

《心理学报》审稿意见与作者回应

题目：中国国民心理健康素养的现状与特点

作者：江光荣 李丹阳 任志洪 闫玉朋 伍新春 朱旭 于丽霞 夏勉 李凤兰 韦辉
张衍 赵春晓 张琳

第一轮

审稿人 1 意见：

意见 1：本文采用的问卷内部一致性信度在 0.64-0.76 之间，信度不是很高；并且在计算问卷总分时，将问卷结果均视为二分变量而非连续变量进行处理，这样做的依据是什么？此外，对于采用二值计分的量表，CTT 理论的适用是有限的，建议换用 IRT 模型重新处理数据，给出题目的难度、区分度、测验总信息以及题目信息等基本指标。

回应：非常感谢专家指出我们在问卷计分及测量学特征分析方面存在的问题，以及提出用 IRT 模型重新分析数据并报告结果的建议。以下我们将针对两个问题分别展开说明。

(1) 对于计算总分时将问卷结果视为二分变量来处理的问题。我们这样做的考虑有两点。一是本研究使用的问卷包括 0、1 计分的判断题和 5 级计分的 Likert 量表题两种题型，由于低测度级别的数据无法转换为高测度级别的数据，因此为了能够得出心理健康素养总分，需要将 1~5 级计分的题目得分转换为 0、1 计分。需要说明的是，该问卷在编制过程中，分别对判断题和 Likert 量表题的测量学特征进行了考察。我们也参考了一些同行的做法。比如，卢家楣等(2018)采用的《中国大学生情感素质问卷》因 6 个分问卷计分方式不同，在统计分析时则将各分问卷实际得分转换成 10 分制计分，以方便统计和理解。

二是 Likert 量表题测量的是维护和促进自己与他人心理健康、应对自己与他人心理疾病的态度和习惯。虽然通常来讲，态度和行为习惯没有正误之分，不能将其转换为 0、1 计分做正误判断，但是心理健康素养实际上蕴含着价值评判。不仅知识和观念有正误之分，而且态度和习惯也有积极和消极之分。态度和习惯趋向于积极一端，意味着个体重视自身和他人的心理健康、对自身和他人的心理状况敏于觉察，能够恰当地自助、求助及助人(江光荣 等, 2020)。态度和习惯趋向于消极一端，则意味着个体对自身和他人的心理健康和心理疾病不够重视，并且缺乏应对心理疾病、维护心理健康的行为习惯和技能。

以上考虑是否恰当，恳请审稿专家进一步批评指正。

(2) 对于使用 IRT 模型分析问卷测量学指标的建议。限于篇幅，我们在内部一致性信度方面仅报告了 6 个分问卷的内部一致性信度范围。实际上，该问卷在编制过程中分别考察了 30 个判断题和 30 个 Likert 量表题的内部一致性，分别为 0.823 和 0.883，在可接受范围内(Wu et al., manuscript)。

但诚如专家所言，CTT 理论的适用性有限，对于二值计分的量表使用 IRT 模型更为恰当。郭庆科、陈英敏和孟庆茂(2005)检验了自陈量表式测验应用 IRT 的可行性，指出对于 0、1 二级计分的量表，采用双参数 Logistic 模型分析是可行的。因此，我们用此方法，对 6 个分问卷采用 0、1 计分时的难度和区分度等基本指标分析如下。

使用 SPSS 20.0，运用主轴因素法(PAF)对 6 个分问卷的数据分别进行因素分析，以进行单维性检验。结果如表 1 所示，各分问卷的第一特征根都大于第二特征根，但比值都没有达到 5:1 的标准(戴海琦, 2006)。不过 IRT 的单维性判断一直没有统一标准，单维性假设被轻微违背的情况下 IRT 仍具有足够的稳健性(郭庆科, 陈英敏, 孟庆茂, 2005)。

使用 Mplus 8.3,采用双参数 Logistic 模型对数据进行分析。结果显示,有 48 个题目(80%)的区分度参数在(0.5, 2)的合理范围(杜文久, 周娟, 李洪波, 2013), 有 5 个题目(8.3%)的区分度过低(<0.3; 曹尚 等, 2016); 53 个题目(88.3%)的难度参数在(-3, 3)的合理范围内(曹尚 等, 2016)。

表 1 各分问卷用因素分析法抽取的第一和第二特征根

分问卷	分问卷 1	分问卷 2	分问卷 3	分问卷 4	分问卷 5	分问卷 6
特征根 1	2.713	3.112	1.711	2.113	1.652	2.231
特征根 2	1.175	1.874	——	1.073	1.024	1.449
比例	2.309	1.661	——	1.969	1.613	1.540

注: 分问卷 1~6 分别为: 心理健康相关知识和观念、心理疾病相关知识和观念、维护和促进自己心理健康的态度和习惯、应对自己心理疾病的态度和习惯、维护和促进他人心理健康的态度和习惯、应对他人心理疾病的态度和习惯; “——”表示该分问卷仅抽取 1 个因子。

基于以上分析,我们对方法部分研究工具的 2.2.1“《国民心理健康素养问卷》”第二段加以修改和补充,为便于专家查阅,特粘贴修改部分如下(正文中的修改见蓝色标记处):

修改稿:

问卷的主要测量学指标如下: 30 个判断题和 30 个 Likert 量表题的内部一致性分别为 0.823 和 0.883, 其中 6 个分问卷的内部一致性信度在 0.64~0.76 之间; 以大学生为样本, 测得问卷总分间隔三周的重测信度为 0.72; 与《心理健康素养量表》(Mental Health Literacy Scale, MHLS)的聚合效度为 0.74; 另外, 问卷总分能有效区分心理健康专业人员和普通公众(Wu et al., manuscript)。由于计算总分时所有题目均转换为 0、1 计分, 因此本研究根据项目反应理论, 通过 Mplus 软件, 采用双参数 Logistic 模型分析问卷的区分度和难度。结果显示, 80% 的题目区分度参数处于(0.5, 2)的合理范围内(杜文久, 周娟, 李洪波, 2013), 88.3% 的题目难度参数在(-3, 3)的合理范围内(曹尚 等, 2016)。可见,《国民心理健康素养问卷》的心理测量学指标较好, 可以用来测量我国国民的心理健康素养。

意见 2: 本文在 SES 指标的合成上, 直接将原始的定序变量转换为标准分, 而后进行主成分分析, 处理不妥。应当注意到, PCA 技术的核心是通过计算数据协方差矩阵的特征值与特征向量, 并以若干较大特征值对应的特征向量组成投影矩阵对原始数据降维。其本质是生成一组原始变量的线性组合, 并以线性组合作为降维后的新变量, 这就要求原始变量之间应当具有可加性。换言之, 传统 PCA 技术处理的对象至少应当是定距变量或可以视为定距变量的变量。本文中这些变量能否简单将其视为定距变量? 如果能, 请作者给出理由, 如果不能, 建议作者将变量重新投影到定距量尺之后再进行处理(例如教育水平可以受教育年限为定距量尺, 收入可以组中点将分组定距化, 职业可以参考 ISEI 等, 具体处理模式可以参考任春荣(2010)的研究, 或范兴华等人(2012)发表在心理学报上的研究)。

回应: 非常感谢审稿专家耐心、细致地指出在进行社会经济地位指标合成时, 数据转换存在的问题, 以及将定序变量重新投影到定距量尺再做处理的建议。

我们参考审稿专家的建议, 尝试将教育水平、职业类型和收入水平转换为定距数据做 PCA, 从而合成 SES。但是, 在数据转换过程中, 我们遇到了诸多问题, 使得原始数据无法可靠地进行转换。经过权衡, 我们最终决定保留原始的转换方法。以下将对此问题做详细论述。

在原稿中, 我们直接将原始的定序变量直接转换为标准分来进行主成分分析, 原因有两点。一方面, 我们参考了周春燕和郭永玉(2013)、周春燕(2018)两篇研究, 将受教育程度、收入水平和职业类型分别赋值 1~6 分、1~6 分和 1~10 分, 转换为标准分后进行主成分分析的做法。另一方面, 我们对教育水平和职业类型的等级划分出于以下考虑和参考。在受教育

水平方面，不同于范兴华、方晓义、刘杨、蔺秀云和袁晓娇(2012)的研究中，采用父母文化程度、父母职业和家庭收入等来衡量青少年的家庭社会经济地位，本研究样本的年龄范围覆盖较广，其中老年群体(60岁及以上)占样本总量的15.3%(1358人)。该群体出生于1960年代之前，其受教育阶段的学制大多与出生于1985年代之后的青年群体(18~35岁)有所不同。对此，若用实际接受过多少年的教育来衡量受教育水平不够恰当(任春荣, 2010)。在职业类型方面，陆学艺(2002)以职业分类为基础，以组织资源、经济资源和文化资源的占有情况为标准，来划分社会阶层的办法，符合当代中国社会的实际情况，即这三种资源的拥有状况，决定着各社会群体在阶层结构中的位置以及个人的综合社会经济地位。因此，我们在数据收集时，对职业类型之一变量的测量即采用了此分类方法，同时增加了“在校学生”这一类别(转换为社会经济地位时排除)。

诚如专家所言，传统PCA技术处理的对象至少应当是定距变量或可以视为定距变量的变量，而本研究中这些变量不能简单视为定距变量。因此，我们采用专家建议的办法，尝试对教育水平、职业类型和收入水平进行转换，具体做法如下。

在受教育水平方面，OECD(1999)教育分类项目中，按照国际教育标准分类(ISCED)对教育水平的划分标准为：ISCED 1在5~7岁开始入学，教育年限一般为4~6年，OECD的教育模式为6年，对应我国的小学阶段；ISCED 2的教育年限一般为2~6年，OECD的教育模式为3年，对应我国的初中阶段；ISCED 3在15或16岁入学，教育年限一般为2~5年，其中ISCED 3A对应我国的普通高中，ISCED 3B/3C对应我国的职业高中/中专/技校；ISCED 4的教育年限为6个月~2年，对应我国高中毕业后的职业培训；ISCED 5的教育年限从2年到6年及以上不等，是高等教育第一阶段，其中ISCED 5A对应我国的本科阶段，ISCED 5B对应我国的专科阶段；ISCED 6的教育年限一般为3年，但多数情况下要超过3年，是高等教育第二阶段，对应我国的研究生阶段(占盛丽, 2009)。因此，结合中国各个教育阶段的实际教育年限，本研究中的教育水平可按受教育年限分别赋值为：初中及以下，8分；中专或高中，11分；大学专科或本科15分；硕士研究生及以上18分。

在职业类型方面，参考PISA(2009)的做法，将职业类型先根据国际标准职业分类编码表(ISCO-88)进行编码，再根据Ganzeboom和Treiman(1996)的国际社会经济地位职业分类索引(ISEI)，转换为ISEI得分。陆学艺(2002)的职业类型划分与ISCO-88和ISEI的具体对应情况如表2所示。结果显示，由于本研究在数据收集阶段对职业类型的测量，采用的是陆学艺(2002)的职业类型划分方法，将原始数据转换为ISCO-88编码时，部分数据面临着“一对多”或“一对无”映射的情况，导致原稿中的1~10职业阶层得分无法准确转换为ISEI得分。仔细研究ISCO-88与ISEI的对应情况，还发现法官(90)、律师(85)、法律专家(82)、医生(88)、牙医(85)等处于职业类型的最高阶层，其得分明显高于政府官员(70)。而这与我国的社会结构和文化背景相差甚远。

表2 陆学艺(2002)职业阶层划分与国际标准职业分类(ISCO-88)、
国际社会经济地位职业分类(ISEI)编码对照表

原稿分值	陆学艺(2002)职业类型	ISCO-88	ISEI
10	国家与社会管理者	1100 立法委员及高级官员	70
9	经理人员	1200 企业经理（大型企业）	68
8	私营企业主	1300 （小企业）总经理	51
7	专业技术人员	2000 专业人才；	70
		3000 技术人员与助理专业人才	54
6	办事人员	3440 海关、税务等政府助理专业人员（包括行政助理、行政公务员、公共管理人员）；	56
		4100 办事人员	45

5	个体工商户	无	无
4	商业服务业员工	4200 客户服务人员； 5000 服务人员、商店及市场销售人员	49 40
3	产业工人	7000 手工业等交易人员； 8000 工厂、机器操作员和装配工； 9300 采矿、建筑、制造业、运输业的劳动者	34 31 23
2	农业劳动者	6000 熟练的农业和渔业工人； 9200 农业、渔业等劳动者	23 16
1	城乡无业、失业、半失业者	无	无

在收入水平方面，我们尝试以组中点将收入水平分组等距化。具体做法是：将“5500 元以下”赋值为 2750，将“5501~13000 元”赋值为 9250，将“13001~21000 元”赋值为 17000，将“21001~32000 元”赋值为 26500，将“32001~60000 元”赋值为 46000，但对于“60000 元以上”，无法使用前述方法赋值。在数据收集阶段，为保证被尽可能确保被调查者作答的准确性，我们采用勾选收入水平范围的办法。但这一做法也使得我们无法将原始数据恰当地转换为定距数据。

综合前述分析，考虑到心理健康素养这一概念不仅仅属于心理学范畴，它还属于社会价值范畴，文明范畴(江光荣 等, 2020)。而陆学艺(2002)以职业分类为基础，以组织资源、经济资源和文化资源的占有情况为标准，来划分社会阶层的办法，贴合当代中国社会结构的实情和发展现状，有独特的社会价值；收入水平则无法进行恰当地转换。

因此，我们仍采用原稿中的社会经济地位指标合成办法，即家庭人均可支配收入的 6 个水平由低到高分别赋值 1~6 分，受教育程度的 4 个水平从低到高分分别赋值 1~4 分，职业阶层的 10 个水平(排除在校学生)从低到高分分别赋值 1~10 分后，将三项指标转换为标准分进行主成分分析。这一做法在最近一项研究中也得到了应用(如，刘广增，张大均，朱政光，李佳佳，陈旭, 2020)。

