

《心理学报》审稿意见与作者回应

题目：人格判断能力的老化：信息量的调节作用

作者：刘雪萍，怀淇琛，龚先旻，彭华茂

第一轮

审稿人 1 意见：

该研究本研究采用 2（判断者年龄组：老年、年轻） \times 3（目标人物信息数量：低、中、高）被试间实验设计，探讨了年龄和目标人物信息数量对人格判断准确性、人格判断普遍准确性和特有准确性等的影响。研究选题有新意，研究发现有较好的理论意义和实践价值，对相关研究领域具有一定的创新性贡献；研究设计科学严谨，统计方法使用正确，所得结果稳定、可靠、可信；论文文献综述详实，写作规范，语言流畅，逻辑性强。但研究也存在一些明显不足之处，需要作者进一步澄清或完善。

意见 1：摘要可读性较差，突出问题包括没有研究背景，关键变量没有界定或说明，语句表达不准确、不通顺、表意不清、病句，研究设计和统计方法不清，最后所得结论与前面介绍结果没有关系（或从目前描述中看不出结论从何而来）。从目前的摘要中不能准确判断研究的内容。

回应：感谢专家的意见。先前的摘要存在的问题过多，目前已根据您的建议进行了重写。我们在摘要部分先简单介绍研究背景，然后将研究目的（增加了对关键变量的界定）与研究方法融合在一起介绍，最后呈现研究结果和所得结论。具体如下（亦可见正文摘要部分）：

人格判断对个体社会功能的发挥具有重要影响。本研究采用 2（判断者年龄组：老年、年轻） \times 3（目标人物信息量：单一、中等、丰富）被试间实验设计，考察了信息量对老年人和年轻人人格判断准确性的影响，包括普遍准确性（对目标人物所在群体人格共性判断的准确性）和特有准确性（对目标人物个体人格判断的准确性）。结果显示，老年人人格判断的普遍准确性高于年轻人，且信息量的多少并不影响这一年龄优势。信息量只影响特有准确性判断的年龄差异：单一信息量（仅实时文字）时，老年人的特有准确性低于年轻人；中等

（实时文字和声音）和丰富信息量（实时文字、声音和图像）时，二者表现相当。综上，老年人的人格判断能力总体保持较好，且能有效利用增量信息来弥补与年轻人在特有准确性上的差距。

意见 2：“社会准确性模型”那一段，大部分内容都是在论证目标群体信息在人格判断中的作用，对于老年人普遍准确性相关内容的论证比较少，还显得很不充分，过于单薄，建议丰富老年人为何能在普遍准确性判断上存在优势的论述。

回应：的确如您所指出的，先前对老年人普遍准确性比年轻人更有优势的论证不足。对此，我们增加了相关文献阅读，但鉴于前言部分有 3500 字的字数限制及参考文献有 50 条的数量限制，我们对老年人在普遍准确性判断上存在优势的相关论述进行了适当加强。具体如下（亦可详见正文“1 前言”第四段）：

随着年龄增长，老年人的社会知识经验（包括人格共性知识）变得愈加丰富，运用社会知识的能力增强(Freund & Isaacowitz, 2014; Hess et al., 1999)。由于运用社会知识经验的能力与社会判断质量正相关(Leclerc & Hess, 2007)，老年人丰富的人格共性知识经验可能使其在人格判断的普遍准确性上更有优势。因为对人格共性知识的提取和运用较少受具体行为信息的影响(Krzyzaniak et al., 2019)，若老年人在人格判断的普遍准确性上存在优势，这种优势应当不受当前信息的特征（如信息量的多寡）影响。

新增参考文献：

Leclerc, C. M., & Hess, T. M. (2007). Age differences in the bases for social judgments: Tests of a social expertise perspective. *Experimental Aging Research*, 33(1), 95–120.

<https://doi.org/10.1080/03610730601006446>

意见 3：为了增加综述部分的可读性和流畅性，建议把穿插在文献综述中的研究问题统一整合到文稿的最后，在对已有文献局限总结的基础上提出研究问题，顺势提出研究假设。

回应：感谢您的建议。目前我们已将研究问题整合到文献综述最后，并在回顾文献的基础上顺势提出了两个研究假设。具体如下（亦可详见正文“1 前言”最后一段）：

由于社会知识经验的积累，老年人可能在普遍准确性判断上比年轻人更有优势，且普遍准确性判断较少受当前信息的影响。因此，本研究提出假设 1：老年人在人格判断上的普遍准确性高于年轻人，这一优势不受信息量的影响。由于认知能力的下降，老年人可能无法像年轻人那样充分利用增量信息来提高人格判断的特有准确性。因此，本研究提出假设 2：

信息量调节特有准确性的年龄差异：老年人在人格判断上的特有准确性低于年轻人，这一年龄劣势随着信息量增加而加剧。

意见 4：建议在研究方法之前，对本文的实验及其相关变量做一个简要的介绍，这样能够对本研究实验能有个大致的预期。尤其对于一些关键的变量，在没有任何预期的情况下，在方法部分突然出现（如什么是目标任务信息数量，如何操作；为什么要关注年龄差异等），会增加读者理解文稿的难度。

回应：非常感谢您的这一建议。根据您的建议，我们在研究方法之前，先对本研究的实验流程及相关变量进行了简要介绍，以帮助读者更好地理解后续的研究内容。具体如下（亦可详见正文“1 前言”倒数第二段）：

综上，本研究通过改变信息通道数量操纵信息量，考察信息量变化对老年和年轻判断者人格判断准确性（以自我-他人一致性为指标）的影响。具体而言，被试将观看相同时长但所含信息量不同（单一信息量：仅实时文字；中等信息量：实时文字和原始音频；丰富信息量：实时文字、原始音频和图像）的目标人物自我介绍视频，随后判断目标人物的大五人格特质。同时，基于社会准确性模型的多层线性分析方法，进一步检验判断者年龄和信息量对普遍准确性和特有准确性两个人格判断子成分的影响，以厘清人格判断准确性年龄差异的来源。

意见 5：研究为何不使用混合设计？信息量水平作为被试内变量似乎更合理。

回应：感谢您对研究设计所提出的意见。确实，把信息量水平作为被试内变量，在逻辑上会更加合理，能够更充分地说明同一个判断者的人格判断准确性变化是由信息量的变化带来的。但是落实到实际实验中，我们发现会存在若干可能的问题：

