

# 《心理学报》审稿意见与作者回应

题目：学前儿童基于物体重量归纳推理的发展：中美跨文化比较

作者：王志丹，周爱保，张荣华，卜梦瑾，李玉雯，王海静，Williamson Rebecca

---

## 第一轮

### 审稿人 1 意见：

语言标签在中美儿童对物体不可视属性类别归纳推理中的作用本文通过重量这个物体的内在不可视属性，研究了中美儿童使用语言标签进行类别归纳推理的发展特点。结果表明 4 岁和 5 岁儿童选择正确重量的物体的频次均显著高于 3 岁儿童，但 4 岁和 5 岁儿童之间没有显著差别。另外，4 岁和 5 岁儿童的表现显著高于随机水平，而 3 岁儿童的表现则与随机水平无差别。最后，中美儿童在根据重量进行类别归纳推理的发展趋势及年龄转折（3 到 4 岁作为关键转折点）上具有跨文化的一致性。总体而言，本文对研究背景的介绍，实验设计，结果都描述地比较清晰。

**意见 1：**引言部分可以重新安排使逻辑更通顺。方法部分有些信息需要补充。引言部分指出基于重量的归纳推理和这方面研究的跨文化比较是目前很少被研究的领域，也是本文独特的地方。但我觉得作者还应该再在讨论部分探讨一下本研究是否有一些实际的比如对教育或养育的指导意义。

**回应：**感谢审稿专家的意见。根据您的意见 4，我们已经对引言部分的逻辑进行了重新调整（详见正文第 14 至 92 行）。在方法部分我们进一步完善了相关信息。另外，在文章的讨论部分，我们也补充本研究的教育意义（详见正文第 301 至 306 行）。

**意见 2：**看标题语言标签似乎是本研究的重点，但读完全文认为发展趋势和文化差异才是。所以作者可以考虑修改一下标题以突出重点避免误导。

**回应：**和审稿专家二的意见一致，我们对文章题目进行了修改，以便更好的反应发展和文化这两个研究重点。新修订的题目是《儿童重量类别归纳推理的发展：中美跨文化比较》（详见正文第 1 行）。

**意见 3：**摘要部分并未提及美国的 2 岁和 6 岁儿童的结果，如果是出于摘要长度要求的考虑，可以忽略这个问题，因为我看到英文摘要里是有详细信息的。

**回应：**感谢审稿专家的详细指导。我们在摘要中补充了这两个年龄组儿童的结果（详见正文第 7 至 8 行）。

**意见 4：**引言部分对语言标签的介绍只在第二段，之后用四段来介绍儿童对重量的认知发展，然后接下来的第七段才是介绍儿童如何根据重量来对物体进行分类（这一段的描述不是很容易让人读懂，作者可尽量修改以表述清楚儿童是怎样进行分类的）。我认为关于语言标签的介绍可以放在第七段之后，段落结尾处指出还没有研究用语言标签研究重量推理，这样和第八段对亟待解决的问题也会衔接地比较好。

**回应：**感谢您提出的宝贵意见。我们已经对引言的逻辑进行了修改，在段落安排上，根据您的意见，我们将关于语言标签的内容移到了对重量的描述之后，另外我们对重量任务方面的

描述也进行了重新组织。首先，我们在文章第一段介绍了物体不可视属性重量研究的重要性，接着第二段我们回顾了皮亚杰关于儿童重量的研究，在第三段我们介绍了最近的一些研究可能与皮亚杰的发现是不一致的。第四至五段我们将重量任务分为两类，重点介绍了学前儿童的重量认知。随后在第六段我们引入了对语言标签的介绍，从而和第六段段尾指出没有研究使用语言标签考察重量推理相承接。到目前为止，我们已经对引言的逻辑和结构进行了较大的调整和修改（详见正文第 14 至 92 行）。

