

《心理学报》审稿意见与作者回应

题目：幼儿友好型内隐联想测验的建构及有效性研究

作者：钱淼 周立霞 鲁甜甜 翁梦星 傅根跃

第一轮

审稿人 1 意见：

意见 1：由于 Banaji 的 Ch-IAT 的被试年龄最小为 6 岁，研究者设计了新的程序 PSIAT 以评估幼儿的內隐态度，研究通过面孔两可分类任务作为效标来确定其方法的有效性。从研究被试来看，研究包含了 6 岁儿童，如果将 PSIAT 和 Ch-IAT 进行关联分析，可能更有说服力。

回应：感谢审稿专家提出的这一宝贵想法，将 Ch-IAT 范式作为效标，与本研究改编的范式进行关联分析，确实更加具有可比性和说服力。本研究构思时只想到模仿“Ch-IAT 的效度证实方法”，即通过花虫內隐态度的结果来证实效度，同时也在种族态度的测量时将面孔两可分类任务作为效标，没有想到直接以 Ch-IAT 范式作为效标。遗憾的是研究中的大班被试已经升学，所以无法找回以追补这一结果。

意见 2：就研究所建构的 PSIAT 来说，还需要探讨该方法的心理计量学特征，例如内部一致性、稳定性；如果能够提供该方法测量结果与被试外显测量或者行为方面的指标的关系，则更为全面。

回应：根据审稿专家的意见，我们在实验一的结果部分加入了心理计量学的特征。参照 Cvencek, Greenwald 和 Meltzoff (2011) 的研究方法，实验一花虫 PSF-IAT 的研究中，我们将每个被试的所有试次进行奇偶分半，并计算出奇数试次和偶数试次的內隐态度 D 值，求出内部一致性系数，克伦巴赫 alpha 系数=0.736；实验二內隐种族态度 PSF-IAT 的研究中，克伦巴赫 alpha 系数=0.713。

另外，诚如专家所建议，本研究实际上在实验一花虫 PSF-IAT 和实验二內隐种族态度 PSF-IAT 测得被试的外显态度，想用外显态度的结果作为一个效标，来佐证范式的效度和可行性。但遗憾的是实验一和实验二的內隐态度和外显态度结果均无显著相关。因此，我们并未报告这部分结果。但是现在经过慎重考虑后，我们觉得內隐态度的测试取得了与外显态度测试相一致的结果，即，偏爱花（实验一）或者偏爱本族人（实验二），从这点上可以证明 PSF-IAT 方法的有效性。至于两者间相关不高的原因可能是外显态度较易受到社会赞许度和社会规范的影响(Greenwald, McGhee, & Schwartz, 1998; Greenwald, Nosek, & Banaji, 2003; Rutland et al., 2005)。我们在慎重考虑审稿专家意见之后，觉得这部分测试结果还是有意义和价值的。因此，我们补充报告了外显态度测试的结果。

审稿人 2 意见：

意见 1：作者在引言中阐述开发幼儿友好型 IAT 的原因和意义时提到，前人的“Ch-IAT 最小只能做到 6 岁儿童。显然 6 岁并不是态度产生的起始年龄，因此，开发和设计出适合更小年龄儿童的內隐联想测验，对于解决有关态度的起源与发展问题具有极为重要的意义。”然而，前人基于 IAT，已有文章测量过 6 岁以下儿童的內隐态度，比如，Thomas, Burton Smith 和 Ball(2007)采用适应性 IAT（全部图片刺激）测量 3-7 岁儿童的花-虫內隐态度和胖-瘦內隐态

度。再比如, Cvencek, Greenwald 和 Meltzoff (2011) 专门开发了幼儿内隐联系测验 (Preschool Implicit Association Test, PSIAT) 用以测量 4 岁儿童的内隐态度。建议作者首先对幼儿内隐态度测量的相关文献进行更为全面清晰的综述, 然后重新阐述开发幼儿友好型内隐态度测验的原因和意义。

