

# 《心理学报》审稿意见与作者回应

题目：交互自然性的结构与作用（工程心理学专栏）

作者：曹剑琴，张警吁，张亮，王晓宇

---

## 第一轮

### 审稿人 1 意见：

该文章针对缺乏人机交互自然性的有效评估方法，设计了新的交互自然性量表。该研究选题有重要的意义，但是对量表的有效性验证过程中和论文写作有一些可以修改提高的地方。相信经过修改，能成为一篇优秀的工程心理学论文。

### 回应：

非常感谢您的宝贵建议，相应地，我们对正文进行了逐一修改，并将修改内容进行了标红。以下是对各建议的逐条回复：

### 检验量表信效度方面：

#### 意见 1：

研究 2 中，在进行探索性因素分析时，建议说明为什么选择了平行分析和正交旋转来抽取因子个数而不是常用的主成分分析和 Promax 斜交旋转的方法的原因。

### 回应：

根据您的意见，我们在正文中“3.4.1 因子分析结果”中进行了相应的补充(p.19)。

首先，我们仍然使用的是主成分分析来做提取。只是在确定因子个数的时候，我们没有依赖特征根大于 1 的经验标准或碎石图检验，而是使用了更加稳健的平行分析。原因在于(1)仅仅是随机噪音就能造成特征根>1 的情况，使得特征根>1 这一标准对变量的数目比较敏感而不稳定；(2) 碎石图在曲线平滑或具有多个拐点难以决定且不够客观(Ledesma & Valero-Mora, 2007)。

相比之下，平行分析先生成大量随机数据集（数据集的条目数量和被试数量与要分析的实际数据是一致的），然后对每个模拟数据集做主成分分析，将所提取的特征根取平均值，就形成了一个模拟数据集特征根的分佈（见图 1 虚线）。可以看到，这个随机数据也能生成不少特征根大于 1 的因子，所以单纯提取特征根大于 1 的因子将引入随机误差。而选取特征根大于平行分析特征根的因子，则能够加好避免这一问题（图中因子 1 和因子 2）。平行分析通过保留特征值高于模拟数据集的所有因子，总体而言，其标准更加合理其可靠性优于特征根>1 和碎石图检验(Crawford et al., 2010; Hayton et al., 2004)。

第二，在因子旋转时，我们采用了正交旋转，因为我们的期望是找到尽可能相互独立的几个因素而不是相关性比较强的因素。一般来说，正交旋转和斜交旋转产生的结果都很相似，尤其在双因素情况下的表现更是如此(Finch, 2006)。在分析本数据集时，我们也发现了二者结果整体上没有明显差异(见表 1)。由于人机交互领域在开发有关测量工具时(e.g. Forster et al., 2020; Reyes et al., 2020)通常汇报正交旋转的结果，故而本论文选择报告正交旋转的结果。

