

《心理学报》审稿意见与作者回应

题目：从效应量应有的性质看中介效应量的合理性

作者：温忠麟，范息涛，叶宝娟，陈宇帅

第一轮

审稿人 1 意见：

该文讨论了效应量的作用和性质，然后具体讨论了中介效应量，简介了中介效应量 κ 平方是如何被终结的，具有一定的理论和实践意义。但该文还存在如下不足，与作者商榷：

意见 1：该文的创新性略显不足。从该文的题目“中介效应量 κ 平方的终结”来看，该文的最大创新之处，在于论证中介效应量 κ 平方是如何终结的。然而这个创新已经充分体现在《Monotonicity of effect sizes: Questioning kappa-squared as mediation effect size measure》上了。因此，本文的创新之处反而在于该文的前三个部分，关于效应量的讨论部分。

回应：赞同审稿专家的评论，结合审稿专家 2 的意见 1，将题目改为“从中介效应量 κ 平方的终结看效应量的性质”。

意见 2：该文的中介模型仅涉及了简单中介模型，即一个自变量，一个中介变量，一个因变量的情况，但是中介模型还包括多重中介、多层中介、纵向中介等模型，建议作者在文中进行一个简单的说明，以避免读者误解。实际上在《Monotonicity of effect sizes: Questioning kappa-squared as mediation effect size measure》中，作者还谈到了，中介效应量 κ 平方在多重中介模型中也是不合适的。

回应：修改稿简要补充了 κ 平方在多重中介模型中也没有单调性。因为本文重心放在效应量的作用和性质，所以没有涉及更多简单中介模型以外的中介模型。

审稿人 2 意见：

意见 1：本文与 Wen & Fan (2015) 的关系可考虑在文章标题上体现，比如加副标题，或者题目改为从...回顾效应量的评价标准，避免两篇文章在标题上距离太近。正文中亦可考虑精简对 Wen & Fan (2015) 研究过程的转述。

回应：非常感谢审稿专家的中肯意见，已将题目改为“从中介效应量 k 平方的终结看效应量的性质”，并精简了对 Wen & Fan (2015)研究过程的转述。增加了一项最新的中介效应量研究的评介。

意见 2：「效应量/Effect Size」在文献中的定义并无公认的统一说法。比如所引文献中地位较特殊的 Wilkinson & TFSI (1999)，可视为协调当时各方说法的公约数。在这篇纲领性的文献中（p.599，上下文细读可知），文中的 Effect Size 既包括本文所特指的标准化效应统计量，也包括未标准化的效应，还包括统计量所对应的总体参数（因此可以作置信区间，可以作特定临界值的 NHST）。考虑这一点，本文可在较前的部分用让步措辞比如「本文所指的效应量定义为...」。

回应：赞同此评论，兼顾到下面的意见 3，文稿做了如下修改，让读者明白什么时候未标准化的效应就可以作为效应量：

“如果涉及的变量有公认的量尺和测量单位，如长度测量使用米尺、抽烟数量以每天抽烟支数计算，则效应本身就是现成的效应量。例如，“男女身高均值差异为 0.1 米”、“男员工比女员工平均每天多抽 2.5 支烟”，意义就相当明确，同类的不同研究之间也有可比性。而且，这个基于原始测量单位的效应量与其他变换后的各种效应量相比，不仅简单，而且更加容易理解。”

“在物理测量中，通常都有公认的测量单位，就算测量单位不同，也可以彼此等值转换（如1米=100厘米）。然而，心理测量通常不像物理测量那样有通行世界的公制单位，不同的测量之间也难以等值转换。……这时，需要一种与测量单位无关（scale-free）的指标——效应量，来衡量效应的大小。”

