

《心理学报》审稿意见与作者回应

题目：改进的认知诊断模型项目功能差异检验方法 ——基于观察信息矩阵的 Wald 统计量
作者：刘彦楼 辛涛 李令青 田伟 刘笑笑

第一轮

审稿人 1 意见：

Hou 等人提出的 Wald 检验法是 CDM 中的 DIF 检验方法之一。本文欲解决这个方法中一类错误膨胀的问题，具体重大改进是提出了一种计算 Wald 检验中标准误的新方法。如果作者的方法成立的话，这个研究就具有一定的理论与实践意义，但是由于作者写作与实验设计的问题，导致我目前无法对其做出准确的判断。建议作者做出如下的修改，以便我重新审稿：

意见 1：从设计角度来说，最大的问题是作者没有把新方法 with Li & Wang, 王卓然、郭磊、边玉芳两篇提出的方法进行比较。如果新方法没有与 Li & Wang 比较，就无法体现其优越性，无法展现本研究对 CDM DIF 研究的贡献。特别是 Li&Wang 的方法也是对 Hou 等人方法的改进，避免了一类错误过大、检验力下降、需要进行多次参数估计、局限于两组的 DIF 检验等弊端。作者没有必要和两篇中提出的老方法都进行比较，但是选择其中一篇的方法也必不可少的。本人推荐选择 Li &Wang 的方法，基于以下几个原因：1) 他们的方法是最新的，代表了目前 CDM DIF 研究中的最高水平；2) 他们的方法是基于 LCDM 的，本文也恰巧采用了这个框架。

回应：感谢审稿专家的宝贵建议，根据您的意见，我们已经在论文中增加了相关表述。本文中所采用的实验设计与 Hou 等人(2014)以及 Li 和 Wang (2015)的相同，因此，模拟结果是可以直接与 Li 和 Wang 的方法进行比较的，本文已经在正文 4.1 中增加了与 Li 和 Wang 的 LCDM-DIF 以及 Wald 结果进行对比的表述，且在文中用红色字体标注。感谢审稿专家这一建设性的意见。

意见 2：就新方法与老的 Wald 检验比较而言，作者的写作非常不清晰，有很多值得改进的地方：

1) 采用 LCDM 值得探讨。本文没有论述在 Wald 检验中采用 LCDM 的必要性与优势。从文稿看来，似乎采用 LCDM 在作者的新方法中没有任何作用与意义。是否有必要引入这

个新的模型，能否直接使用 DINA 来表现新方法的优越性。如果 LCDM 没有限定成 DINA 模型，那么这个本身可能就是一个混淆变量，这就大大削弱了模拟研究的结论。另外，本人如果借用了 Li & Wang 文中的 LCDM—DIF 的理念，应当引用文献。

回应：DINA 模型是一个带有严格约束的模型，在有些情况下，其假设与现实并不相符，因此需要考虑一个更加一般性的模型，这也是 Li & Wang 研究中采用 LCDM 的原因。另外，在 LCDM 框架下，DINA 模型的猜测参数与滑动参数可以与 LCDM 模型中参数进行等价转换，请见正文中的公式（2）与（3）。为了使本研究中提出的方法更具有一般性，方便后续的理论研究及现实应用，我们采用了 LCDM 框架，在论文的综述部分我们增加了相关的论述，且用红色字体标注。

本研究所提出了改进后的 Wald 统计量不仅适用于 DINA 而且适用于一般性的认知诊断模型框架，为方便与以往研究结果进行直接比较我们将 LCDM 限定成 DINA 模型。本研究中所使用是 Wald 统计量及其计算方法与 LCDM—DIF 检验方法的理念不一致，因此在这一部分我们没有引用 Li & Wang 的研究。

2) 本文提出的另一个改进是在计算标准误时不仅要包含项目参数，而且要包含模型参数。我不能理解作者定义的项目参数与模型参数具体指的是什么。希望作者能够以 LCDM（在本文中应该是与 DINA 模型等价的 LCDM 模型）为例，具体指出哪些是项目参数，哪些是模型参数。与此相关的是，作者在第四页指出了两种不同模型的参数个数，但是文章都是用了 F_p ，这就导致我不理解在新 Wald 检验中到底是多少个参数；而 de la Torre 等人提出的方法参数个数又是多少个呢？