一是，在本研究中，被试在不同信息量下需要判断的目标人物是相同的，如正文中的图 1 所示，被试在单一（仅包含实时文字）、中等（实时文字+原始音频）、丰富（实时文字+原始音频+图像）信息量下，都是判断同一个目标人物（共计 24 名目标人物）。假如采用混合实验设计，即信息量作为被试内变量的话，被试将在不同信息量下，对同一个目标人物进行三次人格判断，后两次判断必然会受到第一次判断的影响，那么，得到的判断结果将不再可信。

二是，要想避免上述问题，那就只能通过变换目标人物来实现，即同一被试在三种信息量下对不同的目标人物进行判断。但这样一来，目标人物间的差异也可能会对判断者的人格

判断准确性造成影响，如有的目标人物可能更擅长表达自己，有的目标人物则更倾向于隐藏自己，那么由目标人物的差异产生的效应可能会混淆信息量对人格判断准确性的影响，甚至掩盖信息量的作用。

三是，若将信息量作为被试内变量，各水平下的目标人物不同，那么被试判断目标人物的实验时长将大大增加，被试可能会严重疲劳。目前的实验时长约为 70 分钟，减去社会人口学信息及背景变量的测量，被试判断 24 名目标人物大概需要 50-60 分钟。若采用被试内设计，被试需要判断三种信息量下共计 72 名目标人物的人格，那么将花费近 3 小时时间，被试疲劳对结果产生的影响也不能忽视。

权衡利弊后，我们最终决定采用判断者年龄组和目标人物信息量两因素完全随机的被试间设计，还请专家见谅。

意见 6: 另外，对于信息量水平这个变量，并不是在同一个维度进行操纵的，不同维度信息的呈现对于老年人而言可能并不是认知负荷的增加，也有可能是减少认知负荷（如视觉通道的材料的呈现可能会降低文字材料和听觉通道刺激的负荷），如果用认知功能负荷相关的理论来理解本研究中普遍准确性和特有准确性的年龄差异可能并不太合适。该变量的设置是否合理？至少三个水平是否真的代表了高、中、低三种认知负荷水平？根据双加工理论，图文并茂会促进理解，降低认知负荷。

回应: 非常感谢您对于信息量水平这一关键概念提出的质疑，这帮助我们进一步思考了该变量的设置及本研究设计的核心。本研究关注老年人和年轻人在人格判断准确性上是否存在差异，并进一步探究信息量的变化会如何影响不同年龄个体的人格判断过程。以往研究大多以信息呈现时间的长短界定信息量的多寡(e.g., Beer et al., 2019; Vicaria et al., 2015; Wiedenroth & Leising, 2020)，单位时间内呈现的信息量相对恒定，个体在不同时刻被占用的认知资源量基本相似。因此，这些研究中信息量的操纵难以对判断者的信息觉察和利用产生影响，从而难以检验出判断者年龄和信息量对人格判断的交互影响。本研究采用同时性方法操纵信息量来避免这一研究缺陷。具体而言，本研究通过操纵呈现信息的通道数量来改变单位时间内的信息量，从而考察信息量变化对老年和年轻判断者的人格判断准确性的影响。具体而言，不同信息量条件下的被试将观看相同时长但包含信息量不同（单一信息量：仅实时文字；中等信息量：实时文字和原始音频；丰富信息量：实时文字、原始音频和图像）的目标人物自我介绍视频，随后判断目标人物的大五人格特质。

我们同意专家的说法，本研究三个水平下的信息量未必严格对应认知负荷，但是的确

确反映了线索数量的变化。相比于仅呈现实时文字信息，增加音频后和增加音频与图像后，被试可以利用的信息增加了，增加的信息里可能包含了更广泛的人格相关线索（如目标人物的音色、音调、语速和情绪变化等信息，以及面部表情、注视镜头时长、肢体语言幅度及外貌特征等非言语信息），这些信息可以帮助被试更加全面地判断目标人物的人格(吴婷，郑涌，2019; Funder, 1995; Wiedenroth & Leising, 2020)。同时，增加音频后和增加音频与图像后，也可能会使得被试顾此失彼，错过有效的行为信息，或被无效信息干扰，从而影响人格判断的准确性。因此，随着信息变得愈加丰富，个体的人格判断既可能受到信息量增加的积极影响，也可能受到消极影响。

当通过多个通道同时呈现信息时，年轻人经常表现出对不同通道信息的良好整合(Stevenson et al., 2015)，从而提升其人格判断的特有准确性。但对老年人而言，感知觉功能和注意广度等认知功能的下降可能使其难以觉察、整合和利用不同感觉通道中的有效信息(Freiherr et al., 2013; Hess & Smith, 2014)。另外，由于抑制能力的减弱(Hess & Smith, 2014)，老年人比年轻人更难抑制无关信息的干扰(Radvansky et al., 2010)，更易受到感觉通道之内及之间信息相互冲突的影响(Freiherr et al., 2013; Jones & Noppeney, 2021)，从而对其人格判断的特有准确性产生不利影响。综上，我们推测信息通道数量增加可能提高年轻人而非老年人在人格判断上的特有准确性，从而使老年人在特有准确性上的年龄劣势加剧。

最后本研究的结果发现，随着信息量的增加，老年人的特有准确性在上升，和年轻人在特有准确性判断上的差距在减小而非增加（与假设 2 相反）。产生这种矛盾的可能原因之一是，声音和图像信息的呈现使得能利用的有效信息更多且更生动，即使老年人不能加工整合来自不同通道的所有信息(Freiherr et al., 2013)，但他们仍然能有效地抓取和利用其中有效的部分线索进行人格判断(Hess & Smith, 2014; Hess et al., 1999; Hess et al., 2005)。而且增加音频和图像信息后的情境与老年人在现实生活中真实的社会互动情境（如打电话、视频或见面沟通）更加类似，有助于发挥其丰富的社会知识和经验(Castro & Isaacowitz, 2019b; Schlegel et al., 2020)，从而弥补了其年轻人在特有准确性判断上的差距。Schlegel 等人(2020)采用生态效度较高的面对面真实互动情境，也发现了特有准确性的年龄相似性，与本研究中等和丰富信息量的年龄结果一致。

综上，我们决定在原有阐述思路的基础上，明确本研究是通过改变信息通道数量来操纵信息量的，并对文献综述和结果讨论部分的阐述逻辑进行适当调整，同时对变量及变量水平的表述进行重新斟酌后修改。将原来的“目标人物信息数量：低、中、高”修改为“目标人物信息量：单一、中等、丰富”（注：信息流中的“信息量”是不可数的，与先前可数的“信息数

量”的表述相比，更为贴切)。即，分别用“单一信息量”、“中等信息量”、“丰富信息量”表示仅实时文字、实时文字+原始音频、实时文字+原始音频+图像这三种信息量下的自我介绍视频。这样，既明确了目标人物信息量变化的操纵方法，还真实反映了信息量的变化，也避免了“低、中、高信息数量”这样的表述给读者造成理解上的歧义，带来认知负荷联想。对此，我们已将研究设计及题目、摘要、前言、研究方法、研究结果及讨论部分的相关表述都进行了修改。

新增参考文献：

Freiherr, J., Lundström, J. N., Habel, U., & Reetz, K. (2013). Multisensory integration mechanisms during aging.