**意见 5：**在实验一的介绍中，作者提到对两组实验材料和语言标签进行了顺序平衡，然后这四种可能性是怎么分配给每一组的 20 个孩子的呢？

**回应：**实验材料有两组，分别是冰淇淋杯子和白色小瓶子，一半的被试先接受冰淇淋杯子后接受白色小瓶子测试，另一半被试先接受白色小瓶子后接受冰淇淋杯子测试。其次，实验人员所使用的语言标签也实现顺序平衡（dax/wug），其中一半的被试测试时实验人员先使用 dax 后使用 wug，另一半的被试测试时实验人员先使用 wug 后使用 dax。我们已经在文中补充了这些信息（详见正文第 166 至 170 行）。

**意见 6：**根据作者的描述，中国研究是基于美国研究的基础上选择的年龄组儿童。希望作者可以补充一下两个研究的实验间隔是多久。

**回应：**美国儿童数据的收集具体时间为 2016 年 1 月至 4 月，中国儿童数据收集的具体时间为 2016 年 10 月，数据收集的时间间隔为半年，我们在论文方法部分补充了这些信息（详见正文第 217 行）。

**意见 7：**关于数据分析，因为因变量是 0-2，只有 3 个取值。作者可以看一下数据分布对正态性的偏离，如果是可接受的范围，那么就用 ANOVA，如果严重违反了正态性分布，那就用 K-W 非参检验。没有必要详细介绍两组结果。折中的办法就时简单提一句另一种方法得到的结论一样。

**回应：**感谢审稿专家的专业性建议。由于数据的分布非正态，且样本每组 20 人不是很大。因此，我们采纳您的意见，使用非参数检验报告我们的实验结果（详见正文第 181 至 188 行，第 226 至 232 行和第 240 至 248 行）。

**意见 8：**另外，研究一的结果中，没有提 3 岁组和 4 岁组是否有差别。

**回应：**感谢审稿专家的细致指导。我们补充了这个结果（详见正文第 186 行）。

**意见 9：**总讨论的第一段作者提到本文对开展后续的研究有一定的启示意义，可以列举几个方向。

**回应：**我们在论文结尾部分介绍完本研究的不足之后，列举了几个未来可能的研究方向（详见正文第 315 至 327 行）。

**意见 10：**对研究缺点也没有充分讨论，比如每组 20 个这样的样本数量也不是很大，中美两组儿童的可比性对研究结论的影响也没有提及。

**回应：**感谢审稿专家的中肯建议。本研究的样本数量确实不大，每个年龄组的被试数量偏少。另外，由于研究数据在两个国家进行，测试环境也未能保证完全一致，这些因素都可能在一定程度上影响本研究结论的效力。最后，两国儿童的背景因素如家庭社会经济地位等方面的信息我们也没有收集，这也在一定程度上也影响了中美儿童的可比性。我们在文章总讨论部分补充了这些研究上的不足（详见正文第 307 至 314 行）。

.....

**审稿人 2 意见：**

《语言标签在中美儿童对物体不可视属性类别归纳推理中的作用》分别在中国和美国幼儿群体中探索儿童使用语言标签对重量进行类别推理的作用，研究具有一定的意义和价值，但仍存在一些问题，具体如下：

**意见 1：**论文的题目与研究的内容存在一些脱节，题目看起来是研究语言标签对推理的作用，但是研究的具体内容并没有涉及对语言标签作用的探索，没有设置无语言标签的对照组。另外建议题目具体化，因为本研究之探讨了一种不可使属性即重量。

**回应：**感谢审稿专家的指导。根据您和审稿专家一的共同意见，我们已经对文章题目进行了调整。新修订的题目是《儿童重量类别归纳推理的发展：中美跨文化比较》(详见正文第 1 行)。

**意见 2：**文中多处地方需要引文出处，建议作者完善引文信息。

**回应：**到目前为止，我们尽可能详细的补充了引文信息。

**意见 3：**在问题提出部分，作者认为目前关于什么年龄开始能够根据重量进行推理没有明确的定论，指出有的研究指出 4 岁儿童能够解决天平任务等，而另外的研究表明 3 岁能够解决这些任务，然后作者提出假设 4 岁能够使用语言标签进行归纳推理。在这里作者缺乏对已有研究背后原因的解释不足，也进一步导致为什么有不一致的研究结果却推论本研究与其中某一些研究结果一致的原因不足，建议作者补充。