回应：在计算模型的协方差矩阵时，要包含项目参数与结构参数，认知诊断模型中的模型参数既包括项目参数也包括结构参数。为使本文表达更加清晰，我们在文中增加了如下表达：

“为行文方便，我们以加粗的英文或希腊字母表示向量以及矩阵，设 $\beta = (\beta'_1, \dots, \beta'_j, \dots, \beta'_j)'$ 为模型中所包含的所有项目参数，其中 $\beta_j = (\lambda_{j,0}, \lambda'_j)'$ ， $\lambda_{j,0}$ 为项目 j 的截距参数， λ_j 为项目 j 的主效应以及交互效应参数向量。”，“在本研究中，我们将 LCDM 中的 λ_j 参数的主效应以及低阶交互效应全部设置为 0，只保留最高阶交互效应项，即 $\beta_j = (\lambda_{j,0}, \lambda_{j,K_j(1,\dots,K_j)})'$ ，以获得 DINA 模型”，“通过公式错误！未找到引用源。与错误！未找到引用源。可以发现，认知诊断模型中除了有项目参数 β 外，还有结构参数 η ，因此，可以将认知诊断模型的模型参数表述为 $\gamma = (\beta', \eta')$ 。”希望以上表述可以回答审稿专家关于“希望作者能够以 LCDM（在本

文中应该是与 DINA 模型等价的 LCDM 模型) 为例, 具体指出哪些是项目参数, 哪些是模型参数。”的问题。

DINA 模型是 LCDM 的特例, LCDM 经过约束可以获得 DINA 模型, 两者并非是完全不同的模型, 在有特殊的约束情景下它们是一致的, 为了便于理解, 我们在修改的论文中, 将模型的参数个数统一用变量 F 来表示。当为 DINA 模型时, Wald 统计量中检验的是项目参数, 其数量为 2。Wald 统计量的计算中需要用到相应项目参数的协方差矩阵, 协方差矩阵的正确计算方式时求模型参数信息矩阵的逆, de la Torre 等人提出的方法参数个数也是 2, 只是他们研究中信息矩阵的计算方法存在问题。

3) 公式 1, 2, 3 中用 λ , 而公式 6 中出现了模型参数向量 η 。不知道他们是否相同, 如果相同请统一。如果不同, 请说明两者之间的关系。

回应: 感谢专家的专业与细致, 对其表述我们已经进行了统一。具体可见正文部分或者是在问题 2) 中对于专家的回复。

4) 通过通读全文, 本人推测项目参数指的是传统 CDM 的猜测 (g) 与失误参数 (s), 而模型参数指的是 LCDM 中的那些 λ (或者 η)。但是根据公式 2, 3, g 与 s 是 λ 的函数, 那么他们的估算标准是不是可以根据 λ 的标准误直接计算得来。如果我的推测是正确的话, 我不知道如何建立一个既包含项目参数也包含模型参数的模型, 至少 LCDM 与 DINA 模型都不是这样的模型。

回应: 与单维项目反应理论不同, 在认知诊断模型中包含两类参数, 项目参数 (本研究中用 β 表示) 以及结构参数 (本研究中用 η) 表示, 模型参数 $\gamma = (\beta', \eta')$ 。这一表述不仅适用于 DINA 而且也适用于 LCDM。

5) 第 4 页上作者指出, 公式 7 的参数估算“详细的推导过程, 感兴趣的读者可以参考 de la Torre (2009, 2011)”。我不能理解的是他的两篇文章都是与 DINA, G-DINA 有关, 并没有涉及 LCDM 的 EM 参数估计, 那么我怎么样才能通过这两篇文章得到 LCDM 的估计方法呢? 既然采用了 LCDM, 是不是更应该引用 Templin 等人的文章呢? 另外, G-DINA 的构造方式与 LCDM 非常相近 (但是两者并不完全等价), 采用了 loglinear 与 log odds 的表达方式, 也都有所谓的模型参数, 所以我不能理解 “它不仅包含项目参数, 而且还要包含自由的模型参数, 而非 de la Torre (2009, 2011) 所认为的仅项目参数需要被考虑” 一句中认为 de la