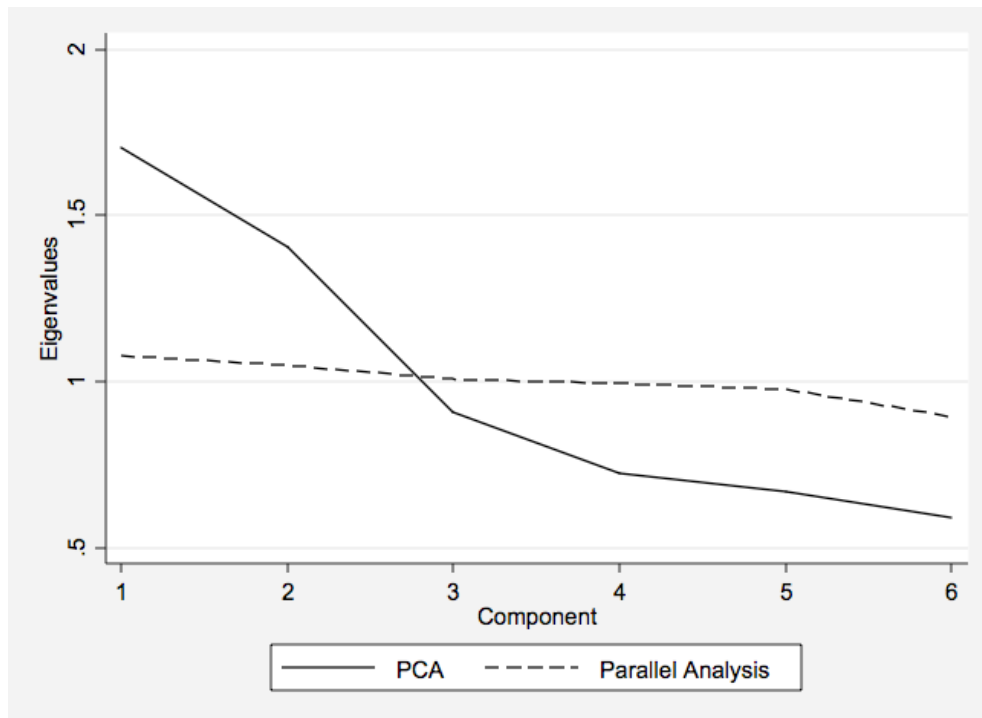


图 1

表 1 正交旋转和斜交旋转结果对比

条目	因子 1: 通达舒畅		因子 2: 随景应人	
	正交旋转	斜交旋转	正交旋转	斜交旋转
1. 流畅的	.68	.80		
2. 自然的	.61	.70		
3. 可学习的	.45	.44		
4. 愉悦感	.43	.41		
5. 合理性	.42	.39		
6. 符合预期	.41	.36		
7. 普适性			.58	.71
8. 情景适宜性			.45	.38
9. 人性化			.42	.38
方差解释百分比	20.0	21.4	12.0	10.6

注：负载小于 0.3 都未显示

**意见 2:**

研究 2 中,在报告探索性因素的结果时,建议报告是否存在具有明显跨因子载荷的条目。

**回应:**

感谢您的意见。研究 2 的探索性因子分析中,不存在明显的跨因子载荷的条目。在正文中“3.4.1 因子分析结果”中进行了相应的补充 (p.19)。

**意见 3:**

不适宜采用同一个样本对量表先后进行探索性因素分析和验证性因素分析。建议在研究 3 中另外报告验证性因素分析的结果。

**回应:**

感谢您的意见。已经删除了研究 2 的验证下因素分析(p.18), 并在研究 3 中单独报告了验证性因素分析的结果(p.23)。

**意见 4:**

研究 2 中, 在分析效标关联效度时, 建议补充自编量表总分与各个效标之间的相关系数。

**回应:**

非常感谢您的意见。已经在表 3 (p.19)和表 5 (p.23)中增加相关的内容。

**意见 5:**

研究 2 中, 对于检验区分效度的方法存在疑问。通常使用比较两因子的平均方差抽取量 (AVE) 和最大共同方差 (MSV) 的大小、两因子间的相关系数与两因子各自与总分的相关系数的大小, 这两条标准来检验量表的区分效度。该论文的研究 2 却是使用了层次回归分析的方法来检验区分效度, 建议说明使用这种方法的原因和文献依据。

**回应:**

非常感谢您的意见。尽管有少部分研究使用了层次回归分析中  $R^2$  的变化量来检验区分效度(Reidy & Keogh, 1997; 游旭群 等人, 2009), 我们认为正文中的区分效度更正为预测效度更加合适。作为校标效度和区分效度的补充, 预测效度不仅验证了校标与因子的预测关系, 还验证了交互自然性区分于可用性能解释更多的方差。

根据您的建议, 在区分效度中, 我们计算每个维度的方差提取量 AVE 值是否大于该维度与其它维度的相关方差, 如果 AVE 值大于两维度间的相关系数平方, 则表示这两个维度具有较好的区分效度。详细修改见正文 p.20。