Frontiers in human neuroscience, 7, 863. <https://doi.org/10.3389/fnhum.2013.00863>

Jones S. A., & Noppeney U. (2021). Ageing and multisensory integration: A review of the evidence, and a

computational perspective. *Cortex*, 138(7), 1-23. <https://doi.org/10.1016/j.cortex.2021.02.001>

Stevenson, R. A., Nelms, C. E., Baum, S. H., Zurkovsky, L., Barense, M. D., Newhouse, P. A., & Wallace, M. T.

(2015). Deficits in audiovisual speech perception in normal aging emerge at the level of whole-word

recognition. *Neurobiology of Aging*, 36(1), 283–291. <https://doi.org/10.1016/j.neurobiolaging.2014.08.003>

意见 7：实验中，研究者使用的目标视频材料的内容（目标人物介绍的事件）与人格特征是否相关，作者是否对这些内容进行过分析？不同类型的内容对于观看者判断的影响是不同的。

回应：感谢您对此提出的疑问和意见。在本研究中，我们让目标人物简要陈述一件近期发生的自己作为重要参与者的印象深刻的事情。我们在视频录制前，已经告知目标人物，这一份自我介绍要让观看视频的人快速了解自己，但尽量避免使用评价性词汇。因此，目标人物对于自我介绍的目的是清楚的。

更为重要的是，本研究关注的是判断者年龄和目标人物信息量这两个自变量对被试人格判断准确性及其子成分的影响，24 名目标人物及这些目标人物所介绍的事件内容，在不同年龄判断者以及不同信息量下都是一致的，具体事件内容与人格特征的相关性和可用性都是固定的。因此，目标人物所介绍的具体事件内容对人格判断结果的影响是恒定的，并不会影响判断者年龄和目标人物信息量对人格判断准确性的结果。

根据本研究的数据分析结果，无论是年轻判断者还是老年判断者，自我-他人一致性这一人格判断指标值均显著大于 0，为 0.37-0.54 之间（Fisher Z 转换后，见正文表 2），即被试对目标人物的人格他评分数与目标人物的人格自评分数之间的相关程度达到了中等。同时，

多层线性分析的模型 1 检验结果也表明：第一，目标人物年龄群体的人格自评分数均值能显著预测被试对目标人物的人格他评分数， $B = 0.872$ ， $SE = 0.039$ ， $p < 0.001$ ，说明被试评判陌生目标人物人格时具有一定程度的普遍准确性；第二，目标人物的自评分数（已减去其所处年龄群体的自评分数均值）同样能显著预测被试对目标人物的人格他评分数， $B = 0.052$ ， $SE = 0.005$ ， $p < 0.001$ ，说明被试评判陌生目标人物人格时同样具有特有准确性。也就是说，被试评判陌生目标人物人格时不仅具有中等程度的自我-他人一致性，而且具有一定程度的普遍准确性和特有准确性。这些数据结果都说明了目标人物介绍的事件是与其人格特征是相关的，判断者能够根据目标人物的自我介绍判断出其人格特征。

另外，的确如专家所说，目标人物阐述不同的事件，可能会让被试产生不同的人格判断结果。例如，讲述近期家庭中所发生的变故与讲述较为日常的生活故事，二者反映的人格特征线索会存在较大差异。感谢专家提醒，目标人物的言行表现可能会影响判断者的人格判断准确性，这是值得进一步研究的相关问题。

意见 8：“中文版十题人格量表”他评版是否经过初测，是否有相关的信度和效度证据。

回应：感谢您的意见。“十题人格量表”原量表是 10 个项目的大五人格量表，每一项人格分别使用 2 个项目测量，如外向性包括 E1: extraverted, enthusiastic（外向的，精力充沛的）和 E2: reserved, quiet（内向的，安静的）。本研究所采用的“中文版十题人格量表”自评版和他评版均是在“中文版十题人格量表”基础上明确了人格评估的对象，自评版为自己（我），他评版为具体的目标人物。我们认为，量表的核心内容并未发生任何变化，仅仅是让评估的对象从大多数研究关注的自己变为他人（目标人物），这并不会影响量表本身的信度和效度。而中文版十题人格量表的信效度已经经过李金德（2013）的验证，验证结果已发表在《中国健康心理学杂志》上，证明了该量表可以作为大五人格可靠和有效的工具。

意见 9：内隐人格信念问卷在本研究中的作用是什么？

回应：感谢您的这一问题。内隐人格信念问卷主要反映的是个体把人格和特定行为联系起来的程度。例如，若个体更加认可“人格实体论”，即人格决定行为，那么，个体就会更倾向于认为高外向性的人，在交流时会有更多的眼神接触。也就是说，这样的个体会更加依赖视频材料的内容去判断目标人物的人格。我们认为这是一个可能会影响被试人格判断准确性的个性特征变量。因此，在研究中作为背景变量进行了测量。但独立样本 t 检验显示，老年人和年轻人的内隐人格观不存在显著的年龄差异($t(161) = 0.81, p = 0.418$)，因此，在后续分析中，

并未进行控制。

意见 10：实验在腾讯会议的环境中展开，其得到的结果是否可以代表实际生活情境中的情况？作者是否进行了相关的检验？

回应：感谢专家对此所提出的疑问。鉴于研究数据收集期间，疫情防控严格，我们为了顺利推进研究，让部分被试通过腾讯会议在主试“一对一”的带领下完成实验。和来到线下实验室参与实验的被试相同，我们要求线上参与的被试在安静、不被他人打扰的单独房间参与实验，并且要能让主试通过视频看到周围的环境，同时，一切实验流程均在主试的引导和掌控之下。本研究 82 名年轻人中，有 24 名为线下实验室参与实验，58 名为线上腾讯会议参与实验，将这两批被试的人格判断准确性（自我-他人一致性为指标）进行独立样本 t 检验，结果显示并无显著差异， $MD = -0.04, t(1, 80) = -1.22, p = 0.225$ 。可见，线上被试的实验环境与线下被试并不存在显著差别，不存在明显的干扰实验进展和结果的因素。