回应: 感谢审稿专家提出的这一重要问题, 并且推荐了 Thomas, Burton Smith 和 Ball (2007), Cvencek, Greenwald 和 Meltzoff (2011) 这两篇重要的文献。我们认真地研读了这两篇文献, 并且重新阐述了开发幼儿友好型内隐态度测验的原因和意义。Thomas, Burton Smith 和 Ball (2007) 采用改编后的 IAT 范式 (Adapted IAT), 实现刺激材料的图片化呈现, 用于测量 3-7 岁儿童的花-虫内隐态度和胖-瘦内隐态度。Cvencek, Greenwald 和 Meltzoff (2011) 专门开发了幼儿内隐联系测验 (Preschool implicit association test, PSIAT) 用以测量 4 岁儿童的花-虫态度和内隐性别态度。这些改编后的幼儿内隐联想测验在幼儿态度研究中取得了一些研究成果, 但是我们质疑其适用性, 认为其存在一定的缺陷。Adapted IAT 和 PSIAT 范式虽然在刺激呈现方式、呈现次数和界面设计上做出了适应性的改变, 但是仍然忠实于 IAT 范式, 保留了其 7 个 block 的特点。在每个 block 的任务中, 要求被试记忆不同的规则, 而且, 在不同组块间, 被试需要做多次任务转换。根据幼儿的认知发展特点和我们的实验经验, 我们认为幼儿难以很好地理解并执行规则。因此, 本研究拟开发和设计出最大限度减少幼儿认知负担, 适应幼儿认知和行为特点的内隐态度测验, 为解决有关幼儿态度的起源与发展问题提供新的工具。

意见 2: 作者提到其开发的幼儿友好型内隐联系测验是基于 IAT 的基本原理的, 简化程序后的 IAT。然而, 审稿人认为作者开发的测验并不符合 IAT 的原理, 本文作者设计的测验和 IAT 在测量内容和实验程序上都有较大的差异。IAT 测量的是目标概念和属性概念间的内隐联系, 实验程序是将四种概念 (一对目标概念和一对属性概念) 的代表性样例作为刺激词, 并用两个键进行目标概念-属性概念的联合分类任务, 即以目标概念和属性概念为分类标签 (呈现在屏幕的上方) 进行分类。作者设计的测验则是测量的目标概念和笑脸或哭脸的联系, 实验顺序是将其中一个目标概念的代表性样例用笑脸键反应, 另一个目标概念的样例则用哭脸键反应, 即直接以笑脸和哭脸作为分类标签对刺激图片进行分类。作者设计的 PSIAT 测量的与 IAT/Ch-IAT 测量的是否是同一个结构?

回应: 诚如专家所言 IAT 测量的是目标概念和属性概念间的内隐联系, 实验程序是将四种概念 (一对目标概念和一对属性概念) 的代表性样例作为刺激词, 并用两个键进行目标概念-属性概念的联合分类任务, 即以目标概念和属性概念为分类标签进行分类。其测试过程的本质是为了获取一对目标概念与属性概念交叉配对后的反应时, 以测量被试对目标概念的内隐态度倾向。以往的 IAT 均涉及目标概念 (花/虫)、属性概念 (消极/积极) 及按键 (左/右) 三种不同的刺激, 因此一般需要 7 个组块的操作, 其中涉及到多项规则与多个任务转换。本研究设计的 PSF-IAT 中目标概念与以往 IAT 相同 (如, 花、虫), 而最大的改变是将属性概念与按键合并, 即笑脸和哭脸既是属性概念又同时作为反应按键, 以获取反应时。这样不但能继续进行目标概念-属性概念的交叉联合分类, 而且使任务难度大大降低, 但其本质与原 IAT 相同, 同样能够获得一对目标概念与属性概念交叉配对后的反应时。因此, 我们认为所设计的 PSF-IAT 测量的与 IAT/Ch-IAT 测量应该是同一个结构。从我们将该范式运用于花虫以及种族态度的测量结果也可以得到初步佐证。鉴于采用不同形式的多个笑脸和哭脸与采用单一形式的笑脸与哭脸在实际上并没有区别, 因此, 我们又将属性概念的刺激简化为单一的笑脸和哭脸。我们认为本研究改编的幼儿友好型内隐联想测验最大的特点在于实现了属性概念和按键的结合, 笑脸和哭脸同时代表了属性概念和按键。