以上考虑是否恰当，恳请审稿专家进一步批评指正。

意见 3：本文对缺失数据采用了列均值插补法，但该方法会在该数据点形成人为“峰值”，降低总体方差。习惯上，均值插补法通常在没有辅助信息可用或需要插补的数据量较少的情形才使用，如果本文涉及“需要插补的数据量较少”的情况，建议补充报告含有缺失值的样本占总体样本的百分比，或换用热平台插补、最近邻插补等方法进行缺失值插补。

回应：感谢审稿专家提出数据处理部分需要报告缺失值数量或换用其他更恰当的缺失值插补法的建议。

考虑到本研究在国民心理健康素养问卷上的缺失值数量较少，含有缺失值的样本数量在 2~95 之间，占总样本的百分比均不超过 1.1%，因而采用了列均值插补法，来处理国民心理健康素养问卷的缺失值。但是，我们未在文中报告缺失值样本的比例，这一做法确实不妥。因此，我们对方法部分“2.4 数据处理”加以修改，为便于专家查阅，特粘贴修改部分如下（正文中的修改见蓝色标记处）：

修改稿：

2.4 数据处理

缺失值分析的结果显示，在国民心理健康素养问卷 60 个题目上，含有缺失值的样本数量在 2~95 之间，占总样本的百分比均不超过 1.1%，可见需要插补的数据量较少。因此，国民心理健康素养问卷的缺失值处理采用“列均值插补法”，即以样本总体在该题目上的均值来替补。其他变量的缺失值不做特殊处理，数据分析时自动排除。所有数据分析均使用 SPSS

20.0 进行。

意见 4: 本文 3.1 部分“维护和促进自己心理健康的态度和习惯、……四个维度回答正确率分别为……可见, 公众在心理健康促进方面的素养高于心理疾病应对, 自助心理健康素养高于助人心理健康素养”。首先, 态度和习惯类的变量如何谈及“正确”与“错误”。另外, 根据正确率对不同分量表之间得分进行比对的逻辑基础是被试在各个分量表的得分上具有可比性, 这要求各个量表的难度至少是相似的, 而难度相似或分量表之间具有可比性是需要给出相应的证据进行论证的, 不能够当作默认条件。不能简单因为“某个班级的数学平均分比语文平均分高”就得出“该班级的人擅长数学”这种结论。

回应: 非常感谢审稿专家指出我们对态度和习惯类变量进行“正确”与“错误”评判, 以及在给出各分问卷难度是否相似的情况下, 直接对各个分问卷正确率进行比对存在的问题。我们对结果分析与描述欠妥深表歉意, 以下对这两个问题分别展开说明。

(1) 关于对态度和习惯类变量做正误判断的问题。正如我们在意见 1 的 (1) 中所述, 心理健康素养这一概念实际上蕴含着价值评判。高心理健康素养意味着, 个体有充分的科学认识(知识和观念层面); 不歧视且愿意关怀心理疾病患者, 能够在保护好自己的前提下, 为他人提供力所能及的心理帮助(对待他人心理健康与心理疾病的态度和习惯); 重视并积极促进自身心理健康, 能够接纳自己的状态, 必要时愿意寻求心理帮助(自对待自己心理健康与心理疾病的态度和习惯)。因此, 对维护和促进心理健康、应对心理疾病的态度和习惯进行“正确”和“错误”判断, 有着理论合理性。但是, 我们未在文中对此进行详细论述, 实有不妥, 故对这一部分的表述加以修改。

(2) 关于各个分问卷回答正确率的比较问题。在文献梳理阶段, 我们发现卢家楣等(2017)的研究比较了大学生情感素质在不同具体情感上的水平高低, 得出道德、生活情感发展水平较高, 人际、审美、理智情感和情绪智力等相对较低的结论。类似的, 辛志勇和金盛华(2005)在研究新时期大学生的价值取向时, 根据不同价值取向在 Likert 量表得分均值的高低, 得出个人性目标价值、社会性目标价值、超然性目标价值下, 不同取向的相对重要性。

诚如专家所言, 只有在各个分问卷之间难度相似, 或分问卷之间具有可比性的情况下进行比较, 才是恰当做法。我们起初没有考虑到, 卢家楣等人(2017)在比较具体情感的水平时, 是通过将归属于本体性情感维度的生活情感、道德情感、审美情感、理智情感和人际情感的均分, 分别与本体性情感均分进行差异 t 检验, 进而指出某些情感素质较强, 某些情感素质较弱。金盛华、郑建君和辛志勇(2009)的研究则是通过对每个价值观维度下的所有项目求均值, 进行维度间的比较, 从而分析当代中国人价值观的特点。这一做法实际上确保了分量表之间具有可比性。

因此, 我们采用同样的办法, 先将 0、1 计分的心理健康促进素养(即维护和促进自己心理健康的态度和习惯、维护和促进他人心理健康的态度和习惯)、心理疾病应对素养(即应对自己心理疾病的态度和习惯、应对他人心理疾病的态度和习惯)、自助心理健康素养(即维护和促进自己心理健康的态度和习惯、应对自己心理疾病的态度和习惯)、助人心理健康素养(即维护和促进他人心理健康的态度和习惯、应对他人心理疾病的态度和习惯)四个方面的实际得分, 以及心理健康素养总分除以各自的题目数, 确保所有得分范围在 0~1 之间, 即最低为 0 分, 最高为 1 分。然后采用配对 t 检验, 比较心理健康促进素养、心理疾病应对素养、自助心理健康素养、助人心理健康素养四个方面的平均分, 分别与心理健康素养总平均分之间的差异。结果显示, 心理健康素养总均分($M=0.60$)显著高于心理疾病应对素养均分($M=0.57$)和助人心理健康素养均分($M=0.54$), 但显著低于心理健康促进素养($M=0.66$)和自助心理健康素养($M=0.68$)。

基于上述分析, 我们对结果部分“3.1 我国国民的心理健康素养水平”加以修改, 为便于

专家查阅，特粘贴修改部分如下（正文中的修改见蓝色标记处）：

修改稿：

3.1 我国国民的心理健康素养水平

表 3 呈现了我国国民心理健康素养总分及各二级维度的得分情况。心理健康素养总分的偏度系数为-0.58，峰度系数为 0.70，基本符合正态分布。心理健康素养总分均值为 35.81，这个分数相当于百分制的 60 分。由于高心理健康素养意味着，个体有充分的科学认识(知识和观念层面)；不歧视且愿意关怀心理疾病患者，能够在保护好自己的前提下，为他人提供力所能及的心理帮助(对待他人心理健康与心理疾病的态度和习惯)；重视并积极促进自身心理健康，能够接纳自己的状态，必要时愿意寻求心理帮助(自对待自己心理健康与心理疾病的态度和习惯)。因此，将 Likert 量表题目得分转换为 0、1 计分后，进一步分析六个二级维度的回答情况，发现心理健康相关知识和观念、心理疾病相关知识和观念两个维度的回答正确率分别为 65.23%和 56.56%；¹维护和促进自己心理健康的态度和习惯、应对自己心理疾病的态度和习惯、维护和促进他人心理健康的态度和习惯、应对他人心理疾病的态度和习惯四个维度，回答正确率分别为 75.12%、63.79%、58.23%和 51.89%。与心理健康素养总均分($M=0.60$)相比，心理健康促进素养($M=0.66, p<0.001, d=0.39$)、自助心理健康素养相对较高($M=0.68, p<0.001, d=0.51$)，心理疾病应对素养($M=0.57, p<0.001, d=0.22$)、助人心理素养相对较低($M=0.54, p<0.001, d=0.35$)。可见，公众在心理健康促进方面的素养高于心理疾病应对，自助心理健康素养高于助人心理健康素养。这个结果从图 2 可以直观地看出来。值得注意的是，与自杀有关的两个题目——“和有自杀念头的人谈论自杀会增加他们的自杀风险”、“假如身边有人流露出自杀念头，我会劝他不要瞎想”——回答正确率最低，仅为 13.3%和 4.3%，说明我国公众极度缺乏与自杀应对相关的知识。表 3 还有一点值得注意，比较 6 个维度的变差系数，可看出知识和观念类维度的变差系数大于态度和习惯类维度。

意见 5：同样，3.1 部分“表 3 还有一点值得注意，比较 6 个维度的变差系数，可看出知识和观念类维度的变差系数远大于态度和习惯类维度”。在结果报告部分，应当慎重使用“远大于”这类词汇，通常情况下，“a 远大于 b”指 a+b 与 a 近似相等，这要求构成“远大于”关系的两个值之间至少相差若干数量级，而非简单的不足 2 倍关系。

回应：非常感谢审稿专家的耐心指正以及提出慎重使用“远大于”这类词汇的建议。

如专家所言，知识和观念类维度的变差系数与态度和习惯类维度的变差系数相差不足 2 倍，“远大于”之说确实不妥当，也不严谨。因此，我们对结果部分“3.1 我国国民的心理健康素养水平”加以修改，将“远大于”改为“大于”。为便于专家查阅同时缩短篇幅，修改后的内容见意见 4 末尾的蓝色标记部分。

意见 6：表 3 重点报告的是描述性统计结果，文中对其均值并无进行推论的意图，报告 95% 置信区间意义不大，反而可能容易引起读者的困惑，建议删除。

回应：感谢审稿专家提出删除表 3 中 95% CI 列的建议。起初考虑到本文是基于大样本进行的全国性调查，心理健康素养得分情况可以作为我国国民心理健康素养水平的常模，呈现得分的 95% 置信区间，可以为后续比较心理健康素养水平的相关研究提供参考。

但确实如专家所言，这一做法会引起读者的困惑。因此，我们在修改稿中删除了表 3 中 95% CI 这一列的数据。

意见 7：通常情况下，当研究的样本容量很小、测量的变量很多以及总体的效应量很小时，才会使用 ω^2 作为对 η^2 的矫正(有文献指出自变量与样本数之比小于 1:10 的情况下建议使用

¹指维度内各条目通过率的平均值。

ω^2), 对大样本而言, 建议直接报告 η^2 即可。

回应: 感谢审稿专家指出报告 η^2 的建议。为了确保本文的数据呈现方式和内容符合《心理学报》一贯的做法, 我们在数据表格的呈现上, 主要参考了分别在 2017 年和 2018 年发表于《心理学报》的两篇有关情感素质的大样本调查研究(卢家楣, 刘伟, 贺雯, 王俊山, 陈念劬, 解登峰, 2017; 卢家楣, 刘伟, 贺雯, 王俊山, 陈念劬, 解登峰, 雷开春, 2018)。如专家所言, 本研究的自变量与样本数之比超过 1:10, 可以直接报告 η^2 。

因此, 我们采纳专家的建议, 将表 4 中的 ω^2 改为报告 η^2 , 并将结果部分“3.2 我国国民心理健康素养的差异分析”第一段的相关表述“对我国国民心理健康素养的差异进行了考察。 t 检验的效应量以 d 值估计, 方差分析的效应量采用 ω^2 估计”, 改为“对我国国民心理健康素养的差异进行了考察。 t 检验的效应量以 d 值估计, 方差分析的效应量采用 η^2 估计”。

意见 8: 表 6 中, 心理卫生背景对模型 1 的 ΔR^2 未标注括号。

回应: 感谢审稿专家的提醒。我们已重新检查了全文的数据表, 并结合重新分析后的数据结果, 对表 6 进行了规范化。

意见 9: 根据本文所述, 本研究中年龄、GDP、地域分布应当为 3 分类变量, 在回归模型中, 3 分类变量应当编码为 2 个哑变量纳入方程, 为何整体方程中对应的位置只有一个回归系数?

回应: 非常感谢审稿专家指出回归模型中对 3 个三分类变量处理不当的问题。我们将 3 分类等级变量未编码为 2 个哑变量就直接纳入回归方程, 这一做法着实欠妥, 对此深感歉疚。同时, 我们也检查了其余变量, 发现虽然性别、城乡来源是二分类变量, 但原稿在做回归分析时, 同样直接将 1、2 计分(1=男性, 2=女性; 1=城镇, 2=农村)的变量直接纳入了回归方程。

因此, 在修改稿中, 我们将 GDP、地域分布和年龄分别编码为 2 个哑变量后, 将所有连续变量标准化, 重新进行回归分析, 并在数据表中增加了回归系数的标准误。结果发现, 新结果相比于原稿没有发生实质性改变, 仅在部分回归系数和解释贡献率的具体数值和显著性水平上有细微改变。由于 GDP 水平、地域分布、年龄三个变量的编码和分析方式改变, 因而重新分析后的数据变化主要体现在这三个变量上。表 6 的具体变化情况如下。

模型 1 中, GDP 水平的回归系数由不显著, 变为中 GDP 水平的回归系数在 0.001 水平上显著, 高 GDP 水平的回归系数不显著; 地域分布的回归系数由在 0.05 水平上显著, 变为中部城市的回归系数不显著, 东部城市的回归系数在 0.01 水平上显著; 年龄的回归系数由在 0.05 水平上显著, 变为青年和中年的回归系数均在 0.001 水平上显著。

模型 3 中, GDP 水平的回归系数由 0.01 水平上显著, 变为中 GDP 水平的回归系数在 0.001 水平上显著, 高 GDP 水平的回归系数 0.01 水平上显著; 地域分布的回归系数由在 0.001 水平上显著, 变为中部城市的回归系数不显著, 东部城市的回归系数在 0.001 水平上显著; 年龄的回归系数由不显著, 变为青年的回归系数不显著, 中年的回归系数在 0.001 水平上显著。

模型 4 中, GDP 水平的回归系数由不显著, 变为中 GDP 水平的回归系数在 0.001 水平上显著, 高 GDP 水平的回归系数不显著。

模型 5 中, 城市特征(GDP 水平)的 ΔR^2 由 0.000 变为 0.003($p < 0.001$); SES 的 ΔR^2 数值没变, 但显著性水平由 0.01 变为 0.05; GDP 水平的回归系数由不显著, 变为中 GDP 水平的回归系数在 0.01 水平上显著, 高 GDP 水平的回归系数不显著; 年龄的回归系数由在 0.01 水平上显著, 变为青年和中年的回归系数均在 0.001 水平上显著。