**意见 6:**

在报告验证性因素分析的结果时, 建议进一步报告单因子模型和双因子模型的 $\chi^2$ 、df、 $\chi^2/df$ 、SRMR、AIC、BIC 等指标。

**回应:**

感谢您的建议, 已经在正文中 p.23 增加了相关指标的报告。

**意见 7:**

建议对自编问卷进行条目可理解性检验。

**回应:**

非常感谢您的建议, 已经增加“3.4.1 条目可理解性检验”(p.19-20)。

**意见 8:**

建议对自编问卷进行项目分析, 可以采用题总相关系数和 27%高低分组决断值的方法。

**回应:**

非常感谢您的建议, 已经在“3.4.2 项目分析”增加了相关内容。

**意见 9:**

建议对自编问卷进行信度分析, 包括: 内部一致性信度、重测信度和分半信度。

**回应:**

感谢您的建议, 已经在研究 2 的 3.4.4 (p.20)和研究 3 的 4.4.2 (p.23)分别补充了信度分析。

**写作规范方面：**

**意见 10：**

在报告使用的研究工具时，建议介绍量表名称，介绍评分最大和最小数字的含义，并报告最终的结果解读方式，比如，研究 2 中的关键消费行为和传统可用标准。

**回应：**

非常感谢您的意见。已经在介绍研究 2 和研究 3 研究工具时增加了相应的描述(p.19、p.22)。

**意见 11：**

在介绍研究工具时，建议明确说明哪些变量是效标。

**回应：**

非常感谢您的意见，已经在介绍研究工具时明确指出关键消费行为量表和传统可用性量表所涉及到的变量是校标变量(p.19、 p.22)。

**审稿人 2 意见：**

针对现有研究缺乏测量人机交互自然性的有效工具的问题，该论文描述了一个包含三个研究步骤，旨在构建一个测量自然性的用户体验的用户主观报告测评工具的研究。研究者在第一个研究步骤里，采用了文献与访谈的方法，确定了描述自然性的 9 个条目，形成了自然性的量表；在第二个研究步骤里，采用了问卷法，结合因子分析和效度检验的方法，得到了自然性量表的两个主要维度，并初步验证了量表的效度；在第三个研究步骤里，在新的样本里，采用与研究步骤二相似的方法，进一步验证了自然性量表的结构的有效性及其效度。

整体而言，该论文对相关领域的文献做了较好的综述与讨论，指出了现有研究与实践中缺乏对自然性体验的有效测量工具的问题。解决这个研究问题对工程心理学、用户体验设计有一定的理论与实践意义。研究符合《心理学报》的征稿范围和读者的兴趣。论文自检报告所提及的内容与论文正文未见冲突之处，整体行文也大致符合相关撰稿要求。

我的意见是该论文如果经过一些必要的修改，有望在《心理学报》上发表。在修改过程中，请作者们回答以下问题：

**意见 1：**

研究 1 中，文献回顾与总结部分，请更清楚地说明检索的过程。文中列举了一系列的关键词，所列举的是搜索时使用的所有关键词吗？这些关键词在搜索时的逻辑关系是什么？仅根据文中的描述，我们无法重复搜索结果。

**回应：**

非常感谢您的意见，文中列举的关键词是检索时使用的所有关键词，搜索时的逻辑关系，更加详细的检索、筛选和综合过程已经在研究 1 的 2.2.2(p.16)进行了相应的补充。

**意见 2：**

根据我的理解，该研究只通过专家审阅和修改的方法来进行量表条目的评价。传统的量表构建研究会采用专家法的同时，采用量化的方法，如内部一致性或结构方程，来对条目进行筛选。请作者解释为什么在本研究里没有使用这些方法。

**回应：**

感谢您的意见，我们实际上是进行了这些分析，并报告了结构方程的结果，只是遗漏了内部一致性系数的汇报，因为这些指标都已经达到了基本要求。

因此，我们在研究 2 中增加了对量表内部一致性系数检验(见 3.4.4, p.20)，增加了条目的可理解性检验和项目分析(见 3.4.1-3.4.2, p.20)，结果发现量表的条目易理解，且具有良好的鉴别水平。通过探索性因子分析发现条目在两个因子上都没有双重载荷，无需删除条目。通过结构方程模型分析发现，双因子结构比单因子结构的拟合指标更好( $\Delta\chi^2 = 11.5$ ,  $\Delta df = 1$ ,  $p < .001$ )。

**意见 3:**

量表的形式是 5 点李克特量表吗？研究者如何确定这种形式？

**回应:**

非常感谢您的意见。所开发的交互自然性量表是 5 点李克特量表。依据有 2 点： 1)以往研究中关于自然性单条目的测量中常使用 5 点评分(Almeida et al., 2019; Bailey et al., 2018; Hsiao, 2017); 2)不同的李克特量表不会直接给统计结果带来明显的差异，在进行分析时可以对不同的评分进行标准化，使它们的系数成为统一化(Aiken, 1987)。

**意见 4:**

研究 2 确定了量表为双因子结构。双因子之间的相关性如何？量表是否为不相关的双因子结构？如果双因子之间相关性较低，对自然性这个概念来说意味着什么？

**回应:**

非常感谢您的意见。我们的分析发现，这两个双因子的相关较低( $r = .19$ ,  $p < 0.001$ )，这表明两因子是独立的(见 3.4.5, p.21)。这两个因子之间的关联性较弱，说明在自然性这个问题上其实存在不同的组成部分，特别是我们发现的随境应人这个维度，不仅对消费者的购买意愿起到了独特的作用，其产生也更多地与交互产品的智能功能有关。这一发现超越了传统可用性维度，它的出现可能与产品的智能化、自适应和环境感知等技术的发展有关，也与新时代下人们对产品的需求有了更高的要求 and 期望有关。

.....