意见 3: 本文的测验将笑脸键和哭脸键位置随机化处理, 额外增加了个体的认知加工 (比如, 在选择按键前对按键刺激的识别, 对按键刺激位置的加工), 而增加的认知加工过程所需要的反应时间也包含在被试的反应时中, 可能会混淆内隐态度测验的指标。请作者说明为什么这样设计? 按键位置为什么不采用被试间平衡?

回应: 本研究的实验设计是将属性刺激和按键结合, 因此笑脸和哭脸同时具有双重功能。审稿专家提议笑脸和哭脸按键的位置可以采用被试间平衡的方法。鄙人认为如果将按键固定化呈现 (如笑脸呈现在左边, 哭脸呈现在右边), 让被试看到花按笑脸, 看到虫按哭脸, 被试会将规则简化为看到花按左边, 看到虫按右边, 失去了将花和虫与积极和消极属性配对的过程, 而变成了对概念进行左右按键分类的过程。如果是这样, 被试的内隐态度将无从测得。

审稿专家提到将笑脸哭脸键位置随机化会增加个体的认知加工, 从而增加反应时, 这个现象的确存在。根据 Cvencek, Greenwald 和 Meltzoff(2011)的反应时结果, 6 岁儿童在内隐种族态度测试中, 相容任务情况下的平均反应时为 1600ms, 不相容任务情况下的平均反应时为 1800ms; 而在本研究中大班儿童在内隐种族态度测试中, 相容任务情况下的平均反应时为 2483ms, 不相容任务情况下的平均反应时为 2930ms, 正如审稿人所说, 增加的认知加工过程所需要的反应时间确实包含在被试的反应时中。但是这一增加的反应时间不会混淆内隐态度测验的指标。因为, 内隐态度测验的指标是 D 值, 其计算方法是不相容任务的反应时间与相容任务的反应时间的差值比上总体标准差。这一增加的反应时间因同时存在于相容任务和不相容任务中, 所以在计算差值时可以相减掉。

意见 4: 本文在第二部分介绍 PSIAT 的实验过程过于简略, 应该将 PSIAT 的一个基本程序进行详尽地介绍, 以便其他有兴趣的研究者进行重复研究。同样地, 在两个实验研究中, 不仅要阐述实验程序, 还需要对实验刺激进行相应的说明, 比如, 图片刺激的尺寸, 图片背景如何处理, 是否统一, 照片刺激的吸引力程度是否进行前测, 是否进行匹配。最好正文中呈现一些刺激的样例。

回应: 感谢审稿专家二的宝贵意见, 我们在介绍 PSF-IAT 实验过程时确实过于简略。因此, 根据审稿专家的意见, 本文在实验介绍部分加入了具体实验程序, 对实验刺激材料进行详述, 并在文中将字体标红。同时, 加入了刺激的样例 (图 5)。

意见 5: 作为一篇开发新测量工具的研究, 检验信度和效度是必不可少的。信度指标, 可采用内部一致性, 重复测量一致性等指标; 效度指标, 包括实验内部效度, 内容效度, 预测效度, 区分效度, 聚合效度等指标。本文没有信度检验信息, 两个内隐态度实验也都没有进行外显态度的测量用以进行效度的检验, 更没有采用前人已有的幼儿内隐态度测量工具同时施测后进行比较, 此外, 没有设计一些合理的指标进行预测效度的检验。作为测量工具的开发研究, 本文所进行的心理测量学检验是不够充分的。作者需要提供进一步的信效度检验信息。