模型 6 中, GDP 水平的回归系数由不显著, 变为中 GDP 水平的回归系数在 0.001 水平上显著, 高 GDP 水平的回归系数不显著; 地域分布的回归系数由在 0.01 水平上显著, 变为

中部城市的回归系数在 0.01 水平上显著，东部城市的回归系数在 0.001 水平上显著。

模型 7 中，GDP 水平的回归系数由不显著，变为中 GDP 水平的回归系数不显著，高 GDP 水平的回归系数在 0.01 水平上显著；地域分布的回归系数由不显著，变为中部城市的回归系数不显著，东部城市的回归系数在 0.01 水平上显著；年龄的回归系数由不显著，变为青年的回归系数不显著，中年的回归系数在 0.01 水平上显著。

基于前述分析，我们修改并完善了表 6 的数据及呈现，以及“3.3 我国国民心理健康素养与基本背景信息的相关分析”表 6 和表 7 之间相应的结果描述部分。因只有个别数值发生了细微改变，考虑到篇幅问题，修改后的内容没有在此呈现。

意见 10：回归部分，解释贡献率（ ΔR^2 ）的计算方法为何？

回应：感谢审稿专家的意见。深入思考了审稿专家提出的意见，审稿专家的疑惑应该在于本文仅呈现了回归分析结果的汇总表，那么 ΔR^2 究竟是如何得出的？我们对此回应如下。

回归分析的指标为 R^2 ，又叫决定系数(coefficient of determination)，取值在 0~1 之间，表示回归平方和对总平方和的贡献(张厚粲, 徐建平, 2003)，即：

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (y_i - \bar{y})^2}$$

代表预测变量的变异解释了被预测变量变异的百分之多少(吴翰林, 于宙, 王雪娇, 张清芳, 2020)。本研究采用的是多元分层回归分析，考虑到篇幅问题，“表 6 各变量对心理健康素养总分及各维度得分的回归系数(β)和解释贡献率(ΔR^2)”整合了分层回归分析每一步输出的结果，没能直观体现出 ΔR^2 是如何得来的。实际上，仅纳入第一层变量时，回归模型 R^2 即为第一层变量的解释贡献率(ΔR^2)；纳入第二层变量后回归方程 R^2 增加的部分，即为第二层变量的解释贡献率(ΔR^2)，表示当控制了第一层变量后，第二层变量对预测变量的解释量；依此类推。

诚如专家所问，原稿确实容易给读者造成困惑。因此，我们在结果部分“3.3 我国国民心理健康素养与基本背景信息的相关分析”第一段，结合下述意见 11，增加了对 R^2 的相关说明。为便于专家查阅，特粘贴修改部分如下（正文中的修改见蓝色标记处）：

修改稿：

3.3 我国国民心理健康素养与基本背景信息的相关分析

将心理健康素养总分及各维度得分与社会经济地位、接触频率和熟悉程度做相关分析，结果见表 5。**相关分析的效应量**即为相关系数本身，其中 0.10~0.29 为小效应，0.30~0.49 为中等效应，0.50 及以上为大效应(Cohen, 1988, 1992)。**回归分析的效应量为 $R^2(\Delta R^2)$** ，取值在 0~1 之间，表示回归平方和对总平方和的贡献(张厚粲,徐建平, 2003)，其中 0.01 为小效应量，0.06 为中等效应量，0.14 为大效应量(Wampold & Imel, 2015/2019)。

意见 11：本文 3.3 部分，“仍然是两个知识观念类维度——心理健康相关知识和观念、心理疾病相关知识和观念，解释率分别达到了中等效应量和大效应量水平(6.9%和 8.6%)”。为何 6.9%为中等效应量，8.6%为大效应量？划分标准为何？

回应：感谢审稿专家提出补充效应量大小判定标准的建议。

我们在判定效应量大小的标准时，参考了 Cohen(1992,1988, 引自 Wampold & Imel, 2015/2019)的划分标准。经审稿专家提醒，我们又查阅了相应文献(Wampold & Imel, 2015/2019)，并核查了数据分析结果。结果发现，对 R^2 而言，0.01 为小效应量，0.06 为中等效应量，0.14 为大效应量。仔细分析原稿描述错误的原因，推测误将 Cohen's d 效应量大小判别标准与 R^2 的效应量大小判别标准相混淆了（Cohen's $d=0.80$ 表示大效应量）。我们对此疏漏深感歉疚。

诚如专家指出的那样， R^2 为 6.9% 和 8.6%，实际上均为中等效应量。因此，我们结合意见 9 中重新处理数据后进行回归分析的结果，将原稿中的相关表述改为“仍然是两个知识观念类维度——心理健康相关知识和观念、心理疾病相关知识和观念，解释率均达到了中等效应量(5.9% 和 7.9%; Wampold & Imel, 2015)”，并在相应位置补充了 R^2 效应量大小的判定标准，为免赘述，具体补充内容见意见 10 的修改稿。

意见 12：本文 4.1 部分，“在调查问卷的量尺上，我国国民心理健康素养总分的均值为 35.81 分。若以百分制的思路，人为划定 60 分为及格，则这个分数刚刚达到及格线(36 分)”，本问卷总分 36 分代表何种判定标准？本问卷大于 36 分的意义是什么？能代表该被试处于何种心理健康素养水平？该判断标准是否有外部效标佐证？另外，知识有正误之分，可以谈及“划定及格线”，但态度是否也能如此操作？另外，后文称“无论如何，有一点是明确的：这个数值所代表的国民心理健康素养现状，无论其在本调查问卷量尺上的位置来看，还是从其与健康中国行动目标的差距来看，都存在较大距离”本问卷得出的结果，能否等同于“健康中国行动目标”？也需要进一步解释。

回应：感谢审稿专家的意见。以下我们将针对审稿专家的意见依次回应。

(1) 关于本问卷总分 36 分代表何种判定标准、有何意义、代表被试处于何种心理健康素养水平，以及有无外部校标佐证的问题。我们在描述心理健康素养现状水平时，借鉴通俗意义上的“及格”含义，即考试或学科成绩达到规定的最低标准，一般指得分为 60%，采用人为划定及格线的办法，实属无奈之举。这一做法在卢家楣等人(2017)的研究中也得到了应用。

诚如专家所言，严格来讲这一判断标准没有外部校标佐证，我们再次查阅、对比了同类文献(如，O' Connor & Casey, 2015)，仍无法找到恰当的外在指标。对于我国国民的心理健康素养水平到底如何这一问题，一方面受制于当前心理健康素养领域在国内外的研究现状。由于不同调查在调查工具、划界标准、样本人群等属性上往往存在差别，故对调查结果(即心理健康素养水平)的比较需要谨慎。对此，我们在引言部分有所说明。另一方面，本研究是基于江光荣等人(2020)心理健康素养新概念框架，首次进行的心理健康素养水平调查，因而尚且没有可参考的比较标准。因此，只能通过将心理健康素养总分的均值放在本调查问卷总分的取值范围这一量尺上，进行描述。

但是，这一做法及原稿中的表述确实容易引起读者的困惑。因此，为方便读者理解，我们参考范兴华等(2012)描述流动儿童的社会支持情况时，将社会支持得分与理论均值做比较的办法(社会支持得分为 2.52 ± 0.95 ，低于理论中值 3.00)，在修改稿中增加心理健康素养总分与理论中值的相对水平描述。

(2) 关于态度可否与知识一样，做出正误判断从而“划定及格线”的问题。对此，我们在意见 1 的 (1) 和意见 4 的 (1) 中做了较为详细地回应，这里就不再赘述。此外，我们在计算心理健康素养问卷总分时，将 Likert 量表的 5 级分数转换为类似正误判断的二分类分数，高分方向的两个选项(表示态度和习惯较积极)计 1 分，其他三个选项计 0 分，从而凸显态度和习惯的积极与消极之分，使其能够如知识观念一样，体现出正误之别，从而使心理健康素养总分能够划定及格线。

(3) 关于本问卷得出的结果，能否等同于“健康中国行动目标”的问题。深入思考了审稿专家提出的意见，推测审稿专家的疑惑可能在于本文具体阐述“健康中国行动目标”的内容，也没有说明本调查问卷的内容和结果与“健康中国行动”具体内容和结果之间的关系。因此，在对两者进行比较时，需要解释这一做法的依据。

经审稿专家提醒，我们又仔细研读了相关文件。《“健康中国 2030”规划纲要》指出，要“加大全民心理健康科普宣传力度，提升心理健康素养。加强对抑郁症、焦虑症等常见精神

障碍和心理行为问题的干预，加大对重点人群心理问题早期发现和及时干预力度。加强严重精神障碍患者报告登记和救治救助管理。全面推进精神障碍社区康复服务。提高突发事件心理危机的干预能力和水平。到 2030 年，常见精神障碍防治和心理行为问题识别干预水平显著提高。”但没有提及具体目标。2019 年发布的《国务院关于实施健康中国行动的意见》则提出“到 2022 年和 2030 年，居民心理健康素养水平提升到 20% 和 30%，心理相关疾病发生的上升趋势减缓。”但没有提及衡量的标准是什么。因此，诚如专家所言，我们在原稿中贸然称与“健康中国行动目标”相差较大实属不妥，本调查的结果确实无法与“健康中国行动目标”进行恰当地比较。为了确保科学研究的准确性，我们决定删除与“健康中国行动目标”做比较的相关表述。

基于上述分析，我们对讨论部分“4.1 关于我国国民心理健康素养的基本情况”第一段加以修改，为便于专家查阅，特粘贴讨论的修改部分如下（相应结论的修改未在下文呈现，正文中的修改见蓝色标记处）：

以上考虑是否恰当，请审稿专家进一步批评指正。

修改稿：

4.1 关于我国国民心理健康素养的基本情况

本调查的一个主要目的，是希望得到我国国民心理健康素养现状的一个基本判断。在调查问卷的量尺上，我国国民心理健康素养总分的均值为 35.81 分，**略高于 30 分这一理论中值**。若以百分制的思路，人为划定 60 分为及格，则这个分数刚刚达到及格线(36 分)。因此，可认为我国国民心理健康素养处于一个中偏低的水平。一个很自然的疑问是，这个分数如果放在各国之间比较，会排序在什么位置？遗憾的是，因为不存在国际比较的条件，所以目前无法回答这个问题。

删除内容如下：

但有一点是明确的：这个数值所代表的国民心理健康素养现状，从其在本调查问卷量尺上的位置来看，还是从其与健康中国行动目标的差距来看，都存在较大距离。

意见 13：本文 4.2 部分“本调查表明，社会经济地位对心理健康素养具有首要影响作用……比较表 3 中六个二级维度的变差系数，……再比较表 6 中模型 2~模型 7 的调整 R^2 项，可发现知识观念类维度(模型 2、3)受自变量影响的程度远高于态度习惯类维度(模型 4~7)……一般来说，变差系数较大者可变性较高，较易受外部因素的影响；而模型 2~模型 7 的调整 R^2 项的比较对此又提供了完美的支持……把上面两个发现结合起来，可知社会经济地位对心理健康素养的影响作用，主要是作用于个体心理健康素养的知识观念部分。”

首先，本调查至多只能表明“在本文模型中涉及到的所有变量中，社会经济地位对心理健康素养的解释力最高”，使用“首要影响”可能引起不必要的误会。

其次，变异系数只表示控制量纲影响后变量的离散程度，并不能说“变差系数较大者可变性较高”，更不能进一步推论“较易受外部因素影响”。应当注意到，变异系数的大小取决于两个值——样本均数与样本标准差，变量 a 的变异系数大于变量 b 的变异系数可能是因为 a 本身均值小于 b，而二者标准差相等；即使是二者标准差不相等的情况下，对于一个通过测量得到的随机变量而言，其分数由真实值（真分数）、系统误差、偶然误差和粗差四部分组成，即使在粗差被完全控制并剔除的情况下，最终测得随机变量的方差仍然包括真分数方差、系统误差方差、偶然误差方差三部分，此时，不妨假定 a 与 b 均为测量得到的随机变量，即使 a 的方差大于 b 的方差，也不意味着 a 的真分数方差就一定大于 b 的真分数方差（也就是说不能说明谁的“可变性”更高）；更进一步地，即使二者确实真分数存在差异，a 的真分数方差大于 b 的真分数方差，也可能仅仅是 a 的分布天然较为离散，不能推论称“较易受外部因素影响”。

再次，除非是同一个因变量的一系列模型/嵌套模型，不同模型之间的调整 R^2 不具有比较的意义，回归方程的决定系数只表示回归方程中的全部自变量能够解释因变量的方差比例，不同回归方程的决定系数不同很可能是某些方程遗漏了其他重要变量或者某些因变量采用的测量方法具有较大的偶然误差，不能说回归方程 A 的决定系数大于回归方程 B 的决定系数，因变量 a 受方程中自变量的影响就要强于因变量 b 受方程中自变量的影响，更不能说因变量 a 受自变量的影响要强于因变量 b 受自变量的影响。

最后，即使前述两个次级推论（①社会经济地位对心理健康素养具有首要影响作用，且②是心理健康素养结构中的知识观念类维度较之态度习惯类维度可塑性更高）均成立，也不能当然推论“社会经济地位对心理健康素养的影响作用，主要是作用于个体心理健康素养的知识观念部分”，次级推论成立只能提示存在前述可能，但是社会经济地位对心理健康素养的影响作用是否通过知识观念部分发挥作用需要进一步的研究证据予以说明。

回应：非常感谢审稿专家耐心、细致地指正。针对上述审稿意见，我们依次回应如下。

（1）对于“社会经济地位对心理健康素养具有首要影响作用”的表述问题。我们实际想表达的意思和审稿专家的意见一样，是“在本研究考察的所有变量中，社会经济地位对心理健康素养具有首要影响作用”，但原稿确实存在表述不严谨的问题。