对于腾讯会议环境下开展的实验，结果能否代表实际生活情景中的情况。我们认为，这更取决于研究所使用的材料。在本研究中，无论是线上还是线下，被试都是观看目标人物的自我介绍视频，并没有本质区别。在实验室环境中（包括线上腾讯会议开展的实验中），我们会尽可能地降低无关、混淆变量对实验结果的影响（如噪音、他人走动带来的干扰等），以更加纯粹地反映所关注的研究变量对研究结果的作用。在本研究中，我们已经在实验实施过程中进行了控制。另外，我们认为要关注实验的生态效度，则更加需要关注的是实验材料本身。本研究所使用的实验材料，与日常生活中人们初次见面时进行简短的自我介绍情境是类似的。只不过通过严格的实验控制，使得不同的实验条件下，呈现给被试的信息量是存在差异的。因此，我们认为，本研究在腾讯会议环境下开展实验得到的结果，能够代表实际生活情境的情况。

意见 11：研究结果为何要使用多层线性模型来处理？该方法对于本研究来说其优越性主要表现在哪些方面？本研究的被试量是否适合使用该种统计方法（g*power 计算出的被试量是否适用于该种统计方法？）？

回应：感谢专家对统计方法所提出的意见。本研究基于社会准确性模型，将人格判断准确性分解成普遍准确性和特有准确性两个成分分别进行了探究(Biesanz, 2010)，使得对人格判断准确性年龄差异的认识和讨论更加深入。具体而言，本研究以被试在中文版十题人格量表各个项目上的他评分数(Y_{ijk})为结果变量，模型的第一水平（即“判断者-目标人物”层： $Y_{ijk} = \beta_{0i} +$

$\beta_{1i} Norm_{jk} + \beta_{2i} TSelf_{jk} + r_{ijk}$) 上有两个预测变量, 一是目标人物所处年龄群体的各个人格特质的自评分数均值($Norm_{jk}$), 代表了相应年龄群体的人格共性。二是减去年龄群体均值后每名目标人物在各个项目上的自评分数($TSelf_{jk}$)。纳入模型前, 上述两个变量均进行了中心化处理。在第二水平(即“判断者”层)上, 预测变量为被试年龄组(Age_i)、信息量($Info1_i$ 和 $Info2_i$)及二者交互项(详见正文 2.5.2 多层线性分析)。

由上可见, 采用多层线性模型, 不仅可以同时考察 $Norm_{jk}$ 和 $TSelf_{jk}$ 对被试人格判断的影响, 更重要的是, 还能进一步考察不同因素(判断者年龄组 Age_i 和信息量 $Info1_i$ 与 $Info2_i$)以及不同因素之间的交互作用(判断者年龄组和信息量二者的交互项 $Age_i \times Info1_i$ 和 $Age_i \times Info2_i$)对被试人格判断的普遍准确性 β_{1i} 和特有准确性 β_{2i} 的影响。即, 应用多层线性模型的优势在于可以检验判断者年龄组、目标人物信息量和二者的交互项对两个人格判断准确性子成分的作用, 进一步厘清人格判断准确性年龄差异的内在原因。

对于多层线性模型的样本量计算, Scherbaum & Pesner (2019)发现, 2010-2015 年发表在 *Journal of Applied Psychology* 上的 139 篇多层线性研究中, 只有 3 篇报告了统计检验力或根据效应量分析事先计算了最小样本量。可见, 目前对于多层线性研究的样本量估计还是不太成熟的。再回溯以往文献发现, Maas & Hox (2005)采用模拟研究确定了组水平上不同样本量对估计准确性(回归系数和方差)及标准误的影响, 结果发现, 只有第二水平上的样本量小于 50 时, 才会导致对第二水平的标准误估计出现偏差, 在所有其他模拟条件下, 回归系数、方差和标准误的估计都是无偏和准确的。根据社会准确性模型, 每名判断者被试和某一特定目标人物的组合被视为统计检验的基本单元, 在本研究中 163 名判断者被试分别对 24 名目标人物进行人格判断, 一共存在 3912 (163×24) 个不同的“判断者-目标人物”组合, 在第二水平(即“判断者”层)上的样本量为 163, 根据 Maas & Hox (2005), 这一样本量是足以开展多层线性分析的。本研究采用 g*power 计算出的被试量主要是为了满足方差分析的需要, 因此, 主要是基于方差分析的统计量要求进行的样本量估计。

相关参考文献:

- Scherbaum, C. A., & Pesner, E. (2019). Power analysis for multilevel research. In S. E. Humphrey & J. M. LeBreton (Eds.), *The handbook of multilevel theory, measurement, and analysis* (pp. 329–352). American Psychological Association. <https://doi.org/10.1037/0000115-015>
- Maas, C., & Hox, J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology European Journal of Research Methods for the Behavioral and Social Sciences*, 1(3), 86-92. DOI: 10.1027/1614-2241.1.3.86

意见 12: 研究结果部分, 被试的基本情况的数据建议整合到被试部分, 这一部分的结果与研究问题无关; 数据缺失情况可以整合到其他部分。

回应: 感谢专家的建议。目前, 被试基本情况的数据已经整合到“2.2 被试”部分。数据缺失情况的内容已整合到“2.4 研究程序”部分。

意见 13: 讨论部分内容总体思路较清楚, 但请结合前面信息量水平变量的问题进一步思考, 对实验中信息量水平结果的解释是否需要进一步修改。

回应: 非常感谢专家的建议。结合两位专家对信息量水平的意见和建议, 我们在讨论中对“信息量对人格判断及特有准确性的影响”部分做了相应的调整和修改, 主要修改的部分如下(亦可详见正文“4.2 信息量对人格判断及特有准确性的影响”第四段):

首先, 声音和图像信息的呈现使得能利用的有效信息更多且更生动, 即使老年人不能加工整合来自不同通道的所有信息(Freiherr et al., 2013), 但他们仍然能有效地抓取和利用其中有效的部分线索进行人格判断(Hess & Smith, 2014; Hess et al., 1999; Hess et al., 2005)。而且增加音频和图像信息后的情境与老年人在现实生活中真实的社会互动情境(如打电话、视频或见面沟通)更加类似, 有助于发挥其丰富的社会知识和经验(Castro & Isaacowitz, 2019b; Schlegel et al., 2020), 从而弥补了其与年轻人在特有准确性判断上的差距。Schlegel 等人(2020)采用生态效度较高的面对面真实互动情境, 也发现了特有准确性的年龄相似性, 与本研究中等和丰富信息量的年龄结果一致。