**回应：**感谢审稿专家的专业性建议。我们确实在重量任务部分对发展的年龄叙述有一定的不一致，一部分原因是因为即使是同类的任务，可能具体的测试情境不同，进而导致解决任务年龄也不尽不同。在修改稿中，我们重新组织了表达，将研究儿童重量认知的任务分成两类，一类是重量因果推理任务，一类是重量分类任务。在每一种任务中我们分别叙述了儿童解决的年龄发展特点，同时我们修改了文章的表述，使前后保持了一致（详见正文第 50 至 76 行）。

**意见 4：**对研究进行跨文化比较的论证不充分，既然作者也认为对于重量的理解较少受到文化的影响，那开展跨文化研究的必要性和意义在哪里？另外，假设不存在文化差异，在假设检验的逻辑上似乎有问题。

**回应：**非常感谢审稿专家提出的宝贵意见。我们认识到了这个问题，并在修改稿中对这个假设进行了重新组织。我们认为可能存在以下两种可能性。一方面，研究表明英语中包含更多关于物体的词汇，汉语中包含更多社会习惯方面的词汇 (Tardif, Gelman, & Xu, 1999)。从教养方式上看，美国母亲更多的为儿童命名周围的物体，相反，中国母亲更多的使用动词，而名词基本上是被忽略的 (Kim, McGregor, & Thompson, 2000; Tardif, 2006)。由于英语的语言特点及美国的教养方式，很可能美国儿童具有更多关于自然界事实方面的知识，而以汉语为母语的儿童则具有更多关于社会习惯与社会交往方面的知识。因此，第一种可能性是美国儿童比中国儿童具有更多关于物理世界方面的知识，进而使他们在物体属性推理任务上的表现高于中国儿童。另一方面，重量是一种物体属性，侧重的是儿童对自然界客观事物属性的理解，不论哪一种文化背景中的儿童都可以通过与周围物体的直接性的感知接触而获得这些知识，并且前人的研究测试美国和中国儿童在按照重量对物体分类任务中的表现，结果没有发现文化差异 (Wang et al., 2015)。因此，第二种可能性重量类别归纳推理可能较少受到文化和社会因素的影响，结果导致美国和中国儿童在本实验任务中表现不存在差异。我们在文中补充了这些信息（详见正文第 106 至 117 行）。

**意见 5:** 方法部分，两个实验的名称都没有体现语言的作用，与本文的标题之间缺少呼应。

**回应:** 文章研究的重点是从跨文化的角度来探讨儿童重量类别归纳推理的发展，语言标签在研究中仅作为类别线索，所以并非研究的重点。因此，根据您的意见 1，我们已对标题进行了修改。同时，为了进一步让读者明确每一个实验中的目的，我们在文章引言部分的结尾新增了一小段，再次重申了每一个实验考察的具体内容（详见正文第 118 至 120 行）。

**意见 6:** 研究 1 被试的描述信息不够充分，比如每一个年龄组被试的年龄范围和平均年龄

**回应:** 本研究美国部分的数据是在亚特兰大 Fernbank 儿童自然博物馆收集的，监管该研究的佐治亚州立大学学术伦理委员会认为在公共场合询问并记录儿童的准确年龄有透漏隐私之嫌，因此建议我们只要求父母或监护人报告儿童的年龄但不报告具体日期，所以，我们研究中报告的年龄无法玩去统计出具体的平均年龄，而特定的年龄氛围包括所有跨度的儿童，例如，3 岁包括 36 个月至 47 个月的儿童。我们以脚注的形式标明了这些信息（详见脚注 1）。

**意见 7:** 研究中使用两个计分员对儿童的成绩进行记录，最终儿童的成绩采用的哪一个记分员的结果？

**回应:** 感谢审稿专家的指导。实验一中美国儿童数据部分，一个实验人员记录全部儿童反应数据，第二个实验人员记录了随机 20% 儿童的实验数据，目的是为了计算评分的一致性，最后选择了第一个实验人员记录的实验数据。实验二中中国儿童数据部分，两名实验人员同时记录了全部实验数据，基本上全部一致，对于不一致的地方，我们进行了重新的核对，并进行了意见的统一，所以最终使用的是数据是完全一致的，所报告的信度是在未进行统一意见之前两者之间的一致性。