诚如专家所言，本研究考察的变量，并不能涵盖所有能够影响心理健康素养水平的变量。因此，我们按照审稿专家的建议，对表述进行修改。

（2）对于“变差系数较大者可变性较高，较易受外部因素的影响”的推论问题。经审稿专家提醒，我们重新审视了变差系数在本文中的含义。

诚如专家所言，本研究属于横断面水平的调查研究，变差系数实际上代表样本总体的离散程度，即被试间的水平差异，而非个体内水平的“可塑性”。仅从变差系数大小本身，我们只能得出“心理健康素养结构中的知识观念类维度较之态度习惯类维度离散程度更高，即不同个体之间在知识观念类维度上的差异，较之于在态度习惯类维度上的差异更大。因此，我们重新修改了摘要、讨论及结论部分的相关表述(摘要和结论部分的修改内容未在下文呈现，在修改稿中以蓝色标记)。

（3）对于“不同模型之间的调整 R^2 比较”的问题。本研究的回归分析部分，采用同样的四组变量作为预测变量，来考察其对心理健康素养六个二级维度的影响。因而结果仅能反映本研究所有考察的自变量对心理健康素养每个维度的影响情况，通过 R^2 大小简单进行横向比较确实不妥，且易给读者造成误解。因此，我们补充了相关分析的结果来说明社会经济地位与心理健康素养六个二级维度之间的关系，并修改补充了结果描述与讨论部分的相应表述，仅描述对应的结果。

（4）对于“社会经济地位对心理健康素养的影响作用，主要是作用于个体心理健康素养的知识观念部分”这一推论的合理性问题。本研究对心理健康素养的概念界定与测量中，心理健康和心理疾病的知识观念是心理健康素养的一部分，因而社会经济地位对知识观念类维度得分的影响，势必会在心理健康素养总分上有所体现。

但是如专家所言，原稿中“作用于个体心理健康素养的知识观念部分”这一表述，确实暗含了作用机制这一涵义，而本研究尚无法回答作用机制问题。实际上，我们想表达的是“社会经济地位对心理健康素养的影响，主要是体现于个体心理健康素养的知识观念部分”。因此，我们按照审稿专家的建议，对表述进行了修改。

基于上述分析，我们对讨论部分“4.2 关于心理健康素养的影响因素”加以修改，为便于专家查阅，特粘贴修改部分如下（正文中的修改见蓝色标记处）：

修改稿：

4.2 关于心理健康素养的影响因素

在本文模型涉及到的所有变量中，社会经济地位对心理健康素养的解释力最高。较之年

龄、性别以及城乡居民身份等，社会经济地位对心理健康素养的解释力高出数倍。国外研究也有类似发现(Holman, 2014)。从表 5 相关分析的结果可以发现，社会经济地位与心理健康素养知识观念类维度的相关系数达到了中等效应量，与态度习惯类维度的相关系数则为小效应量。

本调查的另一个重要发现，是心理健康素养结构中的知识观念类维度较之态度习惯类维度个体差异更大。比较表 3 中六个二级维度的变差系数，可发现心理健康素养的六个二级维度中，知识观念类维度的变差系数显著大于态度习惯类维度的变差系数(前者平均为 32.03%，后者平均为 13.89%；前者是后者的 2.3 倍)；再结合表 6 中模型 2~模型 7 的调整 R^2 项，可发现，城市特征、基本人口学特征、社会经济地位和心理卫生背景四组自变量，对知识观念类维度(模型 2、3)的解释力接近或达到了大效应量(12.6%和 15.9%)，而对态度习惯类维度(模型 4~7)的解释力为小效应量或刚达到中等效应量(2.7%~6.4%)。可见，所考察的自变量对知识观念类维度的变异解释率高，对态度习惯类维度的变异解释率低。一般来说，变差系数较大者说明离散程度较高，不同个体之间的水平差异较大；而模型 2~模型 7 的调整 R^2 项的效应量大小对此又提供了支持。因此我们可以有把握地做出上述断定。

把上面两个发现结合起来，可知社会经济地位对个体知识观念部分的影响较大，对态度习惯部分的影响较小。这从表 6 中相应的 ΔR^2 项也很容易看出来。如何解释这个作用？一个可能的解释是，社会经济地位中核心的影响成分是个体的教育程度，而教育程度跟心理健康/心理疾病的知识相关较高，跟心理健康/心理疾病的态度和习惯相关较低。而我国普及的国民教育，很可能也是使国民心理健康素养在不同 GDP 水平和地域等方面，仍能均衡发展的重要因素。

上述分析同时也指向另一个结论：与知识观念类维度个体差异较大，态度习惯类维度个体差异较小。而心理健康素养这一概念不仅仅属于心理学范畴，它还属于社会价值范畴，文明范畴(江光荣 等, 2020)。这些都提示着，心理健康素养的知识和观念相较于态度和习惯，可能更易受个体内因素和社会文化等外部因素的影响。国外研究也有类似发现(Gulliver, Griffiths, Christensen, & Brewer, 2012; Morgan et al., 2018; Morgan, Reavley, Ross, Too, & Jorm, 2018; Xu et al., 2018)。这对于心理健康素养提升的实践来说，真是一个坏消息。因为从心理健康促进行动的需要来说，公众表现于态度和习惯的心理健康素养才是更有含金量的目标。

意见 14：抽样框的设定方面，以“小学及相应班级”中的家庭为代表能否真正代表所在区域的普通家庭？应当注意到，子女与父母之间的年龄差通常在一个较小的范围内波动，而小学生的年龄又相对集中，以小学及相应班级为抽样是否会导致年龄出现集中分层的情况。换言之“小学生及其家长”能否代替“初中生家庭”、“高中生家庭”、“大学生家庭”、“已婚未育”与“子女年龄尚不足 6 岁的家庭”，世代效应是否影响了研究结果？

回应：感谢审稿专家指出的以“小学及相应班级”中的家庭为代表能否真正代表所在区域的普通家庭的问题。限于篇幅，我们在原稿中简要说明了数据采集方式，但显然表述地不够明确。对此，我们做以下说明。

在抽样阶段，基于“所抽取的学校及班级的学生家庭能最好地代表所在区域的普通家庭”这一立意准则，选定抽样班级后，由班主任给每位学生发放装有调查问卷的信封，并讲述问卷填写要求。信封的封面印有对被调查者的身份要求：城乡、性别、年龄段。学生需按此要求“寻找 1 位身边的人”进行填写，不局限于父母，还可以是祖父母、兄姐、邻居等。我们以小学班级学生的家庭为调查单位，是考虑到小学生群体学业任务较之初中和高中略轻，小学生相对更“听老师的话”，学生家长也比较配合。在某种意义上，学生充当了我们的调查员。此外，前期的预研究显示，用此方法采集的数据质量较高。大调查取样比较困难，我们尽可能使调查所得的数据可靠、有效。

但如专家所言，在青年群体中，18~30岁者可能调查到的样本较少。因此，若在调查时能够收集被调查者的具体年龄，则可知晓样本在年龄上的确切代表情况，从而在出现取样偏差时及时补充数据。很遗憾本研究未能做到这一点，使本调查样本代表性的说服力稍显欠缺。

对此，我们在讨论部分“4.4 研究的局限和未来研究的展望部分”的第一段补充了此不足之处，为便于专家查阅，特粘贴修改部分如下（正文中的修改见蓝色标记处）：

修改稿：

4.4 研究的局限和未来研究的展望

本研究还存在一些局限。一是原设计的4组自变量对心理健康素养的解释力还不是很高。这很可能意味着影响心理健康素养的因素主要还是个体内变量。对这个可能性在调查设计阶段没有预计到。不过这可以在后续研究中再去探索。二是**未能呈现具体的年龄分布情况，使本调查样本代表性的说服力稍显欠缺，未来调查样本和覆盖性还可以更大更全一些，取样方法还可以更精妙一些。**

意见 15：用词方面，本文以“影响因素”为题，实际却为相关研究，这可能造成不必要的误会。并且，社会人口学变量如何“影响”心理健康素养，其与心理健康素养之间的相关关系究竟是“影响”还只是恰好“耦联”，需要进一步甄别。

回应：感谢审稿专家指出以“影响因素”为题不妥的问题。

本研究中，心理健康素养是指“个体在促进自身及他人心理健康，应对自身及他人心理疾病方面所养成的知识、态度和行为习惯”。从概念来看，心理健康素养属于后天形成的、基于社会建构的、受文化环境等影响的变量。理论来讲，本研究所考察的城市特征、基本人口学特征、社会经济地位和心理卫生背景四组变量，先于个体的心理健康素养水平而存在。正是个体所处的城市发展水平、人文环境，社会赋予其的性别角色，以及既往与心理疾病患者的接触史、对专业心理健康服务的熟悉情况，影响了个体的心理健康素养水平。但是如专家所言，本文只是对社会人口学变量与心理健康素养展开的相关研究，对于影响机制及其他可能存在的影响因素，并未涉及。因而以“影响因素”为题，确实会给读者造成误解。因此，我们参考审稿专家的建议，将本文的题目改为“中国国民心理健康素养的现状与特点”。

意见 16：引言部分第二段“国内学者明至君和陈祉妍(2020)也强调重视心理健康的意识”，作者名字写错，应为“明志君”。

回应：非常感谢审稿专家指出原稿中的疏漏，对于文章中出现的错别字我们深表歉意。我们已在修订稿中改正，并重新检查了全文。

意见 17：虽然本文使用了新的测量工具，发展了概念框架，但本文得出的主要结论缺乏新意，大多是既往研究已经发现的内容，对数据的分析和讨论不足，整体略显粗糙。

回应：感谢审稿专家指出本文结果创新性和分析讨论不足的问题。我们对此做以下回应。

一是研究结果的创新性问题。如审稿专家所言，本研究得出的主要结论与一些既往研究发现的内容相同，似乎并无新意。对此，我们与审稿专家的看法略有不同。一方面，虽然以往对心理健康素养的研究有很多发现，但既往研究结果之间存在很大差异，且因调查对象和测量工具不同，难以进行比较。而国内多数调查的内容如本文引言中提到的，多为单独考察心理疾病知识、心理疾病污名态度或求助行为，未能提供心理健康素养现状的总貌。本研究基于整合的心理健康素养概念框架，以有代表性的大样本调查取样为基础，研究结果可信，能够为当前我国国民的心理健康素养水平提供一个基线参考。此外，研究发现对于心理健康素养提升政策、干预方法和干预内容等，也能够提供科学的指导。这在本文的讨论部分“4.3 关于对国民心理健康素养提升实践的启示”有详细论述，是本研究的一个重要贡献之处。

二是数据的分析与讨论深度问题。本研究的目的是基于心理健康素养新概念框架,对我国国民的心理健康素养现状进行调查。因此,本文要解决的核心问题是,当前我国国民的心理健康素养水平如何,其不同维度的发展情况是否均衡,在不同社会学指标分组见的差异如何。我们要解决的核心问题,决定了数据分析需要以描述性统计、差异检验、相关和回归分析为主。在讨论部分,我们也对统计结果进行了系统地分析和比对,从而得出了有价值的结论。的确,本调查是一项基础性研究,未来还需要在此研究发现的的基础上,进行更深入、更细致地探索,如我们在文末“研究的局限和未来研究的展望”中所言。此外,我们结合前述审稿专家意见,对本文尤其是讨论部分的语言表述进行了修改,以求语言表达更加客观、严谨,研究结论更加可信。

.....

审稿人 2 意见:

本研究以中国国民心理健康素养为研究对象,采用的调查问卷的理论较为新颖,全面,调查样本全面,有助于对心理健康素养的概念和中国国民心理健康素养有进一步的了解。

回应:非常感谢审稿专家对调查问卷的理论基础、调查样本以及研究选题的意义所给予的肯定。

意见 1:由于本次调查都是基于问卷进行,对于调查来说,真实作答很重要,请作者提供问卷真实作答的检验判断方法。如果不能提供,说明如何减少不是认真作答的数据对研究结果的影响。

回应:非常感谢审稿专家对问卷的真实作答方面提出的问题和建议。限于篇幅,我们在原稿中的方法部分,仅简要介绍了调查的抽样、施测方式及问卷的有效回收情况,但这对于确定问卷的真实作答性确实不够。对此,我们做以下补充说明。

由于调查问卷未设置测谎题,故无法直接判断问卷是否真实作答。因此,我们在抽样、施测、废卷筛选三个环节,分别采取相应措施,以尽可能确保问卷的填写质量和作答的真实性。具体做法如下。

抽样阶段,如我们在审稿专家 1 的意见 14 中的回应,以小学班级学生的家庭为调查单位,经过培训的班主任以家庭作业的方式,给每位学生各发一个装有调查知情同意书和调查问卷的信封,由学生充当调查员,按班主任要求寻找符合条件的家人或邻居等填写问卷。调查对象填写完成后装回信封,由学生交给班主任。信封上印有“年龄段 | 性别”来限定被调查对象的身份。此抽样方法可从源头上,确保调查对象在很大程度上能够认真作答。

施测阶段,由学生告知被调查对象,在填写问卷前先阅读《知情同意书》。从而让被调查者详细知晓问卷填写结果的匿名性和保密性,以及亲自认真、如实填写的重要性。

废卷筛选阶段,将出现大量缺失值、规律作答、作答时间过长或过短(网络问卷)、矛盾作答等任何一种情况的问卷,视为无效问卷。具体筛选标准为:(1)心理健康素养问卷缺失值 ≥ 10 ;(2)心理健康素养问卷的后 4 个分问卷共 30 题(Likert 5 点计分),连续选择同一选项的题目数 ≥ 10 ;(3)网络问卷作答时间 ≤ 210 秒或 ≥ 6000 秒;(4)“最近一年居住地”选择“农村”,并且“职业类型”选择“在校学生(包括在读全日制专科/本科生和研究生)”;(5)“年龄”未勾选但自行填写的年龄 < 18 岁。需要说明的是,对于标准(3)中,作答时间的下限按照“3s/题,共 70 题”这一经验来判定,作答时间的上限是综合 9 个城市的情况做出的判定,低于 210 秒或超过 6000 秒极有可能是异常值;而作答时间在 210 秒~6000 秒之间的网络问卷,则结合缺失值数量和作答规律逐一筛查。对于标准(4),由于本调查中的“在校学生”并不包括中学生,而在读全日制专科/本科生和研究生最近一年的居住地通常不可能为农村,因此将符合此标准的问卷判定为问题问卷。