意见 14: 很少的一部分句子表达不够凝练、简洁, 如“这是因为更多的行为信息中往往包含了更多人格相关的线索”一句中包括了两个“更多”, 虽无语法错误, 但句子的流畅性不足; 再如“因变量为人格判断准确性, 以“自我-他人一致性”为整体衡量指标, 同时利用多层线性模型将其分解为普遍准确性和特有准确性, 考察自变量对两个子成分的影响”, 后面几个分句的主语并不清楚, 建议分成几个完成的句子来写; 还如“目标人物各项人格的自评分数分布在年龄组内部都相对广泛且五项人格特质自评分数均不存在显著的年龄差异”一句太长, 句子成分切分困难; “即使是初次见面的情境, 也需要判断他人人格以决定是否与其交往、如何与其交往”一句缺少主语。建议通篇检查上述表述问题, 一一修改、完善。

回应: 感谢专家指出上述表达上的问题, 我们一一进行了修改和完善。

上述提及的句子分别修改如下:

(1) 这是因为更多的行为信息往往意味着更广泛的人格相关线索, 有助于判断者充分综合

不同信息来形成更准确的人格判断(Funder, 1995)。

(2) 采用 2 (判断者年龄组: 老年、年轻) \times 3 (目标人物信息量: 单一、中等、丰富) 两因素完全随机被试间设计, 因变量为人格判断准确性 (以“自我-他人一致性”为整体衡量指标)。同时, 利用多层线性模型将人格判断准确性分解为普遍准确性和特有准确性, 进一步检验判断者年龄组和信息量对这两个人格判断子成分的交互影响。

(3) 选取规则为: 目标人物各项人格的自评分数分布在年龄组内部都相对广泛, 且外向性 ($t(22) = -1.03, p = 0.317$)、宜人性 ($t(22) = 0.47, p = 0.642$)、尽责性 ($t(22) = -0.91, p = 0.375$)、神经质 ($t(22) = 1.78, p = 0.089$) 和开放性 ($t(22) = 0.12, p = 0.909$), 这五项人格特质的自评分数均不存在显著的年龄差异。

(4) 个体在整个生命历程中都需要与他人进行社会互动, 经常需要对他人人格进行判断 (Castro & Isaacowitz, 2019a), 以决定是否及如何与其交往 (Funder & West, 1993; Nestler & Back, 2013)。

同时, 我们对文章进行了通篇检查, 并修改完善。其他修改的地方如:

(1) 将“然而, 值得注意的是上述结论是在默认判断者认知资源充足、有能力充分整合和利用不同信息的情况下得出的。信息数量的增多对认知能力下降、认知资源不足的老年群体来说却未必有利。这是因为信息数量增加, 一方面意味着要耗费更多的认知资源去从大量的信息中觉察出有效的行为信息, 另一方面, 正确利用所觉察到的有效信息并忽略无效信息来形成判断也更具有挑战性。”修改为“然而, 上述结论的前提条件是判断者有充足的认知资源和能力来充分整合和利用不同信息。随着信息量增加, 判断者需要耗费更多的认知资源从大量信息中觉察出并正确利用有效的人格相关线索, 同时抵抗无效信息的干扰。由于老年人的认知能力下降、认知资源不足, 信息增多对其而言未必有利, 反而可能带来挑战, 从而妨碍其人格判断的准确性。”(详见正文“1 前言”第五段最后)

(2) 将“但我们同时也应该警惕这种对群体人格共性的认识演变为一种群体刻板印象, 值得庆幸的是, 信息数量增加可以提升老年人对目标人物个体的人格判断, 弥补其与年轻人在特有准确性判断上的差距。”修改为“不过我们同时也应该警惕这种对群体人格共性的认识演变为一种群体刻板印象。值得庆幸的是, 信息量提升可以提高老年人对目标人物个体的人格判断, 助其避免刻板印象式的人格判断, 弥补其与年轻人在特有准确性上的差距。”(详见正文“4.1 人格判断及普遍准确性表现出稳定的年长优势”第二段最后)

意见 15: 部分内容表述形式欠规范, 如“外向性: $t(22) = -1.03, p = 0.317$; 宜人性: $t(22) = 0.47$,

$p = 0.642$; 尽责性: $t(22) = -0.91, p = 0.375$; 神经质: $t(22) = 1.78, p = 0.089$; 开放性: $t(22) = 0.12, p = 0.909$ ”与下面类似内容“发现目标人物两次自评的外向性($r = 0.804, p < 0.001$)、宜人性($r = 0.407, p = 0.048$)、尽责性($r = 0.687, p < 0.001$)、神经质($r = 0.775, p < 0.001$)和开放性($r = 0.442, p = 0.031$)分数均显著正相关”形式不一样, 其中上一句冒号的使用在对应句子中的使用可能不规范, 建议全文对于类似的句子统一表述形式。

回应: 感谢专家指出这一问题, 已检查全文类似的句子, 并进行了统一、规范化表达。上述提及的表述修改如下 (亦可详见正文“2.3.1 目标人物视频材料”第一段):

且外向性 ($t(22) = -1.03, p = 0.317$)、宜人性 ($t(22) = 0.47, p = 0.642$)、尽责性 ($t(22) = -0.91, p = 0.375$)、神经质 ($t(22) = 1.78, p = 0.089$) 和开放性 ($t(22) = 0.12, p = 0.909$), 这五项人格特质的自评分数均不存在显著的年龄差异。

意见 16: 关键词最后的句号可以删掉。

回应: 感谢专家指出, 已删除。

意见 17: 文稿中还存在校对不认真、不到位的情况, 如“反映了个体将某种人格与特定行为相联系的程度”, 中间多了一个*。

回应: 感谢专家指出, *已删除。同时, 我们在文章定稿上传前, 对文稿全文进行了认真校对, 确保不再出现此类低级错误。

.....