Torre 不涉及模型参数的论断。准确地说，DINA 中确实不涉及模型参数，但是 G-DINA 涉及了模型参数。本文涉及 DINA，所以作者才这么说的吧。

回应：G-DINA 中的 loglinear 构造方式等价于 LCDM，因此，可以得到 LCDM 的 EM 参数估计，在 Templin 等人（根据专家的意思我猜想应该指的是 Henson, Templin 和 Willse (2009) 的论文）的文章中，采用的是 MCMC 的参数估计方法，因此我们没有引用。de la Torre (2009, 2011) 关于采用 EM 算法去模型参数的方法是正确的，但是 de la Torre 在这两篇论文中关于信息矩阵的计算方法是错误的，他认为在信息矩阵中“仅项目参数需要被考虑”，实际上在信息矩阵的估计中不仅需要考虑到项目参数而且需要考虑到结构参数，DINA 以及 G-DINA 模型中均涉及了模型参数。

6) 交叉相乘信息矩阵概念使用不当。XPD 应该指的是 empirical cross-product information matrix，即经验交叉相乘信息矩阵。这种方法用经验交叉相乘信息矩阵来估算标准误。

回应：多谢专家指正。XPD 的英文全称是 empirical cross-product information matrix，可以译为经验交叉相乘信息矩阵，亦可简称为 cross-product information matrix，交叉相乘信息矩阵。为了表达的严谨性，根据审稿专家的建议我们已经将“交叉相乘信息矩阵”修改为“经验交叉相乘信息矩阵”，感谢专家的严谨与细致。

7) 文章指出 LCDM 的交叉相乘信息矩阵是公式 9，那么请明确说明 Hou 等人采用的是不是也是同样的公式（只存在项目参数个数不同？）。

回应：Hou 等人采用的公式类似于公式 9，但是她们研究中信息矩阵的计算仅仅包含了项目参数而非模型参数，即她们忽略了模型的结构参数。本研究中的项目参数个数与 Hou 等人的项目参数个数是一致的。

8) 公式 9 中的 π_0 与公式 10 中的 $L(x)$ 是不是都是期望概率。如果是，请统一术语与符号。如果不是，请详细说明 π_0 具体是什么的期望。

回应：感谢专家的建议，我们已经统一了术语与符号，并且进一步对其进行了的解释与说明。具体修改如下，

$I_{Obs} = N \sum_{v=1}^{v_0} \frac{p_v}{\pi_v} \left[\frac{\partial \pi_v}{\partial \gamma} \frac{\partial \pi_v}{\partial \gamma'} - \pi_v \frac{\partial^2 \pi_v}{\partial \gamma \partial \gamma'} \right]$ 在公式错误！未找到引用源。中 v_0 是观测到的作答模式

的总数； p_0 与 π_0 分别代表第 0 个观察到的作答模式所占的比例以及期望的概率。再次感谢这一建设性的意见。

9) 第7页作者指出“本研究中所提出的改进的Wald统计量计算方法所获得的结果并不存在过度膨胀的现象，这与Hou等人(2014)以及王卓然等人(2015)的结果恰好相反，说明本研究中所提出的Wald统计量的计算方法明显优于以上两个研究所使用的方法。”王卓然等人的文章提出的是LR检验法，但是本人并没有把LR纳入模拟研究，所以这句话是不严谨的。

回应：感谢专家指出本文中表达不够严谨之处。我们已经将文中相关表述修改为“说明本研究中所提出的Wald统计量的计算方法明显优于以上两个研究所使用的Wald统计量的计算方法”。

10) 第7页作者指出“健壮宽松准则(当显著性水平为0.05时经验一类错误控制率在0.025与0.75之间)”。个人认为合理的区间应当是0.025到0.075之间，所以这里可能是一个笔误，请作者查证。

回应：感谢专家的细致与严谨。我们已经对论文进行了修改。

11) 第9页作者指出“如，Hou等人(2014)以及王卓然等人(2014)的模拟研究发现Wald统计量会存在一类错误控制率膨胀的问题，Li等人(2015)的模拟研究却发现Wald统计量在使用MCMC计算时有着良好的一类错误控制率。Li等人(2015)将这一原因归为MMLE/EM算法的问题，但本研究通过模拟发现，这主要是由于Wald统计量计算方法的差异引起的，而非MMLE/EM算法的问题。”我的理解是，除了LCDM-DIF方法之外，Li等人已经从MCMC的角度提出了一个改进的Wald法（也就是说，这也是一种计算Wald统计量的新方法），很好地控制了一类错误的问题，而本文是基于EM法的新解决方案。那么，作者也应当展现新的Wald改进方法与这种基于MCMC法的Wald改进方法的优势。另外，作者对Li等人提出的解释的解读过于负面。个人认为他们的从算法的角度是非常恰当的，从他们的基于MCMC的改进Wald法结果，可以看出MCMC可以很好的计算标准误，从而改进了Wald检验，避免了Hou等人的问题，这确实是算法造成。当然，这与本文的贡献并不矛盾。本文更加清晰指出了传统EM的弊端，给出了一种更理想的基于EM的解决方案，加深了对这个问题的理解。个人认为，没有必要为了突出本文的研究意义，而刻意对过去的研究做负面的评价。更重要的是，本文并没有与MCMC改进的Wald比较，我们对最后的