回应: 感谢审稿专家提出的这一点宝贵意见, 根据您的意见, 我们在实验一的结果部分加入了信度指标。我们报告了实验一中花虫 PSF-IAT 的内部一致性信度, 我们分别算得每个被试奇数次和偶数试次的 D 值, 计算克伦巴赫 alpha 系数 $=0.736$; 实验二内隐种族态度 PSF-IAT 的内部一致性信度, 克伦巴赫 alpha 系数 $=0.713$ 。

根据审稿专家的意见, 我们同时增加了一些效度的指标。1.内部效度, 我们借鉴 Cvencek, Greenwald 和 Meltzoff (2011)采用的 Cohen's d 值进行比较的方法, 结果发现, 实验一中花虫态度的效应值(effect size), $d=0.57$, 这与 Cvencek, Greenwald 和 Meltzoff(2011)的结果一致 ($d=0.50$)。2.效标效度和预测效度, 我们采用两可面孔分类任务和改编的范式同时测量幼儿被试对黑人的内隐种族态度, 发现两者存在显著的正相关, $r=0.320, p=0.003$ 。这一结果可以作为效标效度和预测效度的证据。3.内容效度, 一方面, 我们采用统计分析法(折半信度)对

内容效度进行评估；另一方面，我们在实验设计之初，曾将实验的构思（将属性概念和按键相结合）与两位国际专家和两位国内专家进行讨论，并得到他们的肯定。另外，我们已经将此范式用于非洲幼儿内隐态度等国际合作的 研究中，并将文章投稿于 *Child Development* 和 *Cognition*。目前文章在送审当中。

另外，审稿专家提到两个内隐态度实验都没有进行外显态度的测量用以进行效度的检验。关于这一点的回复见对审稿专家一意见 2 的回复。

总之，根据审稿专家的意见，本文在修改稿中保留了原来提供的测试有效性证据，如被试以高正确率通过两个测试，选取两可面孔分类任务作为效标取得较高相关等。另外还补充了内部一致性、态度的效应值等测量学指标。当然我们也知道，虽然 PSF-IAT 测试方法简单，尤其体现出其对于低龄幼儿的友好性，但测量结果的有效性还需要在今后不断的增加证据来进一步验证或修改。

意见 6: 本文在参考文献引用方面有些随意，有的文献正文中并没有引用，但是列在文末的参考文献中（如，Thomas, S., Burton Smith, R., & Ball, P. (2007). Implicit attitudes in very young children: An adaptation of the IAT. *Current Research in Social Psychology*, 13(7), 75-85.）；还有的文献被错误引用，详见文中批注。

回应: 谢谢审稿专家的意见，我们对此疏忽表示歉意，已对发现的错误进行更正。

意见 7: 将第五页第 2 部分（幼儿友好型内隐联想测验的建构）“和 Ch-IAT”去掉。

回应: 已将第五页第 2 部分（幼儿友好型内隐联想测验的建构）“和 Ch-IAT”去掉

意见 8: 第四页第 2 部分（幼儿友好型内隐联想测验的建构）第二段，审稿专家认为“从操作界面上看，屏幕上只呈现刺激和按键刺激，并没有像 Cvencek 等(2011)设计的 PSIAT 那样有清晰地按键规则和按键提示，界面设计还不够友好。并且对实验过程的介绍不够清楚，比如，指导语是怎样的，被试正确反应后呈现什么反馈信息，错误反应后呈现什么反馈信息，是要纠正反应才能继续还是直接进入下一个 trial 等等。反馈信息的呈现最好也用图片展示出来。”