**审稿人 3 意见:**

本研究论文能在较为全面总结和梳理以往有关交互体验的评估方法、自然交互现状的基础上，构建出交互自然性量表，并通过两次大样本的用户调研，一方面修订量表工具，同时还能较为深入探讨其与可用性对于用户购买意愿和行为的影响关系，具有较好的科学意义和实践价值。总体上，本研究思路清晰，逻辑合理，方法较为得当，语言表述精炼。建议在以下方面进行完善：

**意见 1:**

研究 1 本身是对交互自然性的质性研究，表 1 中的字典文献总结中的条目描述第二点用的“自然性”，概念循环解释，建议用另一个能表达使用很顺手的条目或概念；

**回应:**

非常感谢您的意见。但这里我们的目的不是为了解释自然性本身，而是从常用的语言表达和用户心智中提取可用于测量的条目，所以“自然”一词实际上是如实地反映了我们调研得到的结果。必须要指明的是，这种做法在心理学量表构建过程中是难以避免。比如，在信任量表中总会用到“信任”一词，在情绪量表中也都会直接用“生气”、“愉快”来作为测量情绪体验的条目。因此，我们认为在测量自然性的条目中，保留“自然”一词是可以接受的。

### 意见 2:

第 7 页中提到的 287 篇文献提炼出 48 篇核心文献，希望解释核心文献的判定标准是什么？

### 回应:

感谢您的意见，已经详细在第 16 页解释了核心文献的筛选过程和筛选标准。

### 意见 3:

研究 3 中，交互自然性中的随景应人对推荐意愿预测良好，而对使用意愿预测效应不显著的解释似乎逻辑不清晰。

### 回应:

非常感谢您的意见。已经在正文中把解释移到了讨论 p.25 中，进行了相应的修改。

当用户在决定自己是否喜欢或者使用某个产品时，更关注功能流畅好用与否，与预期是否相符合，这些在某种层面上更多地体现为工具性动机的满足(Khalid et al., 2012)。但当用户决定忠诚于某个品牌或者将产品推荐给他人时，不仅看中产品对工具性的动机满足，还需要能满足某种情感性或象征性的动机(Helander et al., 2013; McDonagh & Lebbon, 2000)。随景应人这一维度，反映了交互产品能够理解用户的意图并提供人性化服务的程度，在目前的阶段，它可能更多地与科技感、智能感、人性化设计这些概念联系在一起(Hsiao, 2017)，从而给消费者带来了更多不同动机的满足。

### 意见 4:

本研究在前面介绍时提到选用智能网联汽车作为使用产品的意义，但在后面的描述中，却很少看到交互自然性在网联汽车上的关联性，比如汽车哪些功能和特点与用户体验到的顺达通畅有关，哪些又与随景应人有关。应该有更深入地分析。

### 回应:

非常感谢您的意见。我们实际上是在研究工具中设计了这块内容，但我们的这篇文章，主要是围绕量表开发进行的工作，所以更加关心量表的构建方法，内在结构的合理性（结构效度）和作为一个工具的有效性（预测效度），对量表各维度的影响因素的探讨，本是我们计划在未来的研究中系统考察的内容。

但现在已经根据审稿人的要求增加了车机功能满意度与交互自然性的相关分析(见正文表 7, p.24)，并在研究工具(p.19)、讨论(p.25)中进行了补充说明。我们的基本发现是随景应人这一维度更多地与智能化的功能，如语音，自动泊车等有关。

## 参考文献

- Aiken, L. R. (1987). Formulas for equating ratings on different scales. *Educational and Psychological Measurement, 47*(1), 51–54.
- Almeida, L., Lopes, E., Yalçınkaya, B., Martins, R., Lopes, A., Menezes, P., & Pires, G. (2019). Towards natural interaction in immersive reality with a cyber-glove. *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 2653–2658.
- Bailey, S. K., Johnson, C. I., & Sims, V. K. (2018). Using natural gesture interactions leads to higher usability and presence in a computer lesson. *Congress of the International Ergonomics Association*, 663–671.
- Crawford, A. V., Green, S. B., Levy, R., Lo, W.-J., Scott, L., Svetina, D., & Thompson, M. S. (2010). Evaluation of parallel analysis methods for determining the number of factors. *Educational and Psychological Measurement, 70*(6), 885–901.