表现仍然需要拭目以待。

回应：感谢专家所提出的建议及对我们研究的肯定。Li 等人(2015)在 MCMC 算法下，所使用 LCDM-DIF 方法以及 Wald 统计量确实也有着较好的一类错误控制率(除了有稍许膨胀外)，这与 Hou 等人(2014)以及王卓然等人(2014)的模拟研究结果产生了重大的不一致，这是本文的研究动机之一。抱歉给专家造成了“没有必要为了突出本文的研究意义，而刻意对过去的研究做负面的评价”的误解。Li 等人所使用的 MCMC 算法，确实可以较好的计算标准误，从而改进了 Wald 检验，避免了 Hou 等人的问题。Hou 等人(2014)以及王卓然等人(2014)的研究结果的不足，是不恰当的信息矩阵计算方法造成的，而非 EM 算法本身的问题。在 EM 算法中，信息矩阵（或者是其逆协方差矩阵）并不是伴随产生的，需要去单独计算，因此使得 Li 等人(2015)有了“可能是 MMLE/EM 存在问题”的假象。另外需要指出的是 Li 等人在计算 MMLE/EM 算法下的 Wald 统计量时使用的是 R 软件中的 CDM 包(Robitzsch, Kiefer, George, & Uenlue, 2014)，这一软件包在计算 Wald 统计量时所使用的协方差矩阵计算方式也是存在问题的，专家如果感兴趣的话可以下载相关的源代码进行查看，网址为：<https://cran.r-project.org/web/packages/CDM/index.html> 或者是联系我们由我们指出其中的问题。因此，我们认为应该是从信息矩阵估计方法的角度去探讨问题。为了能更好的贴近 Li 等人(2015)的原意以及避免产生“解读过于负面”的误解，我们在讨论部分已经将可能引起误解的表述进行了修改与删除，以图表达客观与准确。根据专家的建议，在新修改的论文中，我们已经在 4.1 中加入了改进的 Wald 统计量计算方法与 MCMC 改进的 Wald 统计量以及 LCDM-DIF 结果的比较。

总结上述意见，本文最突出的问题有两个：第一，模拟研究设计存在较大的问题，没有加入重要的基线方法（LCDM—DIF，以及基于 MCMC 改进的 Wald 法），可以考虑加入王卓然等人提出的 LR 法。第二，行文不清晰。作为一篇量化方法论文，作者完全可以使用公式清晰地说明研究中最重要概念，例如 Hou 等人的经验交叉相乘信息矩阵，项目参数与模型参数。加上作者引入了 LCDM 框架，同时涉及 DINA，G-DINA，并且在两者之间随意切换论述角度。很多重要概念可能产生不同的对等形式，但是作者没有明确说明他们是否等价，也没有明确说明在哪个框架下展示公式与概念，但是所以我只好凭借自己的背景知识、上下文以及 Li&Wang、de la Torre 的文章来推测作者行文的意义。一篇合格论文应该清晰地展现核心信息，让一位具有基本背景知识的研究者也能够基本掌握文章的思路。本文在改进模拟研究设计之后，也需要思考如何准确清晰的呈现研究成果。

回应：感谢专家宝贵的建议，我们的研究设计与 Hou 等人(2014)以及 Li 等人(2015)的研究设计是一致的，抱歉因为行文不清晰的原因给专家所造成的“模拟研究设计存在较大的问题”的误解。也正是因为本研究设计与以往研究一致，研究结果之间是可以直接进行比较，因此，在正式的模拟中我们并没有继续重复前人的研究。而且以往研究中已经证实了 Wald 法要优于 MH, SIBTEST 或者是 LR 法，因此，我们并没有考虑将这些方法加入到我们的研究中。我们对于行文中的不清晰而给审稿专家造成的不便表示歉意，在新的修改稿中，我们已经尽力使得表达清晰，力图使具有基本背景知识的研究者也能够基本掌握文章的思路。