意见 3：Wilkinson & TFSI (1999)这段引文——Always present effect sizes for primary outcomes. If the units of measurement are meaningful on a practical level (e.g., number of cigarettes smoked per day), then we usually prefer an unstandardized measure (regression coefficient or mean difference) to a standardized measure (r or d)，实际上说明了对于「报告标准化效应量」不宜笼统地夸大其必要性。本文或可考虑平衡介绍可能的重要性（或者说标准化效应量也可能在另一方向产生的误导性）。

回应：赞同此评论。参见上面回应意见 2 所做的修改，相信已经能够平衡未标准化效应量与标准化效应量的作用和重要性，并且也知道什么情况下使用什么效应量比较好。

意见 4: 本文的表 1 将两个逻辑合在了一起反而不容易说清楚。一件事情是点估计附带标准误（与置信区间大体反应出同样多的信息） vs 定性的 NHST 报告，前者会比后者的统计意义多表达出一层「实际/实质/Substantial」意义；另一件事情是效应量（用标准差作衡量单位来）标准化 vs 报告带有测量单位的未标准化量。此外特别地，表 1 的检验力其实无法考察，通常能考察的是标准误大小、置信区间宽窄。Hoenig & Heisey (2001)的文献 *The Abuse of Power* 指出，在数据收集之后 Power 的计算报告通常被误用误读有弊无利。这篇文献之后，报告 Observed Power 成为公认的统计误用。此后国际学术期刊发表的研究完全不再有报告 Power 的实证案例。

回应: 同意审稿专家有关检验力的评论，修改补充了表1的说明部分：“只有当检验结果是不显著时，才需要考虑检验力。不过，如果效应量小，通常都没有必要看检验力高低，可以直接作出没有效应的结论；如果效应量中上，检验不显著会令人怀疑是检验力不够高所致，文章投稿难以被接受，此时适当增加样本容量，通常都会得到显著结果。所以，虽然研究者需要有检验力的概念，知道增加样本容量是提高检验力的途径，但是可以说在文章中是不必报告检验力的。”

在得到效应检验结果、效应估计值及其置信区间、效应量估计值（效应量也可以计算置信区间）后，根据表 1 对统计结果做出判断还是很实用的。尽管不需要报告检验力，但无论是理论上的完整还是研究过程中的思考，检验力还应当包括在里面。

审稿人 3 意见:

《中介效应量 κ 平方的终结》一文主要介绍了中介效应模型中中介效应量计算中常用的几种估计量，重点介绍了 Preacher 和 Kelley(2011)提出的 κ 平方存在的问题。研究问题是目前中介效应研究探讨的主要话题，但是文章存在以下问题需要商榷:

意见 1: 文章中的观点与 Wen 和 Fan (2015) 文章的观点基本类似，内容上还不如 Wen 和 Fan (2015) 深入，不知是否有翻译成中文发表的价值；

回应: 正如审稿专家 1 所说，本文的主要创新之处在于前三个部分，特别是讨论效应量应当具有的性质。审稿专家 2 明察本文与 Wen 和 Fan (2015) 文章的区别，建议修改标题（已修改）。

意见 2: 文章的整体写法像教科书的介绍, 缺少必要的推导和深入的分析;

回应: 本文重点讨论效应量应当有的性质, (1) 分析了为什么这些性质是不可少的; (2) 指出如何知道一个效应量有这些性质; (3) 对例子逐一分析具有哪些性质, 必要的地方对公式做了变形展示。

国际上重要的方法期刊 (如 *Psychological Methods*—国际上对社科研究方法影响很大的 APA 期刊, $IF=7.3$), 都会希望作者用通俗的文字教会读者方法的东西, 所以才会有那么多人引用。

意见 3: 文章很多内容基本都是已有文献中的内容, 似乎没有必要介绍如此详细 (如 P1-P9)。整体来看, 认为有了 Wen 和 Fan (2015) 这篇文章, 本文就没有发表的价值。

回应: 无论国内国外, 都未见到文献专门讨论效应量应当具有什么性质。国内对效应量的研究更加罕见。就算是研究方法、测量统计方面的专家, 对有关报告效应量和报告检验力的议题都还存在疑惑。