审稿人 2 意见:

我认为这是一项有意思且有意义的研究。这篇论文从成年发展的视角研究人格判断, 发现了人格判断这一社会认知能力上的年长优势, 并分析了年龄差异的原因。这一研究增加了对社会认知能力的增龄变化的理解。

意见 1: 我对这个研究最主要的疑惑是关于“信息数量”这个自变量的。

首先, 个人认为, “信息数量”这个说法可能并不准确。这个变量更像是信息媒介形式的不同 (文字/声音/视频) 或者生态性的高低, 而并非信息“数量” (quantity) 的差异 (数量的差异应在同一类型或同一标尺下比较)。

其次, 作者立论的出发点是, 老年人注意资源有限、认知控制能力较差, 因此在信息数量增

加时，无法有效识别并利用有效信息，并引用了一些无关信息降低老年人人格判断成绩的研究作为支持论据，并假设老年人在信息数量高条件下人格判断无法提升反而可能下降。这一思路逻辑是顺畅的，但对信息数量的操作定义没有体现这一思路，并不是信息数量高（视频条件）下对认知能力要求越高，反而是信息数量低（文字条件）可能更考验老年人的认知功能，因此并没有成功地考察基础认知能力或者说信息数量差异后的认知负荷是否会影响人格判断的年龄差异。

回应：非常感谢您对“信息数量”这个自变量所提出的质疑，您对于这一变量的看法帮助我们更深入地思考了该变量的内涵及表述。本研究关注老年人和年轻人在人格判断准确性上是否存在差异，并进一步探究信息量的变化会如何影响不同年龄个体的人格判断过程。以往研究大多以信息呈现时间的长短界定信息量的多寡(e.g., Beer et al., 2019; Vicaria et al., 2015; Wiedenroth & Leising, 2020)，单位时间内呈现的信息量相对恒定，个体在不同时刻被占用的认知资源量基本相似。因此，这些研究中信息量的操纵难以对判断者的信息觉察和利用产生影响，从而难以检验出判断者年龄和信息量对人格判断的交互影响。本研究采用同时性方法操纵信息量来避免这一研究缺陷。具体而言，本研究通过操纵呈现信息的通道数量来改变单位时间内的信息量，从而考察信息量变化对老年和年轻判断者的人格判断准确性的影响。具体而言，不同信息量条件下的被试将观看相同时长但包含信息量不同（单一信息量：仅实时文字；中等信息量：实时文字和原始音频；丰富信息量：实时文字、原始音频和图像）的目标人物自我介绍视频，随后判断目标人物的大五人格特质。

我们同意您的看法，用“低”、“中”、“高”信息数量来分别代表上述信息是不准确的，因为它们的确不是在同一类型或同一标尺下的。但本研究三种信息量下的自我介绍的的确反映了线索数量的变化。相比于仅呈现实时文字信息，增加音频后和增加音频与图像后，被试可以利用的信息增加了，增加的信息里可能包含了更广泛的人格相关线索（如目标人物的音色、音调、语速和情绪变化等信息，以及面部表情、注视镜头时长、肢体语言幅度及外貌特征等非言语信息），这些信息可以帮助被试更加全面地判断目标人物的人格(吴婷，郑涌，2019; Funder, 1995; Wiedenroth & Leising, 2020)。同时，增加音频后和增加音频与图像后，也可能会使得被试顾此失彼，错过有效的行为信息，或被无效信息干扰，从而影响人格判断的准确性。因此，随着信息变得愈加丰富，个体的人格判断既可能受到信息量增加的积极影响，也可能受到消极影响。

当通过多个通道同时呈现信息时，年轻人经常表现出对不同通道信息的良好整合(Stevenson et al., 2015)，从而提升其人格判断的特有准确性。但对老年人而言，感知觉功能

和注意广度等认知功能的下降可能使其难以觉察、整合和利用不同感觉通道中的有效信息(Freiherr et al., 2013; Hess & Smith, 2014)。另外, 由于抑制能力的减弱(Hess & Smith, 2014), 老年人比年轻人更难抑制无关信息的干扰(Radvansky et al., 2010), 更易受到感觉通道之内及之间信息相互冲突的影响(Freiherr et al., 2013; Jones & Noppeney, 2021), 从而对其人格判断的特有准确性产生不利影响。综上, 我们推测信息通道数量增加可能提高年轻人而非老年人在人格判断上的特有准确性, 从而使老年人在特有准确性上的年龄劣势加剧。

最后本研究的结果发现, 随着信息量的增加, 老年人的特有准确性在上升, 和年轻人在特有准确性判断上的差距在减小而非增加(与假设 2 相反)。产生这种矛盾的可能原因之一是, 声音和图像信息的呈现使得能利用的有效信息更多且更生动, 即使老年人不能加工整合来自不同通道的所有信息(Freiherr et al., 2013), 但他们仍然能有效地抓取和利用其中有效的部分线索进行人格判断(Hess & Smith, 2014; Hess et al., 1999; Hess et al., 2005)。而且增加音频和图像信息后的情境与老年人在现实生活中真实的社会互动情境(如打电话、视频或见面沟通)更加类似, 有助于发挥其丰富的社会知识和经验(Castro & Isaacowitz, 2019b; Schlegel et al., 2020), 从而弥补了其年轻人在特有准确性判断上的差距。Schlegel 等人(2020)采用生态效度较高的面对面真实互动情境, 也发现了特有准确性的年龄相似性, 与本研究中中等和丰富信息量的年龄结果一致。

综上, 我们决定在原有阐述思路的基础上, 明确本研究是通过改变信息通道数量来操纵信息量的, 并对文献综述和结果讨论部分的阐述逻辑进行适当调整, 同时对变量及变量水平的表述进行重新斟酌后修改。将原来的“目标人物信息数量: 低、中、高”修改为“目标人物信息量: 单一、中等、丰富”(注: 信息流中的“信息量”是不可数的, 与先前可数的“信息数量”的表述相比, 更为贴切)。即, 分别用“单一信息量”、“中等信息量”、“丰富信息量”表示仅实时文字、实时文字+原始音频、实时文字+原始音频+图像这三种信息量下的自我介绍视频。这样, 既明确了目标人物信息量变化的操纵方法, 还真实反映了信息量的变化, 也避免了“低、中、高信息数量”这样的表述给读者造成理解上的歧义, 带来认知负荷联想。对此, 我们已将研究设计及题目、摘要、前言、研究方法、研究结果及讨论部分的相关表述都进行了修改。

新增参考文献:

Freiherr, J., Lundström, J. N., Habel, U., & Reetz, K. (2013). Multisensory integration mechanisms during aging.

Frontiers in human neuroscience, 7, 863. <https://doi.org/10.3389/fnhum.2013.00863>

Jones S. A., & Noppeney U. (2021). Ageing and multisensory integration: A review of the evidence, and a

computational perspective. *Cortex*, 138(7), 1-23. <https://doi.org/10.1016/j.cortex.2021.02.001>

Stevenson, R. A., Nelms, C. E., Baum, S. H., Zurkovsky, L., Barense, M. D., Newhouse, P. A., & Wallace, M. T.