总的来说,我们从数据收集和剔除数据两大方面,力求减少不是认真作答的数据对研究结果的影响。以上考虑是否恰当,请审稿专家进一步批评指正。

意见 2: 另外,由于问卷都是用自我报告的形式,需要用统计方法说明共同方法偏差的影响,证明结果的有效性。

回应: 感谢审稿专家提出用统计方法检查共同方法偏差影响,以证明结果有效性的建议。起初未进行共同方法偏差检验,是考虑到本研究采用的自我主观报告问卷只有《国民心理健康素养问卷》,而《基本背景信息问卷》是用来收集客观信息的。但是经审稿专家提醒,我们意识到,无论是主观报告还是客观填写,本质上均是通过调查对象自我报告获得的数据,有必要进行共同方法偏差检验。

因此,我们将《基本背景信息问卷》中采用等级计分的 5 个题目(教育程度、收入水平、职业类型、接触频率、熟悉程度)和《国民心理健康素养问卷》60 个题目的数据,进行 Harman 单因素检验。根据检验结果,修改了《论文自检报告》的问题 3,并在方法部分的“2.4 数据处理”后增加“2.5 共同方法偏差检验”。为便于专家查阅,特粘贴修改部分如下(正文中的修改见蓝色标记处):

修改稿:

2.5 共同方法偏差检验

为了减少自陈问卷带来的共同方法偏差,本研究在数据收集过程中,通过部分项目反向计分,以及强调匿名、保密等原则进行程序控制(周浩,龙立荣,2004),并采用 Harman 单因素检验法,对数据进行共同方法偏差检验。结果发现,第一个公因子解释的总方差为 11.15%,小于临界值 40%(汤丹丹,温忠麟,2020)。因此,本研究的数据不存在共同方法偏差问题。

最后,再次感谢两位审稿专家的悉心审阅和专业建议。

回应涉及的参考文献:

- 曹尚,曹荣祥,孙昕雯,郭海建,李小宁,徐勤. (2016). 项目反应理论在居民健康素养标准参照测验中的应用研究. *中国卫生统计*, 33(1), 31–34+38.
- 戴海琦. (2006). 基于项目反应理论的测验编制方法研究. *考试研究*, 2(4), 31–44.
- 杜文久,周娟,李洪波. (2013). 二参数逻辑斯蒂模型项目参数的估计精度. *心理学报*, 45(10), 1179–1186.
- 范兴华,方晓义,刘杨,蔺秀云,袁晓娇. (2012). 流动儿童歧视知觉与社会文化适应: 社会支持和社会认同的作用. *心理学报*, 44(5), 647–663.
- Ganzeboom, H. B. G., & Treiman, D. J. (1996). Internationally Comparable Measures of Occupational Status for the 1988 International Standard Classification of Occupations. *Social Science Research*, 25, 201–239.
- 郭庆科,陈英敏,孟庆茂. (2005). 自陈量表式测验应用 IRT 的可行性. *心理学报*, 37(2), 275–279.
- 国务院:《国务院关于实施健康中国行动的意见》,国发[2019]13号文件,2019年6月24日。
- 江光荣,赵春晓,韦辉,于丽霞,李丹阳,林秀彬,任志洪. (2020). 心理健康素养: 内涵、测量与新概念框架. *心理科学*, 43(1), 232–238.
- 金盛华,郑建君,辛志勇. (2009). 当代中国人价值观的结构与特点. *心理学报*, 41(10), 1000–1014.
- 刘广增,张大均,朱政光,李佳佳,陈旭. (2020). 家庭社会经济地位对青少年问题行为的影响: 父母情感温暖和公正世界信念的链式中介作用. *心理发展与教育*, 36(2), 240–248.
- 卢家楣,刘伟,贺雯,王俊山,陈念劬,解登峰,雷开春. (2018). *心理学报*, 50(5): 528–538.
- 卢家楣,刘伟,贺雯,王俊山,陈念劬,解登峰. (2017). *心理学报*, 49(1), 1–16.
- 陆学艺. (2002). 当代中国社会十大阶层分析. *学习与实践*, (3), 55–63.
- O’ Connor, M., & Casey, L. (2015). The mental health literacy scale (MHLS): A new scale-based measure of

- mental health literacy. *Psychiatry Research*, 229(1–2), 511–516.
- OECD. (1999). “Classifying Educational Programmes”, *Manual for ISCED-97 Implementation in OECD Countries*. OECD Publishing.
- OECD. (2012). PISA 2009 Technical Report. PISA, OECD Publishing.
- 任春荣. (2010). 学生家庭社会经济地位(SES)的测量技术. *教育学报*, 06(5), 77–82.
- 汤丹丹, 温忠麟. (2020). 共同方法偏差检验: 问题与建议. *心理科学*, 43(1), 215–223.
- Wampold, B. E., & Imel, Z. E. (2015). *The Great Psychotherapy Debate: The Evidence for What Makes Psychotherapy Work (second edition)*. New York: Routledge.
- 吴翰林, 于宙, 王雪娇, 张清芳. (2020). 语言能力的老化机制: 语言特异性与非特异性因素的共同作用. *心理学报*, 52(5), 541–561.
- Wu, J., Wang, C. X., Zhu, X., Zhang, L. K., Li, Y. Q., Liu, G. D., & Jiang, G. R. (manuscript). Development and Validation of the Mental Health Literacy Questionnaire for Chinese Adults.
- 辛志勇, 金盛华. (2005). 新时期大学生价值取向与价值观教育. *教育研究*, (10), 22–27.
- 占盛丽. (2009). 从个人和学校视角看家庭社会经济地位对学生学业成绩的影响——国际学生评估项目(PISA)的启示. *上海教育科研*, 12, 10–13.
- 张厚粲, 徐建平. (编). (2003). *现代心理与教育统计学*. 北京: 北京师范大学出版社.
- 中共中央、国务院: 《“健康中国 2030”规划纲要》, 2016 年 10 月 25 日。
- 周春燕, 郭永玉. (2013). 家庭社会阶层对大学生心理健康的影响: 公正世界信念的中介作用. *中国临床心理学杂志*, 21(4), 636–640.
- 周春燕, 万丽君, 宋静静, 黄海, 李林, 刘陈陵. (2018). 家庭社会经济地位与儿童自尊的关系: 父母卷入的中介作用. *中国临床心理学杂志*, 26(6), 1186–1190.
- 周浩, 龙立荣. (2004). 共同方法偏差的统计检验与控制方法. *心理科学进展*, 12(06), 942–950.

第二轮

审稿人 1 意见:

意见 1: 对于第一轮审稿意见 1 的回应中, “实际上, 该问卷在编制过程中分别考察了 30 个判断题和 30 个 Likert 量表题的内部一致性, 分别为 0.823 和 0.883, 在可接受范围内”, 既然分了 6 个分问卷, 为什么后面强调判断题内部和 Likert 量表内部的一致性? 分问卷/分量表和某量表的分维度是不一样的, 既然说是分问卷, 那么若干个问卷组成的组合问卷的内部一致性意义是什么? 另外, 文末所附问卷显示, 知识类题目是有正确、错误和不知道三种可能的, 那么不知道和错误回答是否等价? 建议补充说明此部分。此外, “有 5 个题目(8.3%)的区分度过低(<0.3 ; 曹尚 等, 2016)”, 针对这部分区分度不足的题目, 为何要留下? “53 个题目(88.3%)的难度参数在 $(-3, 3)$ 的合理范围内”, 余下 7 题呢?

回应: 感谢审稿专家对问卷的内部一致性、知识类题目回答的等价性, 以及个别题目的区分度和难度提出的疑问和建议。以下我们将对这三个问题分别展开说明。

(1) 关于问卷的内部一致性问题。由于本问卷的题型包括判断题和 Likert 量表题两种, 使得无法直接报告问卷 60 个题目的内部一致性。鉴于判断题考察的是心理健康素养的知识和观念方面, Likert 量表题考察的是心理健康素养的态度和行为习惯方面, 因此我们分别报告了这两部分的内部一致性信度, 以期对问卷整体的内部一致性提供参考。

但确如审稿专家所言, 我们已报告了 6 个分问卷的内部一致性, 再强调判断题内部和 Likert 量表题内部的一致性(即报告若干个分问卷组成的组合问卷的内部一致性), 这一做法确实欠妥。一致性信度主要反映的是测验内部题目之间的信度关系, 考察测验的各个题目

是否测量了相同的内容或特质(Gregory, 2007/2013, p. 91)。因此, 我们修改了《国民心理健康素养问卷》测量学指标的报告, 删除了第一轮修改稿中判断题和 Likert 量表题的内部一致性结果, 重点说明题目的区分度、难度、分问卷的内部一致性信度、问卷的重测信度、聚合效度, 以及对心理健康专业人员和普通公众的区分能力。

(2) 关于知识类题目中不知道和错误回答的等价性问题。由于这些题目考察的都是有关心理健康和心理疾病的知识, 被调查者无论是回答错误, 还是回答不知道, 都说明其未掌握相应的知识, 因而将这两种情况均计为 0 分。原稿限于篇幅未详细说明, 我们在此轮修改稿中, 按照审稿专家的建议做了补充说明。

(3) 关于将区分度不足的题目保留下来的原因, 以及未报告难度的题目。我们将区分度小于 0.3 的 5 个题目保留下来, 是出于两方面的考虑。一是本文报告的区分度和难度, 是基于本调查收集到的数据分析得来, 而非问卷作者在问卷编制过程中得到的结果, 因而在施测前, 我们尚不知道哪些题目区分度高, 哪些题目区分度不足。二是从测量内容来看, 这 5 个题目测量的内容(“长期过量饮酒是心理疾病”; “缺乏同情心、非常冲动等性格问题也属于心理疾病”; “我认为自己的心理健康是最重要的”; “假如身边有人流露出自杀念头, 我会劝他不要瞎想”; “当服用精神科药物出现副作用时, 我会自己减量或停止服药”)与心理健康素养紧密相关, 且是其他题目不可替代的, 有其独特价值。因此, 我们在数据分析和结果报告时, 保留了上述题目。

对于题目难度, 我们在第一轮修改稿的意见回应中, 仅报告了难度参数在合理范围内的题目占比, 这一做法确实容易造成困惑。因此, 我们在此做补充说明, 其余 7 个题目的难度参数值如表 1 所示。同样, 尽管这 7 个题目的难度参数超出了(-3, 3)这一合理范围, 但从其测量内容来看, 有必要在数据分析和结果报告时保留。

表 1 难度超出(-3, 3)的题目及其难度参数

题目	难度
12. 心理疾病越早治疗效果越好。	-3.037
18. 长期过量饮酒是心理疾病。	5.497
23. 和有自杀念头的人谈论自杀会增加他们的自杀风险。	4.518
29. 情绪稳定是心理健康的表现。	-3.201
32. 我认为自己的心理健康是最重要的。	-7.283
49. 我会把心理疾病患者和我说的话当作笑话和其他人讲。	-30.620
53. 当服用精神科药物出现副作用时, 我会自己减量或停止服药。	4.762

综合上述分析, 我们在第二轮修改稿的“2.2.1 《国民心理健康素养问卷》”中做了如下修改: (1) 在第一段的问卷计分方法部分增加: “因为知识观念类题目无论回答错误, 还是回答“不知道”, 均意味着被调查者未掌握相应的知识, 所以”; (2) 删除第二段的“30 个判断题和 30 个 Likert 量表题的内部一致性分别为 0.823 和 0.883, 其中”。

意见 2: 对于第一轮审稿意见 2 的回应中, “因此, 我们仍采用原稿中的社会经济地位指标合成办法, 即家庭人均可支配年收入的 6 个水平由低到高分别赋值 1~6 分, 受教育程度的 4 个水平从低到高分别赋值 1~4 分, 职业阶层的 10 个水平(排除在校学生)从低到高分别赋值 1~10 分后, 将三项指标转换为标准分进行主成分分析。这一做法在最近一项研究中也得到了应用(如, 刘广增, 张大均, 朱政光, 李佳佳, 陈旭, 2020)”, (1) PCA 技术实际上等价于对数据的协方差矩阵进行谱分解, 而后丢弃贡献较小的部分, 而后以某一组变量的线性组合作为新变量, 诚然对任何的输入, PCA 总能强行给出一个输出, 但如果输入变量不是严格等距, PCA 很可能得出不恰当的结果。(2) 在研究中, 使用技术应当注意技术背后的基本原理及物理含义, 如若变量不构成定距, 则其线性组合构成的变量不具有意义, 这给将来的解

释和分析带来了麻烦。(3) 在这里使用 PCA 技术的主要目的是降维, 降维的方式不止 PCA 一种(虽然 PCA 本身操作比较简单), 对于非等距变量, 可以考虑尝试一些分类算法对数据进行。如果从物理角度而言, 甚至可以考虑将 SES 视为自变量, 将其他变量作为因变量而后拟合有序多分类 Logistic 回归的模型。(4) 如果实在要使用 PCA, 建议补充论证 SES 分指标(家庭人均可支配年收入, 受教育程度, 职业阶层)的等距或近似等距。

回应: 感谢审稿专家对此问题的深入剖析和建议。

如专家所言, 对于任何输入, PCA 总能给出一个输出。而对于输入的数据类型, 我们认为以定距数据最为合适, 但并非必须是严格的定距变量。比如使用 Likert 量表来测量的心理学变量, 其纳入 PCA 分析的原始数据实际上也是等级变量, 而非严格意义的定距变量。然而, 尽管 Likert 量表的测量实际上是序次测量(ordinal measure), 但在 CFA 及其他统计模型分析中, 通常将有 5 个及更多点的序次测量作为连续变量分析(王济川, 王小倩, 姜宝法, 2011, p. 41)。而本研究中的家庭人均可支配年收入水平和职业阶层取值分别为 1~6 分和 1~10 分, 因此在统计模型分析中可以将其作为连续变量来分析。诚然, 受教育程度为 1~4 计分, 小于 5 个计分点, 是一个不足之处, 但其数据类型本质上与家庭人均可支配年收入水平和职业阶层相同。