审稿人 2 意见：

DIF 检验是测验中一个比较重要的研究领域，本研究使用观察信息矩阵代替原始信息矩阵，尝试提出改进后的 wald 统计量，目的是能够较好地控制认知诊断 DIF 检验中的一类错误率。作者借鉴在 IRT 中，利用观察信息矩阵能够更精确地获得方差—协方差矩阵的优势，将该想法拓展至认知诊断中，具有一定的创新性，但该论文仍存在以下问题。

意见 1：根据 Wald 检验的本质，它是对不同组被试作答的题目参数是否存在差异的检验。作者在计算协方差矩阵中加入了被试信息是否多余？而且一个关键问题是，在更换计算 wald 值的信息矩阵后，新构建的统计量是否还是 wald 统计量？请作者仔细思考。新的统计量是否服从卡方分布？这一点作者在文章中需要给出明确的说明，必要时还需要数学推导。另外，作者在 2.2 部分说道：“如果我们所构建 Wald 统计量服从渐近卡方分布”，这是否表明，作者其实也不知道新统计量的分布形态呢？

回应：感谢专家的宝贵意见。（1）在认知诊断模型中不仅仅包括项目参数，而且还包含属性掌握模式概率的参数（本文中将其称为结构参数，也就是审稿专家所提及的被试信息），即认知诊断模型包括两类参数：项目参数以及结构参数。在通过对于模型的信息矩阵求逆计算协方差矩阵的时候要包含结构参数，仅仅对于项目参数的信息矩阵求逆所获得的协方差矩阵是不准确的。协方差矩阵是通过信息矩阵求逆获得的，信息矩阵的计算需要包含结构参数。

（2）当通过对模型完整参数的信息矩阵求逆，获得协方差矩阵后，可以通过相应的项目参数的协方差矩阵构建 Wald 统计量，只有这样才是正确的 Wald 统计量的构建方式。所以说，新构建的统计量是 Wald 统计量。（3）在本文中通过模拟研究的结果可以发现，在 DINA 模

型中，新构建的 Wald 统计量服从自由度为 2 的卡方分布。(4) 抱歉给审稿专家的理解造成了困扰，文中 2.2 部分的意思是如果所构建 Wald 统计量是渐近卡方分布的，那么观察到的一类错误应该与理论上的一类错误率相符。本研究的模拟结果证实了这一点。我们已经对相应的表述修改为“如果我们所构建 Wald 统计量是渐近卡方分布的，那么它观察到的一类错误率应该符合预先设置的理论上的一类错误控制率，如 0.05”。

意见 2: 新统计量检验的参数都包括哪些？请作者给出。

回应: 新统计量检验的参数是每个项目所包含的参数。例如，对于第 j 项目而言，具体为对照组的项目参数 $\hat{\beta}_{Rj}$ ，目标组的项目参数 $\hat{\beta}_{Fj}$ 。

意见 3: De la Torre 等人在计算协方差矩阵时，只使用了题目参数作为衡量 DIF 存在与否的依据。作者既然将自由的模型参数也包含在内考虑，那么请问这时的假设检验又是什么？即作者检验的参数是哪些？

回应: 在构建信息矩阵时，需要把模型中所有自由估计的参数考虑在内，此时通过对信息矩阵求逆获得协方差矩阵才是正确的。即计算信息矩阵时，要考虑所有自由估计的参数，获得正确的协方差矩阵后，只需要取其中对应的项目参数部分即可。(1) 假设检验是项目参数之间是否存在显著差异。(2) 检验的参数也只是项目参数，但是项目参数协方差矩阵的计算过程（求信息矩阵的逆）需要包含全部自由估计的模型参数，因为 Wald 统计量的计算中需要用到协方差矩阵。

意见 4: 结果呈现部分，是仅仅对项目参数，即 s 和 g 参数的检验，还是包含了模型参数 η ？

作者没有交代清楚

回应: 仅仅对项目参数，即 s 和 g 参数的检验，不包含结构参数。抱歉给专家的理解造成了困扰，为了表达清晰，我们将文稿中第 5 页的相关表述修改为“可以在项目水平上构建用于 DIF 检验的 Wald 统计量”。

意见 5: 接上一问题，如果作者不仅仅对项目参数 s 和 g 进行了检验，还包括模型参数 η ，那么请问检验时的自由度为何还是 2？如果只是对 s 和 g 进行检验，那么请问，在构建新统计量时将 η 包含在内，又不对它进行检验，是否合理？这种情况下， η 是否还有存在的必要