回应: 在本研究 PSF-IAT 中，因实现了按键和属性刺激的结合（笑脸和哭脸兼有属性刺激和按键的双重功能），所以我们将 7 个 block 简化为 2 个 Block：相容任务和不相容任务。首先，因每个任务中被试的记忆规则极其简单，只需记住一个规则，如看到花按笑脸键，看到虫按哭脸键，所以无需在电脑屏幕中呈现按键规则和按键提示。其次，被试在完成一个任务后只需进行一次规则转换，实验结果发现被试通过 8 个 trial 的练习即可完成规则转换，这一结果可用正确率来进行佐证。然而，在 Cvencek 等(2011)设计的 PSIAT 中，他将属性刺激的提示和目标刺激的提示放置于屏幕两侧，并采用与按键颜色相同的两条宽条作为提醒的设计，是出于他实验本身的需要，被试需要完成多个 blocks，记忆多条规则，并且需要在规则间做多次转换，这不仅加重了幼儿的任务负担，而且对幼儿被试的记忆力、执行功能发展水平要求较高。为了减少对幼儿认知能力的要求，Cvencek 设置了友好界面，但在本实验中被试不需要这种提示即可完成 PSF-IAT。

感谢审稿专家提出了对实验过程描述的要求。根据意见，在文中鄙人在对实验过程的介绍中加入了相应的信息并标红，添加了一张含有反馈信息的完整实验程序图（图 2）。

意见 9: 在文中第六页第一段第四行解释联合标准差的含义

回应: 已在文中注明“总的标准差指的是相容任务和不相容任务所有试次的联合标准差 (pooled SD)”。

意见 10: 第六页第三段, Ch-IAT 范式的任务

回应: 第六页第三段, 重新介绍了 Ch-IAT 范式的任务, 使之更加忠于该范式。在 Ch-IAT 范式中属性刺激为积极或者消极词性的词语(如“开心”, “难过”), 通过录音方式呈现, 要求被试对这些积极或者消极词性的词进行左右按键反应。

意见 11: 第七页第二段, 审稿专家认为 Mierke 和 Klauer (2003)提出方法特异性变异, (Method-specific variance)会影响 IAT 的效应, 但是他们同时也在本篇文章中提到 Greenwald(2003)采用的新算法——D 分数, 可以在较大程度上降低这种方法特异性变异的影响。

回应: 本研究的方法由于实现了属性刺激和按键的结合, 使得实验任务不需对属性刺激进行左右按键反应, 也无需对概念刺激进行左右按键反应, 因此实验任务简化为两个 block: 相容任务和不相容任务。这种设计可以彻底的消除方法特异性变异的影响。方法特异性变异指的是在传统 IAT 中, 被试在相容和不相容任务中对指导语的再编码无法匹配。如, 在相容情况下, 要求被试看到花和积极的词汇按左键, 看到虫和消极的词汇按右键。此时, 被试易对指导语进行再编码来简化反应过程, 认为看到积极的东西(包括花)都按左键, 看到消极的东西(包括虫)都按右键, 此时对四类刺激做出反应的任务, 被简化为对两类刺激做反应; 在不相容的情况下, 这种简化的过程则不能实现, 被试依旧是对四类刺激做出相应的按键反应。而在本研究的实验设计中, 被试始终都是对两类刺激(黑人和中国人)进行反应, 因此可以从根本上消除方法特异性变异的影响。

意见 12: 第 12 页 3.5 讨论部分, 未引用“Cvencek, Greenwald, & Meltzoff, 2011”这篇文献。

回应: 已删除。

意见 13: 第 13 页 4.2 实验程序和实验材料部分, 中国人和黑人面孔图片的平均吸引力程度是否进行过控制?

回应: 我们已经对中国人和黑人面孔图片的平均吸引力程度进行控制。这种控制通过预实验, 用 30 名大学生从 40 张面孔中挑出的 20 张面孔吸引力相匹配的照片来作为实验材料。

第二轮

审稿人 1 意见:

意见 1: 作者在修改稿中对文献梳理部分和实验程序部分进行了补充, 对幼儿友好型内隐态度测验的介绍也更为清晰。然而, 作为一种新的幼儿内隐态度测验, 还有一些本质性的问题, 作者并没有很好地解决。尽管作者在介绍 PSF-IAT 类别的时候, 将“笑脸”和“哭脸”上升为“积极属性”和“消极属性”, 但是其分类标签、属性概念、属性刺激仍然是同一个, 没有区分开来。因此, 其包含的任务(如, 花按笑脸键, 虫按哭脸键)也与 Greenwald 等(1998)设计的测量内隐态度的 IAT 任务(如, 花和积极词按“F”键, 虫和消极词按“J”键)还是有较大的区别。从任务要求的内隐程度上来说, 作者设计的测验测量的态度内隐程度相对较低。也就是说, PSF-IAT 测量的是内隐态度吗?