**意见 8:** 研究中既然使用的语言标签都是无意义的音节，为什么在中国样本中对音节进行了改变？如何保证测试材料在两个样本中的等值性？

**回应:** 英语中的无意义音节如果直接翻译成中文使用就不是无意义音节了，例如，研究中使用的一个英语无意义名词“dax”，如果按照音译为“戴克斯”，似乎成了一个人的名字。所以，我们根据专家的建议，按照汉语的发音特点单独够早了两个无意义音节。我们在文中补充了这些信息（详见正文第 219 至 223 行）。

**意见 9:** 其他问题，如标点符号的规范使用，语言表述的清晰及规范性等。

**回应:** 感谢审稿专家的详细指导。我们已经进一步核对了标点符号，语言表述等方面的问题。但由于儿童研究的难度确实较大，而我们自身水平也相对有限，可能还存在很多不足之处，我们会进一步努力。

---

## 第二轮

**审稿人 1 意见:**

感谢作者非常积极地采纳了我上轮提出的意见。其他一些小的建议:

**意见 1:** 关于分析方法，很多读者也许并不熟悉 K-W。建议作者加上一句话说明 ANOVA 的分析你们也做了，得到的结论一样/类似。

**回应:** 感谢审稿专家的专业性建议。根据您的建议，我们在文中增加了该表述（详见正文第 183 至 184 行，第 233 行）。

**意见 2:** 所有引用文献处和句末 (比如,。。而名词基本上是被忽略的 (Kim, McGregor, & Thompson, 2000; Tardif, 2006) 还有其他括号处 (比如结果部分的括号, 4 岁组 ( $p = .004$ )) 都缺失空格。括号后面的空格也要注意, 例如, (2) 3 岁至 4 岁之间是儿童重量类别归纳推理发展的年龄转折点。

**回应:** 感谢审稿专家细致性的指导。我们已经仔细检查了每处参考文献和文中的括号处, 并添加了空格。

**意见 3:** 开始介绍实验一前的引言部分的文献(102-109 行处), 字体号码和其他部分不一致。

**回应:** 我们已在文中进行了修改。

**意见 4:** 表 2 内容字体太小。

**回应:** 感谢审稿专家的意见。根据心理学报的要求, 表题使用小五号黑体, 表文、图题、图文需使用六号宋体。所以, 图中内容看起来有点小。为了便于您的阅读, 我们将表中的内容更改为小五号宋体 (详见正文表 1 和表 2)。

**意见 5:** 以下  $p$  值看起来很怪, 是不是忘了小数点? “进一步的分析发现文化, 年龄和目标物体匹配得分三者之间的交互作用不显著,  $\chi^2(4) = 6.98, p = .14$ , 年龄与文化之间的交互作用不显著,  $\chi^2(2) = .87, p = .65$ , 文化与目标物体匹配得分之间的交互作用也不显著,  $\chi^2(2) = 3.69, p = .16$ 。”

**回应:** 我们已在文中进行了修改 (详见正文第 242 至 244 行)。

.....

**审稿人 2 意见:**

经过修改, 论文得到了较大的改善, 但仍存在几个小问题:

**意见 1:** 作者关于实验材料顺序平衡的描述还是不很清楚, 就目前的表述来看, 我理解为冰激凌杯子对应 dax, 白色杯子对应 wug, 在被试之间进行平衡, 是这样子吗?

**回应:** 感谢审稿专家的详细指导。我们的做法和您的理解是一致的, 原文中的表达不够清晰, 在修改稿中我们修改了相关表述 (详见正文第 164 至 165 行)。

**意见 2:** 表 1 的名称“各年龄组美国儿童选择正确物体次数”, 实际上表格的内容反映的是选择不同次数的人数, 建议作者修改表格题目, 为方便读者读取结果, 建议增加选择人数在不同年龄组的占比。表 2 同样存在类似问题。

**回应:** 非常感谢审稿专家提出的宝贵意见。我们已经在文中补充了这些信息 (详见正文表 1 和表 2)。

**意见 3:** 实验二样本的抽样方式是怎样的?

**回应:** 感谢审稿专家的意见。实验二采用随机方式抽取儿童。实验人员测试前拿到了班级名单, 对每名儿童进行了数字编号, 然后使用 [www.randomizer.org](http://www.randomizer.org) 网站在每个班级中随机生成了 20 个数字, 每个数字对应的儿童即为测试对象。我们在文中增加了随机抽样的表述 (详见正文第 210 行)。

---

### 第三轮

#### 编委意见:

作者按照审稿人建议进行了认真修改, 编委仍有以下问题和建议供参考:

**意见 1:** “美国 3 岁儿童的表现处于随机水平; 美国 2 岁儿童的表现边缘低于随机水平”, 建议明确此统计结果在儿童认知发展中的含义, 即此结果说明什么。另外, 3 岁组( $p=.16$ ) 选择正确物体的表现与随机水平没有差异, 是说明他们不理解任务, 还是缺乏重量类别归纳推理的能力?

**回应:** 非常感谢您百忙之中抽出时间来处理我们的文章, 并提出了宝贵的修改建议! 本文认为 3 岁儿童不能够进行重量类别归纳推理并不是由于他们不能够理解任务, 因为已有研究表明 3 岁儿童能够理解匹配样本任务 (Gelman & Wellman, 1991)。在该研究中, 实验人员先向儿童展示一张目标图片, 例如一头猪, 然后要求儿童从剩余两张图片中 (例如, 一个猪形状的存钱罐和一头牛) 选择他们认为最像目标图片的一张, 儿童随后还被询问 “你认为哪一张图片和猪有共同的内部构造。” 结果表明 3 岁和 4 岁儿童作出正确回答的次数均高于机率水平。其它一些研究甚至证明即使 14 个月左右的婴儿也能够理解匹配样本任务 (Graham, Kilbreath, & Welder, 2004; Welder & Graham, 2001; Hochmann et al., 2016)。因此, 本文认为 3 岁儿童尚缺乏进行重量类别归纳推理的能力 (详见正文第 276 至 285 行)。

**意见 2:** 对于本研究进行中美跨文化研究的意义和必要性论述不充分。“由于英语的语言特点及美国的教养方式, 很可能美国儿童具有更多关于自然界物体属性的知识, 而以汉语为母语的儿童则具有更多关于社会习惯与社会交往方面的知识。” 依据是什么? 即便 “美国母亲更多的为儿童命名周围的物体”, 那么儿童更多习得的是具体名词, 而物体的属性常常是抽象的。建议进一步加强研究假设的推导逻辑。

**回应:** 再次感谢编委老师的意见。我们查阅了资料, 补充了进行中美跨文化比较的意义 (详见正文第 86 至 95 行)。另外, 我们也修改了假设部分的推理论据。研究表明中美儿童在认知风格上存在差异: 中国儿童更擅长关系-背景性质的推理 (relational-contextual reasoning), 在这类推理任务中, 儿童需要基于物体之间的相互关系进行判断, 例如, 婴儿和妈妈属于一类是由于妈妈需要照顾婴儿。而美国儿童更擅长类别性质的推理 (inferential-categorical reasoning)。在这类推理任务中, 儿童需要基于物体的属性进行类别判断, 例如, 一艘船和一辆吉普车属于一类是由于它们都有发动机 (Chiu, 1972; Richland, Chan, Morrison, & Au, 2010)。由于本实验中的推理任务是基于物体属性的类别任务, 因此, 美国儿童在重量类别推理任务上的表现可能会高于中国儿童 (详见正文第 101 至 107 行)。