- Finch, H. (2006). Comparison of the performance of varimax and promax rotations: Factor structure recovery for dichotomous items. *Journal of Educational Measurement*, 43(1), 39–52.
- Forster, Y., Hergeth, S., Naujoks, F., Krems, J. F., & Keinath, A. (2020). Self-report measures for the assessment of human-machine interfaces in automated driving. *Cognition, Technology & Work*, 22(4), 703–720.
- Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, 7(2), 191–205.
- Hsiao, K.-L. (2017). What drives smartwatch adoption intention? Comparing Apple and non-Apple watches. *Library Hi Tech*, 35(1), 186–206. <https://doi.org/10.1108/LHT-09-2016-0105>
- Ledesma, R. D., & Valero-Mora, P. (2007). Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out parallel analysis. *Practical Assessment, Research, and Evaluation*, 12(1), 2.
- Reidy, J., & Keogh, E. (1997). Testing the discriminant and convergent validity of the Mood and Anxiety Symptoms Questionnaire using a British sample. *Personality and Individual Differences*, 23(2), 337–344.
- Reyes-Fournier, E., Cumella, E. J., Blackman, G., March, M., & Pedersen, J. (2020). Development and Validation of the Online Teaching Effectiveness Scale. *Online Learning*, 24(2), 111–127.
- 游旭群, 姬鸣, 戴鲲, 杨仕云, 常明. (2009). 航线驾驶安全行为多维评价量表的构建. *心理学报*, 41(12), 1237–1251.

---

## 第二轮

### 审稿人 1 意见:

作者对之前提出的意见进行了很好的回应, 没有特别的意见了, 对作者有两个小建议, 修改之后建议发表。

### 回应:

非常感谢您的宝贵意见, 相应地, 我们对正文进行了修改, 并将修改内容进行了批注。以下是对各建议的逐条回复:

### 意见 1:

建议将本研究使用的“传统可用性量表”和“关键消费行为量表”的具体量表内容作为补充材料放在文章的末尾。

### 回应:

根据您的意见, 我们在正文增加了附录, 包括了“交互自然性量表”、“传统可用性量表”和“关键消费行为量表”的具体测量内容(p.27-28)。

### 意见 2:

请作者再仔细校对文稿, 引用正确的表格名字, 比如 23 页的 4.4.4 章节的第一句“结果发现(表 8)”, 但是文中不存在表 8。

### 回应:

非常感谢您的意见, 已经对文稿进行了仔细校对, 您提到的内容也已经更正(p.24)。

.....

### 审稿人 2 意见:

文章的修改稿已经解答了我的所有疑问。修改后的文稿表述清晰, 研究方法、结果、结

论恰当可信。建议发表。

回应：

非常感谢您的意见和认可。

---

### 第三轮

编委意见：

我看了三位专家对该文章的评价意见，和作者对评审专家意见的回复意见。我觉得三位评审专家对文章都做出了正面评价，对作者的回复意见也基本认可。我同意评审专家的意见。同意该文章发表。

回应：

非常感谢您的意见和认可。

主编意见：

学报原投稿指南上明确指明“不接受单纯的量表编制报告”。现在的投稿指南上说：本刊不接受单纯对西方量表修订汉化的报告，接受原创构念和理论的量表编制文章。其本意是：不接受为单纯为编制量表而发表的文章（学报不为“量表”背书），但接受“为回答一个有意义的科学问题而编制量表”的论文。因此，建议作者从文章题目到前言部分均做出修改。旨在说明，您研究的问题是一个很有意义问题，为回答这个有意义的科学问题，须开发、编制一个新的量表。

回应：

非常感谢您的宝贵意见。我们进一步将科学问题提炼为：探索交互自然性的心理结构和作用机制。为了回答这一问题，我们借助于心理测量学方法，结合定性和定量研究，先通过文献分析、词典分析和专家访谈，搜集相关的概念，并通过大规模问卷调查，结合信度分析、结构效度分析和校标关联效度分析来探讨其稳定性，结构特征和作用机制。我们对文章题目、摘要、前言和讨论都做了相应修改。如题目修改为“交互自然性的心理结构及其影响”，同时在中英文摘要(p.14-15)、引言(p.17)、讨论(p.26)和结论(p.28)中都进行了相应的修改。