理, 父母知情同意, 是否培训, 是否有报酬, 样本流失情况。另外, 请详细描述这个样本, 包括性别等因素。

回应:

a) 感谢您宝贵的意见, 我们在修改稿中进一步对采样过程和样本的基本信息进行了描述(见 3.1.1 小节, 5-11 行)。

意见 6: 研究 2 中的, T1-T4 的时间间隔是如何确定的? 包括 1 周到 18 天不等。

回应:

a) 感谢您的评论。T2、T3 同属于密集追踪, 为了更好的探讨认知-情绪作用机制, 使用对半拆分的方法对密集追踪时间间隔进行拆分。T1、T4 与密集追踪的时间间隔为 1 周。在密集追踪研究中, 前测、后测与密集追踪的时间间隔没有统一标准, 一周间隔较为常用, 两周、三周等在实践中也均有使用(Brailovskaia et al., 2020; Hopko & Mullane, 2008; Smith et al., 2019; Voogt et al., 2014)。

意见 7: 研究 2 的样本还是存在数据的嵌套, 学生嵌套于 10 个班级, 请至少算一下 ICC 判断是否需要多层线性模型分析。

回应:

a) 感谢您的具体建议。我们计算了前测和后测自变量与因变量的 ICC, 发现手机压力前测 ICC=0.011、抑郁前测 ICC=0.000、幸福感前测 ICC=0.014、抑郁后测 ICC=0.027、幸福感后测 ICC=0.045。按照前人提出的标准(一般来讲 ICC 多介于 0.05~0.20, ICC 的中位数落在 0.12, 为此, 0.12 被认为是判定 ICC 足够大的标准(邱皓政, 2017, p78)。此外, 经验性研究指出, 多水平稳健估计的样本量标准为 50 个群组及以上(Ali et al., 2019; Maas & Hox, 2005), 样本量较少会引发估计偏差。我们在手稿中报告了 ICC 的情况, 但不再进行多水平模型分析(见 3.1.3 小节 14-15 行)。

意见 8: 另外, 研究 2 的分析非常复杂, 我有些不太理解作者把均值和波动性独立分开分析的做法, 作者能够提供使用均值和波动性独立模型的理由和相应的文献。

回应:

a) 感谢您的宝贵建议! 在动态性数据的指标(如均值、波动性、敏感性等)中, 大部分指标之间存在较高的共线性(有研究发现强度和波动性的共线性最高)(Dawel et al., 2023),

如果放入同一个模型,则会因为共线性问题导致估计偏差。同时,虽然二者存在共线性,但是主成分分析发现均值和波动性从属于两个不同的成分(Dejonckheere et al., 2019)。因此将均值和波动性分开进行中介效应检验是合理的。在经验研究中,研究者也提出不应将均值和波动性指标放入同一个模型(Walerius et al., 2016)。例如,为探讨正念对情绪动态性的影响机制,将情绪动态性多个指标分别放入中介模型中(Keng & Tong, 2016)。我们在综述部分(见**3.1.3 小节, 7~9 行**)对这一要点进行了说明。

参考文献:

- Aarts, A. A., Anderson, J. E., Anderson, C. J., Attridge, P. R., Attwood, A., Axt, J., Babel, M., Bahník, Š., Baranski, E., Barnett-Cowan, M., Bartmess, E., Beer, J., Bell, R., Bentley, H., Beyan, L., Binion, G., Borsboom, D., Bosch, A., Bosco, F. A., ... Zuni, K. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716.
- Ali, A., Ali, S., Khan, S. A., Khan, D. M., Abbas, K., Khalil, A., Manzoor, S., & Khalil, U. (2019). Sample size issues in multilevel logistic regression models. *PLoS ONE*, *14*(11), 1–13.
- Bhugra, D., Till, A., & Sartorius, N. (2013). What is mental health? *International Journal of Social Psychiatry*, *59*(1), 3–4.
- Brailovskaia, J., Ströse, F., Schillack, H., & Margraf, J. (2020). Less Facebook use – More well-being and a healthier lifestyle? An experimental intervention study. *Computers in Human Behavior*, *108*(March).
- Dawel, A., Gulliver, A., Farrer, L. M., Kalokerinos, E. K., Cherbuin, N., Calear, A. L., McCallum, S., Morse, A. R., & Monaghan, C. (2023). Do Emotion Intensity, Variability, Differentiation, Co-Occurrence, and Positive–Negative Ratios Make Unique Contributions to Predicting Longitudinal Change in Psychological Distress and Well-Being? *Emotion*.
- Dejonckheere, E., Mestdagh, M., Houben, M., Rutten, I., Sels, L., Kuppens, P., & Tuerlinckx, F. (2019). Complex affect dynamics add limited information to the prediction of psychological well-being. *Nature Human Behaviour*, *3*(5), 478–491.
- Hefner, D., & Vorderer, P. (2017). Digital stress: Permanent Connectedness and Multitasking. In L. Reinecke & M. B. Oliver (Eds.), *The handbook of Media use and well-being*. Routledge.
- Hopko, D. R., & Mullane, C. M. (2008). Exploring the relation of depression and overt behavior with daily diaries. *Behaviour Research and Therapy*, *46*(9), 1085–1089.
- Huang, S., Lai, X., Ke, L., Qin, X., Yan, J. J., Xie, Y., Dai, X., & Wang, Y. (2022). Smartphone stress: Concept, structure, and development of measurement among adolescents. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *16*(5), Article 1.
- Keng, S. L., & Tong, E. M. W. (2016). Riding the tide of emotions with mindfulness: Mindfulness, affect dynamics, and the mediating role of coping. *Emotion*, *16*(5), 706–718.
- Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*(3), 86–92.
- Massé R., Poulin, C., Dassa, C., Lambert, J., Bédair, S., & Battaglini, A. (1998). The structure of mental health: Higher-order confirmatory factor analyses of psychological distress and well-being measures. *Social Indicators Research*, *45*(1–3), 475–504.
- Masur, P. K., & Scharrow, M. (2020). *specr: Conducting and Visualizing Specification Curve Analyses* (Version 0.2.1). R groups. <https://masurp.github.io/specr/>, <https://github.com/masurp/specr>
- Orben, A., & Przybylski, A. K. (2019a). Screens, Teens, and Psychological Well-Being: Evidence From Three Time-Use-Diary Studies. *Psychological Science*, *30*(5), 682–696.
- Orben, A., & Przybylski, A. K. (2019b). The association between adolescent well-being and digital technology use. *Nature Human Behaviour*, *3*(2), 173–182.

- Selby, E. A., Anestis, M. D., & Joiner, T. E. (2008). Understanding the relationship between emotional and behavioral dysregulation: Emotional cascades. *Behaviour Research and Therapy*, 46(5), 593–611.
- Selby, E. A., Franklin, J., Carson-Wong, A., & Rizvi, S. L. (2013). Emotional cascades and self-injury: Investigating instability of rumination and negative emotion. *Journal of Clinical Psychology*, 69(12), 1213–1227.
- Smith, J. M., Whisenhunt, B. L., Buchanan, E. M., & Hudson, D. L. (2019). Evaluating the effectiveness of ecological momentary assessment and intervention targeting body checking behaviors. *Eating Disorders*, 27(6), 521–537.
- Steele, R. G., Hall, J. A., & Christofferson, J. L. (2020). Conceptualizing Digital Stress in Adolescents and Young Adults: Toward the Development of an Empirically Based Model. *Clinical Child and Family Psychology Review*, 23(1), 15–26.
- Twenge, J. M., Haidt, J., Lozano, J., & Cummins, K. M. (2022). Specification curve analysis shows that social media use is linked to poor mental health, especially among girls. *Acta Psychologica*, 224(November 2021), 103512.
- Voogt, C., Kuntsche, E., Kleinjan, M., Poelen, E., & Engels, R. (2014). Using ecological momentary assessment to test the effectiveness of a web-based brief alcohol intervention over time among heavy-drinking students: Randomized controlled trial. *Journal of Medical Internet Research*, 16(1).
- Walerius, D. M., Fogleman, N. D., & Rosen, P. J. (2016). The Role of ADHD and Negative Emotional Lability in Predicting Changes in Parenting Daily Hassles. *Journal of Child and Family Studies*, 25(7), 2279–2291.
- 刘红云. (2019). *高级心理统计分析*. 中国人民大学出版社.
- 李莹娜, & 李丹. (2022). 非自杀性自伤的相关情绪调节模型研究进展. *心理学进展*, 12(01), 246–252.
- 李金珍, 王文忠, & 施建农. (2003). 积极心理学: 一种新的研究方向. *心理科学进展*, 11(3), 321–327.
- 邱皓政. (2017). *多层次模式与纵向资料分析: Mplus 8解析应用*. 五南图书出版公司.
- 黄顺森, 陈豪杰, 来梟雄, 代欣然, & 王耘. (2023). 多元宇宙样分析: 简介及应用. *心理科学进展*, 31(2), 196–208.