---

### 第四轮

#### 主编意见:

**意见 1:** 论文讨论的最后两段内容, 需要作者压缩。

**回应:** 感谢主编提出的宝贵意见和建议。我们对讨论部分最后两段的内容进行了整合 (详见正文第 285 至 303 行)。

**意见 2:** 科学研究需要实事求是, 如果一项研究中存在三个以上的不足, 会影响论文本本身的质量。

**回应:** 在上一稿中我们提到了本研究的三个不足。第一, 本研究分别在两个国家进行, 测试环境未能保持完全一致。这主要是由于两个测试点数据收集的便利性不同。在美国作者读博

士期间的实验室主要在当地的博物馆测试儿童,没有条件进入当地的幼儿园收集数据。但是,如果在中国的博物馆等公共场所测试儿童会有很多问题,一方面家长会比较排斥,另一方面和博物馆方面达成共识也比较困难。相反,进入幼儿园测试则相对比较便利。为了弥补测试环境方面的这些差异,我们已经保证了测试程序、测试材料、以及测试人员都是完全一致的,这在一定程度上确保了进行跨文化比较的有效性(详见正文第 285 至 287 行)。第二,研究样本数量不大,每个年龄组的被试数量偏少。实际上每个年龄组 20 名儿童已经达到了足够的统计效力。实验一中,我们根据预实验结果使用 G\*Power (Erdfeiler, Faul, & Buchner, 1996) 计算样本数量,检验功效设为.85,误差概率设为.05,效应量设为.20,结果显示样本数量为 100 人,每个年龄组 20 人。实验二的样本数量是根据实验一每个年龄组测试的儿童数量确定的。我们在正文中补充了这些信息(详见正文第 111 至 113 行; 194 行)。第三,虽然本文没有收集两国儿童的背景因素如家庭社会经济地位等方面的信息,但根据美国博物馆方面提供的信息表明该馆的主要访客为当地中产阶级家庭,而中国测试的幼儿园的注册家庭也多数为中产阶级家庭。因此,研究中的测试对象具有一定的可比性。我们在正文中补充了这些信息(详见正文第 287 至 290 行)。

研究人员开展跨文化研究,在不同国家收集数据确实面临很多现实性的困难,在一些细节上未能做到尽善尽美。但是,从总体上本文还是具有一定的研究价值的。第一,本文以美国作为西方文化的代表和中国作为东方文化的代表考察东西方文化背景儿童基于重量的归纳推理的跨文化一致性,这对于更好的理解不同文化背景儿童的认知发展具有重要的理论意义。第二,本文选择了物体的一个内在属性重量作为考察儿童的归纳推理能力,相对于前人研究大多考察儿童基于物体视觉外在属性的归纳推理具有一定的创新性。第三,本文发现 4 岁是儿童基于重量的归纳推理发展的年龄转折点。这有利于抓住关键期更好的对儿童进行归纳推理训练,提高儿童的科学素养。我们在正文中补充了本文的创新之处(详见正文第 290 至 296 行)。

**意见 3:** 反对将一个研究结果无限扩大,根据一项研究就指出未来要研究的许多问题,说服力不强。

**回应:** 我们对主编专家严谨的治学精神表示敬佩。原文中对未来研究方向的阐述确实有大而空的问题。经过审慎思考,我们凝练出了两点未来研究方向,这两点都是基于本文的研究发现和对该领域前人研究结论的分析而提出的。具体描述为:第一,本文只选择了重量作为考察儿童基于物体内在属性进行归纳推理的一个特例,未来的研究可以选择物体的其它内在属性(例如,声音)考察基于不同内在属性的归纳推理是否具有发展的一致性。第二,4 岁既是社会认知领域儿童具备心理理论(e.g., Wellman, 2001; Liu et al., 2008)又是物理认知领域儿童基于物体内在属性推理(e.g., Schrauf et al., 2011; Wang et al., 2018)发展的转折点,未来的研究可以探讨儿童进行社会认知推理和物理认知推理任务的内部机制是否一致(详见正文第 297 至 303 行)。

最后,再次感谢主编专家对本文提出的一系列宝贵意见和建议,这对本文的完善以及我们今后写文章都有重要的帮助,谢谢!