(2015). Deficits in audiovisual speech perception in normal aging emerge at the level of whole-word

recognition. *Neurobiology of Aging*, 36(1), 283–291. <https://doi.org/10.1016/j.neurobiolaging.2014.08.003>

意见 2：上述主要问题之外，另有一些建议。

标题和关键词：这篇论文关注的是老年人和青年人在人格判断准确性上的差异，即人格判断的老化效应，但标题和关键词使用的词语分别是“年龄特点”和“年龄差异”。鉴于《心理学报》是一份面向心理学所有领域的期刊，“年龄特点”和“年龄差异”这样的用词涵义较为含糊，不清楚是儿童/青少年、中老年人还是多年龄段比较，建议在用词上明确老年人这一研究对象，有助于读者快速捕捉论文的核心内容。

回应：感谢专家的建议。的确，本研究主要关注的是老年人和青年人在人格判断准确性上的差异，使用“老化效应”和“老化特点”会比“年龄特点”和“年龄差异”更能体现研究对象和研究关注的核心问题。根据您的建议，已将标题修改为：人格判断能力的老化：信息量的调节作用。同时，将原关键词“年龄差异”修改为“老化特点”。

意见 3：摘要：摘要部分开始就是研究方法，有点不易理解。建议以 1-2 句简短的背景介绍开头，介绍核心概念或明确研究问题，再介绍方法，这样摘要部分背景-方法-结果-结论的结构完整，有助于读者理解。

回应：感谢专家的建议。根据您的意见修改后，我们在摘要部分先简单介绍研究背景，然后将研究目的（增加了对核心概念的界定）与研究方法融合在一起介绍，最后呈现研究结果和所得结论。具体如下（亦可见正文摘要部分）：

人格判断对个体社会功能的发挥具有重要影响。本研究采用 2（判断者年龄组：老年、年轻） \times 3（目标人物信息量：单一、中等、丰富）被试间实验设计，考察了信息量对老年人和年轻人人格判断准确性的影响，包括普遍准确性（对目标人物所在群体人格共性判断的准确性）和特有准确性（对目标人物个体人格判断的准确性）。结果显示，老年人人格判断的普遍准确性高于年轻人，且信息量的多少并不影响这一优势。信息量只影响特有准确性判断的年龄差异：单一信息量（仅实时文字）时，老年人的特有准确性低于年轻人；中等（实时文字和声音）和丰富信息量（实时文字、声音和图像）时，二者表现相当。综上，老年人的人格判断能力总体保持较好，且能有效利用增量信息来弥补与年轻人在特有准确性上

的差距。

意见 4: 正文第一段:“另一方面,老年人拥有【更加】丰富的社会阅历……【更】有经验……【更多】……”,建议在“老年人”之前加上“与年轻人相比”,这样后面的一系列“更加”才有比较的基础。

回应: 感谢专家指出这一表述问题,目前已将对应部分内容修改为(亦可详见“1 前言”第一段):

但另一方面,老年人拥有比年轻人更加丰富的社会阅历,在识人方面更有经验,也掌握了更多关于不同群体人格特征的朴素知识,这些知识经验可能有助于其做出更准确的人格判断(Freund & Isaacowitz, 2014; Hess et al., 2005)。

意见 5: 自检报告问题 8 (实验材料共享)。建议作者考虑使用 OSF 平台或者心理科学数据银行 (<https://www.scidb.cn/psych>) 等开放数据平台共享实验材料。

回应: 感谢专家的建议,目前已将实验材料共享至心理科学数据银行

(<https://www.scidb.cn/psych>), 供专家和读者查阅。实验材料链接为:

<https://www.scidb.cn/anonymous/Wk5WYklm>

第二轮

审稿人 1 意见:

作者已经很好地回到了我提出的问题,也对我前一轮的意见做了非常详细的修改。本轮我只有以下两点小的建议,修改好后推荐发表。

意见 1: 表格 1 信息量较少,与研究问题也无关,建议删除,并将内容或数据以适当形式整合到正文中。

回应: 感谢您的这一修改建议。根据您的建议,我们已将表格 1 删除,并将表中年轻判断者和老年判断者的背景信息整合到了“2.2 被试”的正文当中,具体修改如下(亦可见正文“2.2 被试”第二段):

年轻判断者和老年判断者的家庭人均月收入(年轻: $M \pm SD = 2.02 \pm 1.07$, 老年: $M \pm SD$

= 2.00 ± 0.88 ; $t(161) = 0.16, p = 0.874$)、自评健康(年轻: $M \pm SD = 3.80 \pm 0.94$, 老年: $M \pm SD = 3.84 \pm 0.75$; $t(161) = -0.26, p = 0.795$)和内隐人格观(年轻: $M \pm SD = 0.32 \pm 0.47$, 老年: $M \pm SD = 0.26 \pm 0.44$; $t(161) = 0.81, p = 0.418$)均不存在显著的年龄差异。老年判断者的受教育年限($M \pm SD = 11.42 \pm 2.59$)显著低于年轻判断者($M \pm SD = 15.94 \pm 2.11$), $t(161) = -12.21, p < 0.001$, Cohen's $d = 1.914$, 95%CI = [-5.25, -3.79]。

同时,我们在“2.3.3 背景变量测量工具”部分补充了家庭月收入 and 自评健康的评分标准,具体如下(亦可见正文“2.3.3 背景变量测量工具”第一段):

(一) 社会人口学信息。包括年龄、性别、受教育年限、家庭人均月收入(1=5000 元以下, 2=5000-10000 元, 3=10000-15000 元, 4=15000-20000 元, 5=20000 元以上)和自评健康状况(5 点评分, 1=非常不健康, 5=非常健康)。

意见 2: 内隐人格问卷的信度系数是在中国文化背景下其他研究的数据,还是在本研究中?如果是其他研究的数据,在本研究的信度系数是多少?在这里要明确说明。同样的问题还存在于“中文版十题人格量表”。

回应:感谢您对此所提出的疑问和建议。对于内隐人格信念问卷,我们将之作为背景变量进行了测量,要求被试在观点 A(主张人格实体论,即人格决定行为)和观点 B(认可人格渐变论,即环境决定行为)中选择更为认可的观点(0=人格实体论, 1=人格渐变论)。由于本研究采用的只有 1 道迫选题而非包含多个题项的问卷,因此,无法计算内隐人格信念问卷的内部一致性系数。感谢专家指出,先前文中所报告的中国文化下的内隐人格信念问卷的信度系数为其他研究的数据,且并非采用和本研究一样的迫选题形式。慎重考虑后,我们决定删除先前关于信度系数的表述,以避免给读者造成误导。