对于审稿专家提出的将 SES 视为自变量, 将其他变量作为因变量, 从而拟合有序多分类 Logistic 回归模型的建议, 我们考虑再三后没有采纳。原因在于, 我们在将受教育程度、家庭人均可支配年收入水平、职业阶层三个变量降维, 以折算为一个新指标社会地位(SES)之前, SES 这一变量取值为何, 在本研究中尚未知晓, 因而无法将其视为自变量纳入模型。

因此, 我们当前仍保留初始的分析方法, 后续研究或可采用更为直接的社会经济地位指标。再次感谢审稿专家的意见和解析。

意见 3: 对于第一轮审稿意见 4 的回应, “因此, 我们采用同样的办法, 先将 0、1 计分的心理健康促进素养(即维护和促进自己心理健康的态度和习惯、维护和促进他人心理健康的态度和习惯)、心理疾病应对素养(即应对自己心理疾病的态度和习惯、应对他人心理疾病的态度和习惯)、自助心理健康素养(即维护和促进自己心理健康的态度和习惯、应对自己心理疾病的态度和习惯)、助人心理健康素养(即维护和促进他人心理健康的态度和习惯、应对他人心理疾病的态度和习惯)四个方面的实际得分, 以及心理健康素养总分除以各自的题目数, 确保所有得分范围在 0~1 之间, 即最低为 0 分, 最高为 1 分。然后采用配对 t 检验, 比较心理健康促进素养、心理疾病应对素养、自助心理健康素养、助人心理健康素养四个方面的平均分, 分别与心理健康素养总平均分之间的差异。结果显示, 心理健康素养总均分($M=0.60$)显著高于心理疾病应对素养均分($M=0.57$)和助人心理健康素养均分($M=0.54$), 但显著低于心理健康促进素养($M=0.66$)和自助心理健康素养($M=0.68$)”, 这样做还是无法避免因为分量表之间难度不同而导致的一系列问题, 做差比较的基准必须是二者具有相同的零点, 如果不满足, 则无法进行推论。

回应: 感谢审稿专家的进一步指正。对于审稿专家提出的分量表之间难度不同, 没有相同的零点, 使得无法做差比较的问题, 我们有不同看法。

虽然本研究中的知识和观念为判断题, 且采用 0、1 计分, 态度和习惯从积极和消极的角度也对其做正误判断, 且采用 0、1 计分, 但这些题目本质上与认知能力测验的题目不同, 不存在必须难度相同方可比较的问题。具体来说, 个体在某个题目回答错误或选择消极选项, 则意味着个体尚未具备该题目代表的相关知识和观念或态度和习惯, 但是这与诸如语文、数学、智力等认知能力测验中, 严格采用不同难度等级的题目来测量同一种认知能力的意义不同。比如本调查所用问卷中, 不同难度参数的两个题目“心理疾病越早治疗效果越好”(-3.037)

和“长期过量饮酒是心理疾病”(5.497)之间的难度差别所代表的意义,与韦氏智力测验算术分测验中,不同难度等级的两个题目“拿一元去买六角钱的糖,还可找回多少?”和“6角钱一尺布,3元6角钱可买几尺?”之间的难度差别所代表的意义,即完全不同。

因此,我们在第二轮修改稿中,保留了第一轮修改稿中参考前人研究做法(金盛华,郑建君,辛志勇,2009;卢家楣等,2017)进行的分析过程和结果。

意见 4: 对于第一轮审稿意见 6 的回应,“感谢审稿专家提出删除表 3 中 95% *CI* 列的建议。起初考虑到本文是基于大样本进行的全国性调查,心理健康素养得分情况可以作为我国国民心理健康素养水平的常模,呈现得分的 95%置信区间,可以为后续比较心理健康素养水平的相关研究提供参考。”请注意心理测量学中对“常模”概念的定义,平均数的 95%置信区间如何构成“常模”?

回应: 非常感谢审稿专家指出我们对常模含义的误解。我们重温了常模的相关知识和置信区间的含义,认识到平均数的 95%*CI* 仅表示总体参数有 95%的可能性落在这一范围,并不能构成常模,因而最初的想法是不正确的。由于第一轮修改稿已删除了相关数据表中的 95%*CI*,故此次正文不再做其他修改。

意见 5: 对于第一轮审稿意见 10 的回应,“代表预测变量的变异解释了被预测变量变异的百分之多少(吴翰林,于宙,王雪娇,张清芳,2020)。本研究采用的是多元分层回归分析,考虑到篇幅问题,“表 6 各变量对心理健康素养总分及各维度得分的回归系数(β)和解释贡献率(ΔR^2)”整合了分层回归分析每一步输出的结果,没能直观体现出 ΔR^2 是如何得来的。实际上,仅纳入第一层变量时,回归模型 R^2 即为第一层变量的解释贡献率(ΔR^2);纳入第二层变量后回归方程 R^2 增加的部分,即为第二层变量的解释贡献率(ΔR^2),表示当控制了第一层变量后,第二层变量对预测变量的解释量;依此类推。”想问的就是分层回归的第一层和第二层是如何设置的,如果是分层回归,应当给出每一层的结果,以及方程之间的 ΔR^2 ,至少也应说明第一层模型的型态。

回应: 感谢审稿专家的进一步提问,第一轮意见回应中我们没有对此解释清楚,感到抱歉。我们在此做补充说明。

为了知晓每个预测变量对因变量的解释贡献率和回归系数,以及每组预测变量对因变量的合共解释贡献率,我们对每个因变量分别进行两次分层回归分析。第一次,将城市特征、基本人口学特征、SES、心理卫生背景四组预测变量,分别纳入回归模型的第一层、第二层、第三层、第四层,以得出每组预测变量对因变量的合共解释率。第二次,每层仅纳入一个预测变量,直至所有预测变量均纳入到回归模型,以得出每个预测变量对因变量的解释贡献率和回归系数。需要说明的是,以上预测变量的选取均根据差异分析和相关分析的结果,纳入差异分析和相关分析中结果显著的变量。

考虑到篇幅限制和数据结果呈现的清晰度,我们在正文中保留合并整理后的数据表,但将每个分层回归模型的输出结果以附表的形式,补充在附录部分。为避免篇幅过长,此处不再呈现补充的附表,但在第二轮修改稿文末以红色标记呈现。同时,我们再次核查了数据结果,发现表 6 中有一处数据录入错误,在正文中已修改并用红色标记。

意见 6: 对于第一轮审稿意见 12 的修改稿,“在调查问卷的量尺上,我国国民心理健康素养总分的均值为 35.81 分,略高于 30 分这一理论中值。若以百分制的思路,人为划定 60 分为及格,则这个分数刚刚达到及格线(36 分)。因此,可认为我国国民心理健康素养处于一个中偏低的水平。”确实,“理论中值”可以构成比较对象,但理论中值的理论意义为何?达到理论中值意味着国民心理健康素养处于一个什么水平?如果能解释,请补充解释该标准的具体

含义，如果不能解释，慎重做推论。

回应：感谢审稿专家对心理健康素养总分意义的进一步提问。达到理论中值，意味着国民心理健康素养水平在本调查问卷的量尺上，处于中等水平。而按照通俗意义上的“优良中差”划分标准，“及格线”（百分制的 60 分，本调查问卷中的 36 分）则往往意味着较低水平或表现较差。因此，结合“理论中值”和“及格线”两个参考标准，可认为心理健康素养总分的均值为 35.81 分，意味着我国国民心理健康素养处于一个中偏低的水平。

意见 7：对于第一轮审稿意见 13 的修改稿，首先，“可以推测，表 6 中心理健康素养总分被自变量解释的效应量，大部分来自知识观念类维度。”做如此推测的依据是什么？最好给出分析。其次，“把上面两个发现结合起来，可知社会经济地位对心理健康素养的影响，可能主要是体现于个体心理健康素养的知识观念部分。这从表 6 中相应的 ΔR^2 项也很容易看出来。”参考上一点，请解释该条推论。依据什么从 ΔR^2 就能看出来 SES 对 MHL 的影响主要体现于其中的知识观念部分？最后，“这些都提示着，心理健康素养的知识和观念相较于态度和习惯，更易受个体内因素和社会文化等外部因素的影响。”请考虑这个结果提示的可靠程度，不确定的推论请加上“可能”，请明确区分本文做出的结果和基于此结果联系他人研究进行的解释。

回应：感谢审稿专家提出补充推测依据的建议以及指出本文表述中存在的问题。

经审稿专家提醒，我们再次仔细思考了原稿和修改稿中的相关表述，意识到对于“心理健康素养总分被自变量解释的效应量，大部分来自知识观念类维度”和“社会经济地位对心理健康素养的影响，可能主要是体现于个体心理健康素养的知识观念部分”这两句话，很容易让读者产生“心理健康素养总分被解释的效应量，即为各个方面被解释的效应量之和”的误解，但实际上并非如此。本研究所得结果，只能说明“知识观念类维度的变异被所考察的自变量解释的比例高，态度习惯类维度的变异被所考察的自变量解释的比例低”，且“社会经济地位对个体知识观念部分的影响较大，对态度习惯部分的影响较小。”而对于“心理健康素养的知识和观念相较于态度和习惯，更易受个体内因素和社会文化等外部因素的影响”这一观点，确如专家所言，是结合本研究结果及他人研究进行的解释和推测。因此，遵循审稿专家的建议，在表述上增加“可能”两字。

综合前述分析，我们将摘要部分的“社会经济地位是所考察变量中效应最大的因素，且其影响主要体现在素养的知识观念方面”，改为“社会经济地位是所考察变量中效应最大的因素，且其对素养的知识观念方面影响较大，对素养的态度习惯方面影响较小”（英文摘要也相应做了修改，见红色标记）；将结论部分第（5）条中“社会经济地位和其他影响因素对心理健康素养的影响，主要体现在知识观念方面”，改为“社会经济地位和其他影响因素对心理健康素养的知识观念方面影响较大，对态度习惯方面影响较小”；对讨论部分“4.2 关于心理健康素养的影响因素”第二段、第三段、第四段的相关表述进行修改，为便于专家查阅，特粘贴修改部分如下（正文中的修改见红色标记处，摘要和结论的修改见文章相应部分的红色标记，未在下文呈现）：

修改稿：

4.2 关于心理健康素养的影响因素

在本文模型涉及到的所有变量中，社会经济地位对心理健康素养的解释力最高。较之年龄、性别以及城乡居民身份等，社会经济地位对心理健康素养的解释力高出数倍。国外研究也有类似发现(Holman, 2014)。从表 5 相关分析的结果可以发现，社会经济地位与心理健康素养知识观念类维度的相关系数达到了中等效应量，与态度习惯类维度的相关系数则为小效应量。

本调查的另一个重要发现，是心理健康素养结构中的知识观念类维度较之态度习惯类维

度个体差异更大。比较表 3 中六个二级维度的变差系数,可发现心理健康素养的六个二级维度中,知识观念类维度的变差系数显著大于态度习惯类维度的变差系数(前者平均为 32.03%,后者平均为 13.89%;前者是后者的 2.3 倍);再结合表 6 中模型 2~模型 7 的调整 R^2 项,可发现,城市特征、基本人口学特征、社会经济地位和心理卫生背景四组自变量,对知识观念类维度(模型 2、3)的解释力接近或达到了大效应量(12.6%和 15.9%),而对态度习惯类维度(模型 4~7)的解释力为小效应量或刚达到中等效应量(2.7%~6.4%)。可见,所考察的自变量对知识观念类维度的变异解释率高,对态度习惯类维度的变异解释率低。一般来说,变差系数较大者说明离散程度较高,不同个体之间的水平差异较大;而模型 2~模型 7 的调整 R^2 项的效应量大小对此又提供了支持。因此我们可以有把握地做出上述断定。

把上面两个发现结合起来,可知社会经济地位对个体知识观念部分的影响较大,对态度习惯部分的影响较小。这从表 6 中相应的 ΔR^2 项也很容易看出来。如何解释这个作用?一个可能的解释是,社会经济地位中核心的影响成分是个体的教育程度,而教育程度跟心理健康/心理疾病的知识相关较高,跟心理健康/心理疾病的态度和习惯相关较低。而我国普及的国民教育,很可能也是使国民心理健康素养在不同 GDP 水平和地域等方面,仍能均衡发展的重要因素。

上述分析同时也指向另一个结论:与知识观念类维度个体差异较大,态度习惯类维度个体差异较小。而心理健康素养这一概念不仅仅属于心理学范畴,它还属于社会价值范畴,文明范畴(江光荣 等, 2020)。这些都提示着,心理健康素养的知识和观念相较于态度和习惯,可能更易受个体内因素和社会文化等外部因素的影响。国外研究也有类似发现(Gulliver, Griffiths, Christensen, & Brewer, 2012; Morgan et al., 2018; Morgan, Reavley, Ross, Too, & Jorm, 2018; Xu et al., 2018)。这对于心理健康素养提升的实践来说,真是一个坏消息。因为从心理健康促进行动的需要来说,公众表现于态度和习惯的心理健康素养才是更有含金量的目标。