回应: IAT 测量的是目标概念和属性概念间的内隐联系, 实验程序是将四种概念(一对目

标概念和一对属性概念)的代表性样例作为刺激词,并用两个键进行目标概念-属性概念的联合分类任务,即以目标概念和属性概念为分类标签进行分类。其测试过程的本质是为了获取一对目标概念与属性概念交叉配对后的反应时,以测量被试对目标概念的内隐态度倾向。以往的 IAT 均涉及目标概念(花/虫)、属性概念(消极/积极)及按键(左/右)三种不同的刺激,因此一般需要 7 个组块的操作,其中涉及到多项规则与多个任务转换。本研究设计的 PSF-IAT 中目标概念与以往 IAT 相同(如,花、虫),而最大的改变是将属性概念与按键合并,即笑脸和哭脸既是属性概念又同时作为反应按键,以获取反应时。这样不但能继续进行目标概念-属性概念的交叉联合分类,而且使任务难度大大降低,但其本质与原 IAT 相同,同样能够获得一对目标概念与属性概念交叉配对后的反应时。因此,我们认为所设计的 PSF-IAT 测量的与 IAT/Ch-IAT 测量应该是同一个结构。因此该范式测得的是内隐态度。

意见 2: 作者在二稿中没有提出更有说服力的证据来说明其设计的 PSF-IAT 与 CH-IAT/IAT 测量了同样的内隐态度。前人已发表 Ch-IAT 和 PSIAT 用以测量幼儿的内隐态度,作者期望设计新的 IAT 变式,必然需要与已有的、经过检验的幼儿内隐态度测验进行比较,来验证其与 IAT 测量内容的一致性。然而作者并没有进行相关的验证。

回应: 测量 Ch-IAT 与 PSIAT 范式与 PSF-IAT 范式的相关确实是有需要的,但是非常遗憾,这部分工作由于一些现实原因无法对原有被试重新实测。在本研究中,我们同时增加了一些效度的指标,如内部效度,内容效度,效标效度和预测效度。研究结果为测量工具的有效性提供了初步的证据。在今后的研究中,考察 Ch-IAT 和 PSF-IAT 范式的相关对于验证该范式非常有必要。

意见 3: 对于外显态度测量,作者以幼儿选择某一目标概念刺激(如,花)的百分比作为外显态度的指标,可以反映幼儿个体对该目标概念(如,花)的态度,比如与虫的照片相比,幼儿更多地选择喜欢花的照片,可以说明幼儿喜欢花比虫多。然而,这个百分比的差异并不适合反映不同幼儿间对花-虫喜欢程度的个体差异。因为看起来是 10 道迫选题(第一个实验),实际上合起来就是一个问题:你喜欢花的照片还是虫的照片?比如,两位小朋友,一位 10 张里面 8 张选择了花,另一位 10 张里面 7 张选择了花,这并不能代表前者比后者更喜欢花。所以,该指标也不太适合作为本研究的效度指标。

回应: 本研究的外显态度测试采用 Baron 和 Banaji(2006)发表于 *Psychological Science* 上的方法。采用破选法让被试在一系列配对出现的花虫图片中选择自己更喜欢的一张。选择花的百分比作为其外显态度的指标。

第三轮

编委复审意见:

作者根据一审两位审稿人的意见进行了仔细的修改,并对有关问题进行了答复。作者的态度是严谨的。虽然本研究依然存在一些缺陷(见二轮审稿人的意见),但是,考虑到有些工作已经无法弥补,现有的数据还是为测量工具的有效性提供了初步的证据,具有一定的创新性。同意发表。