对于中文版十题人格量表,我们先前是基于李金德(2013)对中文版十题人格量表的信效度检验所得的数据,五项人格的内部一致性系数($n = 830$ 人)均达到了 0.60-0.67,效度检验也达标,据此我们认为该量表可作为在中国文化下测量大五人格的可靠和有效的工具。在本研究中,我们基于预研究和正式研究共计 276 名被试(131 名年轻被试和 145 名老年被试)的数据,计算得到外向性、宜人性、尽责性、神经质和开放性的内部一致性系数分别为: 0.51、0.24、0.59、0.67、0.54。相对而言,只根据 131 名年轻被试计算得到的内部一致性系数更高些,分别为: 0.71、0.27、0.68、0.67、0.62(除宜人性维度外,其他四个维度信度系数均达到了 0.62 以上)。实际上,十题人格量表(英文原版)上述五个维度的内部一致性系数($n = 1704$ 人)分别为 0.68、0.40、0.50、0.73、0.45,宜人性、尽责性、开放性三个维度

的内部一致性都算不上高。

关于十题人格量表的内部一致性系数相对研究中常见问卷的内部一致性系数结果偏低的问题，我们在查阅心理测量相关书籍以及和心理测量领域专家讨论后，是这样理解的：内部一致性系数评价的是测验内不同项目测量同一能力或特质的程度。如果所设计的测验要测量若干特质的话，得到的信度估计值可能会偏低。虽然十题人格量表共包含十个题目用于测量人格特质，但这十个题目测量的是五个不同的人格特质，这些人格特质彼此之间并不一定会存在高相关。因此，可能不能将这十个题目放在一起算整体的内部一致性系数，或者无法说清一个整体的内部一致性系数结果意味着什么。当然，我们可以分五个人格特质（每个特质 2 题）去计算内部一致性系数，但由于每个维度下的题目数量有限，计算得到的每个维度的内部一致性系数通常也不会特别高。而重测信度估计往往被用于评价同一个测验在两个不同时间实施时所伴随的误差，适用于评价不会随时间而改变的“特质”或特性的测量。不论对于年轻人还是老年人，人格特质都是较为稳定的心理特质。对于人格测验的信度计算，关注测验能否稳定地测量出一个人的人格特质可能更重要，因此，许多人格测验通常使用重测信度报告信度系数。因此，我们补充了十题人格量表的重测信度。

基于上述测量学的考虑，我们共招募了 91 名被试，包括 59 名 18-22 岁年轻人（ $M \pm SD = 19.12 \pm 0.65$ ，女性 33 名）和 32 名 60-76 岁老年人（ $M \pm SD = 67.47 \pm 3.81$ ，女性 17 名），间隔 1 周，进行了重测信度分析。结果显示，中文版十题人格量表的重测信度如下：外向性为 0.84，宜人性为 0.64，尽责性为 0.75，神经质为 0.71，开放性为 0.67。可见，五项人格的重测信度均达到了良好及以上。同时，本研究中 24 名目标人物在作为备选人物时，先使用了中文版十题人格量表进行人格自评。在正式参与自我介绍视频拍摄后，又使用第二版大五人格量表(BFI-2)再次进行自评(Soto & John, 2017; Zhang et al., 2022)。结果发现，目标人物两次自评的外向性($r = 0.804, p < 0.001$)、宜人性($r = 0.407, p = 0.048$)、尽责性($r = 0.687, p < 0.001$)、神经质($r = 0.775, p < 0.001$)和开放性($r = 0.442, p = 0.031$)分数均显著正相关。这表明使用中文版十题人格量表测量各项人格时个体能对自己人格水平进行有效的自我报告。鉴于在实际应用中，短小精悍的量表可以又快又准地完成测量，极大提高研究效率。特别是在本研究中，判断者需要对 24 名目标人物的人格进行判断。若判断者对每名目标人物均采用 60 题的大五人格量表（BFI-2）进行判断，不仅总体耗时长，而且到了每轮判断后期，判断者可能已难以回忆目标人物的表现而难以进行准确判断。根据上述分析和利弊权衡，我们认为中文版十题人格量表是最符合本研究需要的可靠且有效的人格量表。

另外，我们在进行中文版十题人格量表的重测信度测量过程中，也对迫选题形式的内隐

人格信念问卷进行了重测信度测量（我们认为一个人的内隐人格信念也是相对稳定的）。结果显示，基于同样的 91 名被试（包括 59 名 18-22 岁年轻人和 32 名 60-76 岁老年人）间隔 1 周的重测信度测量，该问卷的重测信度为 0.54，达到了可接受的水平。

根据上述分析结果，我们在正文中对中文版十题人格量表的信度进行了补充报告，具体如下（亦可见“2.3.2 中文版十题人格量表(TIPI-C)”）：

通过招募 59 名 18-22 岁年轻人（ $M \pm SD = 19.12 \pm 0.65$ ，女性 33 名）和 32 名 60-76 岁老年人（ $M \pm SD = 67.47 \pm 3.81$ ，女性 17 名）共计 91 名被试，间隔 1 周，进行重测信度测量。结果显示，该量表五项人格的重测信度均达到了良好及以上：外向性为 0.84，宜人性为 0.64，尽责性为 0.75，神经质为 0.71，开放性为 0.67。此外，李金德(2013)对该量表的信效度检验表明，中国文化背景下该量表内部一致性信度、重测信度、效标效度均达到了可接受水平，是测量大五人格可靠有效的工具。

类似的，我们对内隐人格信念问卷的信度也进行了补充报告，具体如下（亦可见“2.3.4 背景变量测量工具”第二段）：

基于 59 名年轻人和 32 名老年人（同“中文版十题人格量表”重测信度检验被试）间隔 1 周的重测信度测量数据，得到该问卷的重测信度为 0.54，表明该问卷的信度达到了可接受水平。

相关参考书籍：

罗伯特·M·卡普兰，丹尼斯·P·萨库佐 著，陈国鹏，席居哲 等 译校.（2010）. *心理测验：原理，应用和争论*（第六版）（pp. 80-81, pp. 84-85）. 上海人民出版社.

审稿人 2 意见：谢谢作者对于审稿意见的回应和相应修改，我没有更多的疑问或建议。

回应：感谢专家先前对本研究所提出的所有宝贵意见和建议，也感谢专家对上一轮修改和回应的肯定！

编委意见：同意发表。

主编意见：该文经过多轮评审和修改，已经达到学报发表要求。同意发表。