意见 8: 对于第一轮审稿意见 17 的回应,“一是研究结果的创新性问题。如审稿专家所言,本研究得出的主要结论与一些既往研究发现的内容相同,似乎并无新意。对此,我们与审稿专家的看法略有不同。一方面,虽然以往对心理健康素养的研究有很多发现,但既往研究结果之间存在很大差异,且因调查对象和测量工具不同,难以进行比较。而国内多数调查的内容如本文引言中提到的,多为单独考察心理疾病知识、心理疾病污名态度或求助行为,未能提供心理健康素养现状的总貌。本研究基于整合的心理健康素养概念框架,以有代表性的大样本调查取样为基础,研究结果可信,能够为当前我国国民的心理健康素养水平提供一个基线参考。此外,研究发现对于心理健康素养提升政策、干预方法和干预内容等,也能够提供科学的指导。这在本文的讨论部分‘4.3 关于对国民心理健康素养提升实践的启示’有详细论述,是本研究的一个重要贡献之处。”但仍然需要注意的是,该研究仅仅是一个基于问卷的全国部分城市的特殊群体的调查结果,样本量确实够大,但在抽样框设计、问卷作答控制等因素都略有欠缺。另外,对于“此外,我们结合前述审稿专家意见,对本文尤其是讨论部分的语言表述进行了修改,以求语言表达更加客观、严谨,研究结论更加可信。”应当注意在结论中写明哪些是本文的发现,哪些是结合他人研究进行的推论,写明的本文发现应当与本研究主题有关,不要有过度推论之嫌。

回应: 感谢审稿专家对本调查的抽样框设计、问卷作答控制等略有欠缺的提醒,以及区分本文的发现与推论的建议。

如审稿专家所言,本研究只是基于大规模样本的问卷调查。尽管我们在数据收集的各个环节都采取了相应措施,以确保调查样本的代表性和所收数据的可靠性,但调查结果确实无法与我国国民的心理健康素养水平完全划等号。有一点需要说明的是,虽然本调查是基于全国部分城市,但这些城市的选择综合考虑了地域分布和人均 GDP 水平,并且每个城市内部

的样本涵盖了不同年龄段、不同居住地，以及各行各业的人士，所得的调查结果具有一定的参考价值。诚然，本调查只是对我国国民心理健康素养现状与特点的一个初步探索，未来还需要在更广范围内开展更加全面、更加精细的研究。

对于研究结论的呈现，我们综合前述所有审稿建议，逐一审视了结论的可靠性和与本研究主题的关联。本研究的主题是考察我国国民心理健康素养的现状与特点，因而结论（1）和（3）根据描述性统计和差异检验结果，呈现了我国国民心理健康素养的现状水平及其在结构发展上的特点；结论（2）和（4）根据描述性统计、差异检验和回归分析结果，呈现了所考察的自变量对心理健康素养的影响特点；结论（5）中“**知识观念类维度个体差异较大，态度习惯类维度个体差异较小**。社会经济地位和其他影响因素对心理健康素养的影响，主要体现在知识观念方面”，根据描述性统计和回归分析结果，呈现了心理健康素养的结构发展特点，但“个体心理健康素养中，知识观念部分与态度习惯部分存在分离的现象”确如审稿专家所言，是结合他人研究进行的推论，而“社会经济地位和其他影响因素对心理健康素养的影响，主要体现在知识观念方面”表述存在问题。对于后者，我们已在意见 7 中做了相应修改，改为“社会经济地位和其他影响因素对心理健康素养的知识观念方面影响较大，对态度习惯方面影响较小”，此处不再赘述。

对于前者，我们结合本研究的结果仔细分析这一推论，发现心理健康素养的知识观念类维度和态度习惯类维度之间的相关分析结果，也可对此推论提供一些支持，但原稿中未呈现这一结果。因此，我们在表 5 中增加了心理健康素养六个二级维度之间的相关系数（如表 5 所示）。可以看到，心理健康的知识和观念与心理疾病的知识和观念之间的相关系数为 0.584，为大效应量；维护和促进自己心理健康的态度和习惯、应对自己心理疾病的态度和习惯、维护和促进他人心理健康的态度和习惯、应对自己心理疾病的态度和习惯之间的相关系数在 0.350~0.486 之间，为中等效应量。然而，两个知识观念类维度与四个态度习惯类维度之间的相关系数在 0.143~0.241 之间，虽然为正相关，但仅为小效应量。可见，心理健康素养的知识和观念，与态度和习惯之间虽然有正向关联，知识观念水平越高，态度习惯也越积极，但两者之间的关联较小。这也提示我们，个体在心理健康素养的知识观念方面，可能与态度习惯方面存在一定程度的分离。

综上，我们补充了表 5 的内容，并删除了“5 结论”中的推论部分。为便于专家查阅，特粘贴修改部分如下（正文和中、英文摘要部分的修改见红色标记处）：

修改稿：

表 5 心理健康素养总分及各维度与社会经济地位、接触频率和熟悉程度的相关系数

	<i>M</i>	<i>SD</i>	<i>SES</i>	熟悉程度	接触频率	总分	维度 1	维度 2	维度 3	维度 4	维度 5	维度 6
SES	0.004	1.00	1									
熟悉程度	2.22	1.56	0.145***	1								
接触频率	1.59	1.16	0.065***	0.380**	1							
总分	35.81	8.06	0.308***	0.207***	0.059***	1						
维度 1	5.87	2.08	0.331***	0.113***	0.036**	0.644***	1					
维度 2	11.88	3.39	0.377***	0.163***	0.080***	0.779***	0.584***	1				
维度 3	19.79	2.80	0.152***	0.160***	-0.003	0.493***	0.145***	0.214***	1			
维度 4	29.70	3.98	0.033**	0.144***	0.002	0.536***	0.143***	0.186***	0.430***	1		
维度 5	21.35	3.22	0.140***	0.137***	0.016	0.539***	0.193***	0.241***	0.407***	0.404***	1	
维度 6	36.86	4.72	0.086***	0.142***	0.057***	0.573***	0.180***	0.236***	0.350***	0.472***	0.486***	1

注：* $p<0.05$ ，** $p<0.01$ ，*** $p<0.001$ ；黑体表示达到中等或大效应量。

删除内容：

“知识观念部分与态度习惯部分存在分离的现象，且”。

意见 9: 回应涉及的参考文献中,“汤丹丹, 温中麟. (2020). 共同方法偏差检验: 问题与建议. *心理科学*, 43(1): 215–223.” 温忠麟, 名字又写错了。

回应: 非常感谢审稿专家的细心审阅! 对于再次出现的低级错误, 我们实在惭愧。我们已在修改稿中改正, 并请课题组其他成员通读全文, 避免再有此类疏漏。

意见 10: 正文“4.3 关于对国民心理健康素养提升实践的启示”中提到, “二是心理健康素养结构中, 容易受干预而改变的是知识观念, 不易改变的是态度和习惯。”凭什么说这是调查结果的发现? 调查是否进行了干预? 再者, “知识观念类素养与态度习惯类素养存在分离或脱节现象。这意味着增加知识不一定导致态度的积极变化。”为什么这么说, 这是调查显示的结果么?

回应: 感谢审稿专家指出上述两处表述存在的问题。

如审稿专家所言, 本研究只是一项调查研究, 并未进行干预实验。“知识观念易受干预而改变, 态度和习惯不易改变”这一观点, 是结合他人研究得到的推论。因此, 当前的表述确实不妥, 容易造成误解。对此, 我们补充了参考文献, 并修改了相关表述。

而对于“本调查提示, 个体心理健康素养中, 知识观念类素养与态度习惯类素养存在分离或脱节的现象。这意味着增加知识不一定导致态度的积极变化”, 我们在意见 8 的回应中补充了知识观念类维度与态度习惯类维度之间的相关分析结果。并且经审稿专家提醒, 我们意识到, 这一观点实则是结合他人研究结果得出的推论, 而当前的表述是以结论的形式呈现出来的, 这一做法实有不妥。因此, 我们修改了相关表述, 并删除了结论中“知识观念部分与态度习惯部分存在分离的现象”这一说法(如意见 8 的回应, 此处不再赘述), 力求不再引起误解。

综上, 我们对“4.3 关于对国民心理健康素养提升实践的启示”部分的第一段和第三段做了修改。为便于专家查阅, 特将修改部分粘贴如下(正文部分的修改见红色标记处)。

修改稿:

4.3 关于对国民心理健康素养提升实践的启示

以下以调查结果为依据, **结合已有研究发现**, 对国民心理健康素养提升实践提出几点看法。第一, 要正确把握国民心理健康素养发展的现状, 充分认识心理健康素养提升任务的艰巨性。**当前, 有两个问题**使得我们不敢掉以轻心。一是我国公众心理健康素养整体水平偏低; 二是心理健康素养结构中, **知识观念类维度个体差异相对较大, 态度习惯类维度个体差异相对较小, 并且**容易受干预而改变的是知识观念, 不易改变的是态度和习惯(Morgan, Reavley, Ross, Too, & Jorm, 2018; 任志洪等, 2020; Shi et al., 2019)。前者意味着提升工程的任务巨大, 后者意味着提升工程的难度巨大。目前党和政府的相关规划中, 对国民心理健康素养提升的要求目标高, 时间紧。相关政府部门和行业人士对此挑战要有清醒和足够的认识。

第二, 要把提升心理疾病应对, 尤其是他人心理疾病应对的心理健康素养作为当前的工作重点。他人心理疾病应对主要属于心理疾病污名这个领域。把这个方面作为当前的主攻方向, 有两个理由。一是调查结果显示, 我国公众有关心理疾病应对尤其是他人心理疾病应对的素养最为薄弱。从干预策略来讲, 最薄弱的地方提升空间最大, 最容易做出成效。二是公众心理疾病应对相关素养低下(污名现象严重)的确是当前制约我国心理健康事业发展的主要障碍之一, 做出针对性的重点安排有其现实迫切性。

第三, 要深入研究心理健康相关态度习惯改变的条件和机制。**从既有研究发现来看**, 个体心理健康素养中, 知识观念类素养与态度习惯类素养**可能**存在分离或脱节现象, 增加知识不一定导致态度的积极变化(Angermeyer & Matschinger, 2004; Hengartner et al., 2013; 任志洪等, 2020; Shi et al., 2019)。当前令本领域的研究者和实践者非常头疼的问题, 就是没有找到有效

改变以污名为首的负面态度和行为的干预方法。这个领域此前的研究比较粗陋。例如，学者们一直认为，比较有效可行的方法，一是公共卫生类教育，二是使受干预对象跟精神疾病患者进行接触。大量的研究只是提出各种教育、接触方案，然后检验效果。但教育和接触导致受干预者改变的条件和机制到底是什么？研究却非常少。近两年国际上已经开始转向研究条件和机制，我国学者也需要根据中国的实际情况，深入探索在我国有效改变污名态度的条件和机制。

.....

审稿人 2 意见：

作者对于评审意见进行了有效的回应，并对相关的内容进行修订。目前没有其他意见和建议。

回应：非常感谢审稿专家对本文修改工作的积极肯定。

最后，再次感谢两位审稿专家的悉心审阅和专业建议。

回应涉及的参考文献：

- 罗伯特 J. 格雷戈里. (2013). *心理测量历史、原理及应用(第 5 版)*. (施俊琦 等 译). 北京：机械工业出版社.
- 金盛华, 郑建君, 辛志勇. (2009). 当代中国人价值观的结构与特点. *心理学报*, 41(10), 1000–1014.
- 卢家楣, 刘伟, 贺雯, 王俊山, 陈念劬, 解登峰. (2017). *心理学报*, 49(1), 1–16.
- Morgan, A. J., Reavley, N. J. & Ross, A., Too, L. S., & Jorm, A. F. (2018). Interventions to reduce stigma towards people with severe mental illness: Systematic review and meta-analysis. *Journal of Psychiatric Research*, 103(5), 120–133.
- 任志洪, 赵春晓, 田凡, 闫玉朋, 李丹阳, 赵子仪, ... 江光荣. (2020). 中国人心理健康素养干预效果的元分析. *心理学报*, 52(4), 497–512.
- 王济川, 王小倩, 姜宝法. (2011). *结构方程模型：方法与应用*. 北京：高等教育出版社.

第三轮

审稿人 1 意见：

意见 1：对于第二轮审稿意见 2 的回应中，“而对于输入的数据类型，我们认为以定距数据最为合适，但并非必须是严格的定距变量。比如使用 Likert 量表来测量的心理学变量，其纳入 PCA 分析的原始数据实际上也是等级变量，而非严格意义的定距变量。然而，尽管 Likert 量表的测量实际上是序次测量（ordinal measure），但在 CFA 及其他统计模型分析中，通常将有 5 个及更多点的序次测量作为连续变量分析(王济川, 王小倩, 姜宝法, 2011, p. 41)。”PCA 确实可以使用不严格的定距变量，但通常情况下要说明或者学科习惯默认其为近似定距的，否则的话建议在文中或 Supplemental materials 中补充一下有关说明。

回应：感谢审稿专家的宝贵建议。我们仅在审稿意见回应中做了补充说明，但未在正文中体现，确实给读者带来了不便。因此，我们按照审稿专家的建议，在正文的“2.2.2 《基本背景信息问卷》”部分做补充说明。为便于审稿专家查阅，将修改内容粘贴如下（正文中的修改标记见紫色部分）：

修改稿：

2.2.2 《基本背景信息问卷》

在文献分析的基础上，编制了由 4 组变量组成、共计 11 个题项(含所在城市)的基本背

景信息问卷，以考察心理健康素养的群体差异和影响因素。这 4 组变量分别是地域和 GDP 水平(依所在城市来界定)、基本人口学变量(包括民族、城乡来源、性别、年龄)、社会经济地位(包括收入、职业类型、受教育程度)和心理卫生背景(包括专业身份、和心理/精神疾病患者的接触频率、对专业心理健康服务的熟悉程度)。各变量的计分方式如表 2 所示。对于社会经济地位(social-economic status, SES)的计算，由于在 CFA 及其他统计模型分析中，通常将有 5 个及更多点的序次测量作为连续变量分析(王济川，王小倩，姜宝法，2011, p. 41)。因此，参照有关研究(任春荣，2010；周春燕，郭永玉，2013)，将收入、职业类型、受教育程度三个指标转换为等级变量后，采用因子分析法进行加权变换而来。具体做法是：将家庭人均可支配收入的 6 个水平由低到高分分别赋值 1~6 分，受教育程度的 4 个水平从低到高分分别赋值 1~4 分，职业阶层的 10 个水平(排除在校学生)从低到高分分别赋值 1~10 分；再将三项指标转换成标准分，进行主成分分析，得到 1 个特征根大于 1 的主因子，解释了 66.23% 的方差；利用三个指标在主因子上的载荷系数，做出 SES 的计算公式： $SES=(0.852 \times Z_{\text{受教育程度}} + 0.727 \times Z_{\text{家庭人均可支配收入水平}} + 0.856 \times Z_{\text{职业阶层}})/1.987$ 。其中，0.852、0.727、0.856 分别是三个因子指标的因子载荷，1.987 为第一个因子的特征根。SES 得分越高，表示社会经济地位越高。