第二轮

审稿人 2 意见: 作者对专家的建议能较好地做出回应, 参考文献格式请按照《心理学报》要求, 如中文期刊要用斜体。

回应:

- 感谢您的宝贵意见, 这对我们修改手稿具有很大帮助, 我们对手稿进行了以下修改:
- 参考文献的中文期刊名字均改为斜体, 其余检查后发现有误的参考文献格式也均已进行了修改。
- 根据《心理学报》要求, 我们还进行了以下修改: 希腊字母(α 、 β 、 χ^2 、 η^2)、上标和下标的数字和文字, 不用斜体; 手稿行距修改为 1.5 倍; 标题字体取消加粗。

审稿人 3 意见: 感谢作者的修改和回复。

回应: 感谢专家的评阅和意见!

第三轮

编委意见:

作者对三位审稿专家的意见做了较好的回应。仍存在以下问题待完善:

意见 1: 请精简文字。根据学报要求, 前言和讨论部分, 各不超过 3500 字; 方法与结果部分也尽量精练。

回应: 感谢编委的意见, 我们对前言和讨论部分均进行了优化和精炼, 在不改变原意和对先前审稿人意见的修改的前提下, 我们将前言和讨论部分各缩减到 3500 字内。同时也将方法与结果部分进行了精简(不改变科学、客观性的情况下减少 250 字左右)。所有的修改均标注为蓝色字体。

意见 2: 本研究的亮点在于使用了“多元宇宙样分析”, 但方法只是工具, 应在引言和讨论充分体现使用该方法获得的新发现、新贡献。

回应: 感谢编委的意见, 在确保前言和讨论部分字数的情况下, 我们进一步在讨论(第 4 节, 第二段, 5-7 行; 第 4 节, 第三段, 8-9 行)和前言部分(第 1.1 节, 第三段, 倒数最后四行)突出使用该方法获得的新发现和新贡献。

意见 3: 英文摘要:

长英文摘要是为增加外库引用, 需注意一些表述:

(1) 首次出现的缩写, 应给全称, 比如: NSRPD.....

(2) 表述问题, 比如: The study concluded that rumination and negative emotion intensity and fluctuation play a mediating role..

回应: 感谢编委的意见, 我们在英文摘要中对增加了 NSRPD 首次缩写的全文。此外, 对于英文摘要的表述问题, 我们对您提出的例子进行了修改, 同时针对类似问题(请专业人士)再润色了一次。所有的修改均使用蓝色字体显示。

第四轮

编委意见: 该研究使用的方法较为新颖, 也有一定的学术发现。经多轮修改, 论文质量有较大的提升。推荐发表。

主编意见: 本研究借助多元宇宙样分析和密集追踪方法, 对青少年手机压力与心理健康的关系进行了考察。本论文的研究框架清晰, 数据处理过程科学规范, 最终所获得研究结论不但可以推动相关理论的发展, 在一定程度上还可以为青少年心理健康的干预提供启发, 具有较为重要的理论和现实意义。