本文明确指出, 报告效应量总是需要的, 而检验力却不需要在投稿文章中出现。根据本文可以得到如下判断: (1) 如果效应显著, 此时检验力不用考虑 (考虑检验力相当于考虑第二类错误率, 显著的时候只有第一类错误)。(2) 如果效应不显著: (2a) 如果效应量小, 不用考虑检验力, 效应量小的时候就算显著都没有什么意义; (2b) 如果效应量中上, 文章不应当投稿, 而是应当增加被试, 提高检验力, 因为效应量中上, 只要样本容量够大, 通常都会得到显著结果。

结合审稿专家 3 和专家 1 的意见, 修改稿精简了对 Wen & Fan (2015) 研究过程的转述, 增加一项最新的中介效应量研究的评价。

第二轮

审稿人意见

意见 1: 该文的文题《从中介效应量 k 平方的终结看效应量的性质》不能充分反映文章的内容。首先, 该文的逻辑顺序是首先介绍效应量的作用 (一是弥补了统计检验的不足, 二是使得效应有可比性) 和基本性质 (测量单位无关、单调性、不受样本容量的影响)。然后利用效应量的性质来审视中介效应量。而根据现在的文章题目, 给人的感觉是首先应该介绍中介效应量 k 平方的终结原因, 然后来看效应量的性质。其次, 该文章将大多数中介效应量都进行

了评价，内容不仅仅是 κ 平方这个中介效应量。因此，建议作者修改文题，是否已可以考虑改为《从效应量的性质看中介效应量的合理性》。

回应：赞同审稿专家的建议，题目改为“从效应量应有的性质看中介效应量的合理性”。同时，因应题目的变化，调整和修改了中介效应量部分的小节及其标题。

意见 2：建议作者在摘要和正文的第一段中，不要一开始就强调“Wen 和 Fan（2015）的研究结果彻底终结了 κ 平方作为中介效应量的合法性”。因为，Wen 和 Fan（2015）文章的核心在于终结 κ 平方作为中介效应量的合法性，而在本文中，终结 κ 平方作为中介效应量的合法性仅是本文内容的一小部分。本文的重点在于效应量的作用和性质，然后用效应量的性质来审视各种中介效应量是否合理。

回应：摘要按照下面第 3 条建议进行了修改。修改后摘要中不再出现“Wen 和 Fan（2015）的研究结果彻底终结了 κ 平方作为中介效应量的合法性”。正文第一段则基本保留，其中 Wen 和 Fan（2015）的研究结果是当作故事来讲的，放在开头方便引出本文的主题；在讨论效应量性质之后，简要介绍他们的研究结果并回应开头提出的问题。这种“插叙”的写法在科研论文中也是常见的，有助引起读者的兴趣。

意见 3：基于第二条的理由，中英文摘要也建议作者修改。摘要是否可以这样开始“效应量的作用有两个方面，一是弥补了统计检验的不足，二是使得效应有可比性。结合统计显著性和效应量，才能得出适当的统计结论。效应量应当具有一些基本性质，包括与测量单位无关、单调性、不受样本容量的影响。中介效应量 κ 平方就是因为缺乏单调性，而被彻底终结了其作为中介效应量的合法性。……”

回应：按照建议修改了摘要。

意见 4：关键词选择“中介效应”是合适的，但为什么选择“间接效应”作为关键词？本文主要用的是简单中介模型。在简单中介模型中，中介效应就是间接效应。是否全文可以统一说法为中介效应呢？这样读者理解起来也更容易。

回应：是的，中介效应和间接效应是近义词，有些地方换用没有问题。但有的场合它们有微小区别，例如，相对于直接效应的时候，使用“间接效应”比较好。几乎所有研究中介效应的方法文章，都会同时使用两种术语。将其列为关键词不仅因为文中多次出现“间接效应”，而且希望用“间接效应”作为关键词能检索到本文。