意见 2: 对于第二轮审稿意见 3 的回应中，“虽然本研究中的知识和观念为判断题，且采用 0、1 计分，态度和习惯从积极和消极的角度也对其做正误判断，且采用 0、1 计分，但这些题目本质上与认知能力测验的题目不同，不存在必须难度相同方可比较的问题。”建议至少要补充说明在理论上这四个维度之间的含义以及他们之间能够构成比较或相对比较的基础，并给出对这种比较具体意义。

回应: 感谢审稿专家的建议。以下我们将对这四个维度的理论含义、比较的基础，以及比较的意义做详细说明。

如引言中提到的，江光荣等(2020)提出了心理健康素养新概念框架，认为心理健康素养在结构上包含“心理疾病应对——心理健康促进”和“自我——他人”两个维度，在内容上包含知识、态度和行为三个习惯三个方面，具体内容包括：心理健康相关的知识和观念、心理疾病相关的知识和观念、促进自身心理健康的态度和习惯、促进他人心理健康的态度和习惯、应对自身心理疾病的态度和习惯、应对他人心理疾病的态度和习惯六个方面，如图 1 所示。

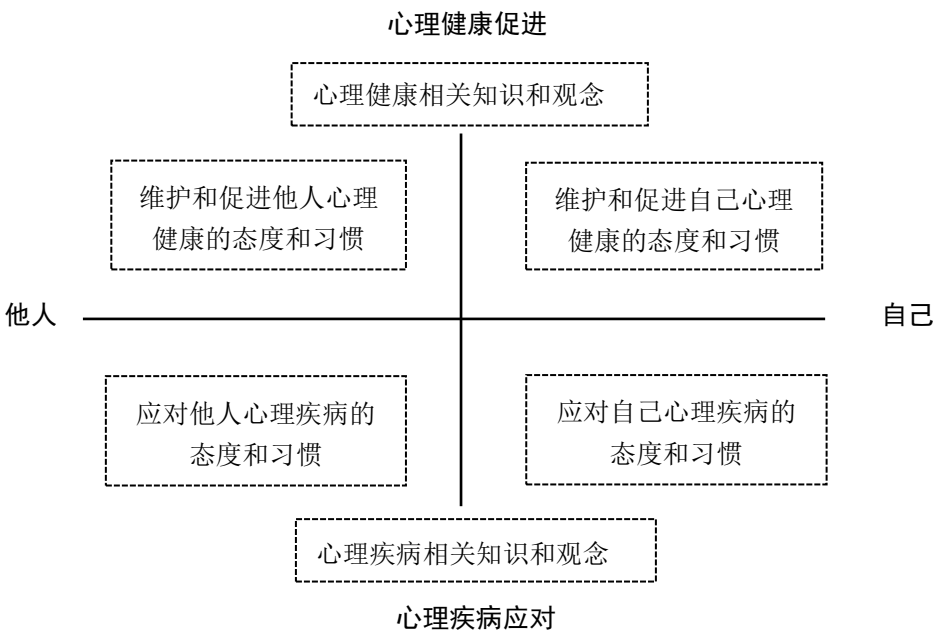


图 1 心理健康素养概念框架

此外，为方便计，江光荣等(2020)将这6个方面分做两大块描述：自助心理健康素养包括维护和促进自身心理健康以及应对自身心理疾病的知识、态度和行为习惯，体现于重视自身健康、对自身心理状况的觉察和了解以及能恰当地自助与求助；助人心理健康素养包括帮助和促进他人心理健康以及帮助和促进他人应对心理疾病的知识、态度和行为习惯，体现为具有心理学思维、重视他人的心理健康、对他人的心理状态敏于觉察，以及恰当地帮助他人应对心理疾病、维护心理健康。相应地，心理健康促进素养包括维护和促进自身以及他人心理健康的知识、态度和行为习惯；心理疾病应对素养包括应对自身和他人心理疾病的知识、态度和行为习惯。

为了解我国国民心理健康素养的结构特点，找出其中的薄弱环节，为今后的心理健康素养干预提升项目提供有针对性的指导，我们将促进自身心理健康的态度和习惯、促进他人心理健康的态度和习惯，视为心理健康促进素养；将应对自身心理疾病的态度 and 习惯、应对他人心理疾病的态度 and 习惯，视为心理疾病应对素养；将促进自身心理健康的态度和习惯、应对自身心理疾病的态度 and 习惯，视为自助心理健康素养；将促进他人心理健康的态度和习惯、应对他人心理疾病的态度 and 习惯，视为助人心理健康素养。通过比较每个方面得分水平的相对高低，来了解我国国民在自助、助人心理健康素养，以及心理疾病应对和心理健康促进素养这四个方面的相对优势和劣势(在测量中，关于心理疾病应对和心理健康促进的知识，是不区分自我和他人的，因而未将心理健康的知识和观念、心理疾病的知识和观念纳入其中)。

理论上，个体在心理健康素养某一方面的得分较高，则意味着其在该方面的素养水平较高。具体比较时，因自助、助人心理健康素养，以及心理疾病应对和心理健康促进素养这四个方面的题目数不同，导致得分的可取值范围不同，即具有不同的计分量尺。因此，我们以平均分为指标，将这四个方面各自的总分除以各自的题目数，使其具有相同的计分量尺，即得分范围均在0~1之间，从而进行后续比较。

综上，我们分别在“1 引言”的第二段和正文的“3.1 我国国民的心理健康素养水平”两个部分做了相应的补充说明。补充内容在正文以紫色标记，并粘贴如下：

修改稿：

1 引言

最初，心理健康素养是指“帮助人们识别、处理及预防心理疾病相关知识和观念”(Jorm et al., 1997)。数年后，Jorm 将其内涵扩展至自助和帮助技能(Jorm, 2012)，但仍局限在心理疾病应对方面。近年来，有学者开始将心理健康促进的内容纳入心理健康素养(CAMIMH, 2007; Kutcher, Wei, & Coniglio, 2016)，并得到了一些学者的支持(Bjørnsen, Eilertsen, Ringdal, Espnes, & Moksnes, 2017)。国内学者明志君和陈祉妍(2020)也强调重视心理健康的意识。事实上，心理健康素养作为认识对象，其关键特征在于两个方面，一是所涉知识和观念的内容，二是所涉态度的对象。这两个方面规定了心理健康素养的基本结构框架。基于此，江光荣等(2020)将心理健康素养界定为“个体在促进自身及他人心理健康，应对自身及他人心理疾病方面所养成的知识、态度和行为习惯”。新概念在结构上包含“心理疾病应对——心理健康促进”和“自我——他人”两个维度，在内容上包含知识、态度和习惯三个方面，具体内容包括：心理健康相关的知识和观念、心理疾病相关的知识和观念、促进自身心理健康的态度和习惯、促进他人心理健康的态度和习惯、应对自身心理疾病的态度和习惯、应对他人心理疾病的态度和习惯六个方面。其中，自助心理健康素养包括维护和促进自身心理健康以及应对自身心理疾病的知识、态度和行为习惯，助人心理健康素养包括帮助和促进他人心理健康以及帮助和促进他人应对心理疾病的知识、态度和行为习惯，心理健康促进素养包括维护和促进自身以及他人心理健康的知识、态度和行为习惯，心理疾病应对素养包括应对自身和他人心理疾病的知识、态度和行为习惯。

3.1 我国国民的心理健康素养水平

由于问卷中关于心理疾病应对和心理健康促进的知识和观念不区分自我和他人,因此将促进自身心理健康的态度和习惯、促进他人心理健康的态度和习惯,视为心理健康促进素养,将应对自身心理疾病的态度和习惯、应对他人心理疾病的态度和习惯,视为心理疾病应对素养,将促进自身心理健康的态度和习惯、应对自身心理疾病的态度和习惯,视为自助心理健康素养,将促进他人心理健康的态度和习惯、应对他人心理疾病的态度和习惯,视为助人心理健康素养,考察我国国民心理健康素养的结构特点。结果显示,与心理健康素养总均分($M=0.60$)相比,心理健康促进素养($M=0.66, p<0.001, d=0.39$)、自助心理健康素养相对较高($M=0.68, p<0.001, d=0.51$),心理疾病应对素养($M=0.57, p<0.001, d=0.22$)、助人心理健康素养相对较低($M=0.54, p<0.001, d=0.35$)。可见,公众在心理健康促进方面的素养高于心理疾病应对,自助心理健康素养高于助人心理健康素养。这个结果从图 2 可以直观地看出来。值得注意的是,与自杀有关的两个题目——“和有自杀念头的人谈论自杀会增加他们的自杀风险”、“假如身边有人流露出自杀念头,我会劝他不要瞎想”——回答正确率最低,仅为 13.3%和 4.3%,说明我国公众极度缺乏与自杀应对相关的知识。表 3 还有一点值得注意,比较 6 个维度的变差系数,可看出知识和观念类维度的变差系数大于态度和习惯类维度。

意见 3: 对于第二轮审稿意见 6 的回应,“感谢审稿专家对心理健康素养总分意义的进一步提问。达到理论中值,意味着国民心理健康素养水平在本调查问卷的量尺上,处于中等水平。而按照通俗意义上的“优良中差”划分标准,“及格线”(百分制的 60 分,本调查问卷中的 36 分)则往往意味着较低水平或表现较差。因此,结合“理论中值”和“及格线”两个参考标准,可认为心理健康素养总分的均值为 35.81 分,意味着我国国民心理健康素养处于一个中偏低的水平。”建议在正文中稍加展开说明,以避免可能的歧义。

回应: 感谢审稿专家的建议。我们未在正文中详细说明,确实不利于读者理解。因此,我们按照审稿专家的建议,在正文的“4.1 关于我国国民心理健康素养的基本情况”第一段做补充说明。为便于审稿专家查阅,将修改内容粘贴如下(正文中的修改标记见紫色部分):

修改稿:

4.1 关于我国国民心理健康素养的基本情况

本调查的一个主要目的,是希望得到我国国民心理健康素养现状的一个基本判断。在调查问卷的量尺上,我国国民心理健康素养总分的均值为 35.81 分,略高于 30 分这一理论中值。这意味着,我国国民心理健康素养水平在本调查问卷的量尺上,处于中等水平。若以百分制的思路,人为划定 60 分为及格,则这个分数刚刚达到及格线(36 分)。而按照通俗意义上的“优良中差”划分标准,及格线往往意味着较低水平或表现较差。因此,结合理论中值和及格线两个参考标准,可认为我国国民心理健康素养处于一个中偏低的水平。一个很自然的疑问是,这个分数如果放在各国之间比较,会排序在什么位置?遗憾的是,因为不存在国际比较的条件,所以目前无法回答这个问题。

最后,再次感谢审稿专家的宝贵建议。

回应涉及的参考文献:

江光荣,赵春晓,韦辉,于丽霞,李丹阳,林秀彬,任志洪. (2020). 心理健康素养:内涵、测量与新概念框架. *心理科学*, 43(1), 232-238.

第四轮 主编终审

作者对审稿人的意见进行了有效的回应和修改，论文可以发表。为使文章更加完善，还需参考以下意见修改：

意见 1：“人为划定 60 分为及格，则这个分数刚刚达到及格线(36 分)”，均值 35.81 分是“刚刚达到”还是“非常接近”36 分的及格线？建议再斟酌用词。

回应：感谢主编的建议。起初使用“刚刚达到”，默认为 35.81 分经四舍五入后为 36 分。但如主编所言，从实际分值来看，确实用“非常接近 36 分的及格线”更为恰当。因此，我们采纳主编的建议，将正文“4.1 关于我国国民心理健康素养的基本情况”中，“若以百分制的思路，人为划定 60 分为及格，则这个分数刚刚达到及格线(36 分)”改为“若以百分制的思路，人为划定 60 分为及格，则这个分数非常接近及格线(36 分)”。

意见 2：“因为从心理健康促进行动的需要来说，公众表现于态度和习惯的心理健康素养才是更有含金量的目标。”这句看不懂，什么叫“更有含金量的目标”？含金量的定义？学术论文建议还是使用更学术一些的用词。谢谢！

回应：感谢主编对论文用词严谨性的建议。

原稿中使用“更有含金量的目标”，是想表达：心理健康素养的态度和习惯相较于知识和观念，对个体的心理健康状况影响更大。因而，从提升心理健康素养、促进心理健康的角度来讲，公众表现于态度和习惯的心理健康素养才是更有价值、更需要被干预的目标。但如主编所言，原文并未对“含金量”做明确的界定，在学术论文中用此表述确有不当之处。

因此，我们采纳主编的建议，将正文“4.2 关于心理健康素养的影响因素”最后一段中，“因为从心理健康促进行动的需要来说，公众表现于态度和习惯的心理健康素养才是更有含金量的目标”，改为“因为从心理健康促进行动的需要来说，公众表现于态度和习惯的心理健康素养才是更有价值、更有助于提升心理健康的目标(郭菲，黄峥，陈祉妍, 2019; Schnyder, Panczak, Groth, & Schultze-Lutter, 2017)”。

最后，再次感谢主编老师的宝贵建议。

回应涉及的参考文献：

Schnyder, N., Panczak, R., Groth, N., & Schultze-Lutter, F. (2017). Association between mental health-related stigma and active help-seeking: systematic review and meta-analysis. *British Journal of Psychiatry: the Journal of Mental Science*, 210(4), 261–268.

郭菲，黄峥，陈祉妍. (2019). 国民心理健康状况调查. 见 傅小兰，张侃(编). *心理健康蓝皮书：中国国民心理健康发展报告(2017~2018)* (pp. 001–055). 北京：社会科学文献出版社.