性？

回应：Wald 统计量只是针对项目参数 s 和 g 进行检验（当为 DINA 模型时），但是在构建它们的协方差矩阵时，需要用到结构参数 η ，因此，自由度是 2。认知诊断模型协方差矩阵的计算需要对整个模型自由估计参数的信息矩阵求逆，在构建信息矩阵的过程中需要用到结构参数 η ，否则计算获得的协方差矩阵是错误的。因此，需要将结构参数 η 包含在信息矩阵内，它要参与到协方差矩阵的计算中，它的存在是合理的，必要的。

意见 6：作者使用的是观察信息矩阵，所有自由估计的参数之间的协方差可能不为 0，这是否会对结果产生影响？具体又是什么影响，请作者思考。

回应：我们同意审稿专家关于所有自由估计的参数之间的协方差可能不为 0 的观点。在本研究中，我们已经假定所有自由估计的参数之间的协方差可能不为 0。只有这样做才是合理的，具体可参考：

Tsutakawa, R. K. (1984). Estimation of two-parameter logistic item response curves. *Journal of Educational and Behavioral Statistics*, 9, 263-276.

意见 7：作者虽然解决了 wald 统计量本身的问题（假设本文作者提出的新方法是对的前提条件下），但缺少和已有方法的比较，例如 LCDM-DIF 和 LR 法，如何体现出新方法的优势呢？假如新方法仍不如 LCDM-DIF 和 LR 法，那么这种改进意义就不大了。

回应：感谢审稿专家这一专业性的建议，在新修改的论文中我们已经采纳了这一建议，增加了相关内容的表述。（1）为了方便与已有研究结果进行直接比较，本研究中所用的实验设计与 Hou 等人(2014)以及 Li 等人(2015)所采用设计相同，因此，结果直接是可以直接进行比较的。结果发现，我们所提出的方法同 LCDM-DIF 方法一样都具有较好的一类错误控制率，不同的是 LCDM-DIF 是在 MCMC 框架下，计算量大耗时长，而本研究中所用的 MMLE/EM 算法耗时量非常小（根据 Li 等人(2015)的说法，MCMC 每次计算需耗时 5 小时左右，而本文中的 EM 算法耗时量在 1 秒左右）。（2）本研究之所以没有考虑 MH, SIBTEST 或者是 LR 法，是因为它们的统计检验力受到测验中 DIF 项目比例的影响，而 Wald 统计量则没有这一缺点，这是我们没有采用以上统计量的主要原因。以上两点我们在原文增加了相应的表述。

意见 8：作者是基于 EM 算法提出的新统计量，但根据 Li 和 Wang（2015）的研究表明，他们使用 MCMC 算法是 wald 统计量同样具有较好的一类错误控制率，作者是否还应该考虑

新的方法和 MCMC 的比较呢？

回应：感谢专家的建设性意见，我们同样接受了专家的建议。为了便于研究结果间的直接比较，本研究中的实验设计 Hou 等人(2014)以及 Li 等人(2015)所使用设计一致。因此，本研究结果是可以直接与 MCMC 的结果进行比较的。我们已经在正文的 4.1 中增加了我们所改进的 Wald 方法与 MCMC 中的 Wald 统计量法以及 LCDM-DIF 检测方法的比较的内容。

意见 9：作者在 4.3 部分说道：“由于本研究所采用的是对于 LCDM 模型进行约束而获得的 DINA 模型，因此，可以很方便的进行扩展”。但其实作者还是按照 DINA 模型进行的研究，并不是以 LCDM 为模型进行研究，因为参数给定的是 s 和 g 参数，并非 $\lambda_{j,0}$ 和 λ_j 等参数。因此，这不能算作是基于 LCDM 的研究。

回应：DINA 模型可以被认为是 LCDM 的特例，对于 LCDM 中的一些参数进行限制便可以获得 DINA 模型，具体的转换公式可参考文中公式(2)与(3)。遵循以往研究惯例(Li & Wang, 2015)，在研究中我们使用的是对 LCDM 进行约束后获得的 DINA 模型，将通过(2)与(3)转换后的参数称为 s 和 g 参数，抱歉给专家造成了困扰。具体可参考：

Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, 74, 191-210.

意见 10：本文使用了一些专业术语，如“经验一类错误”，“经验分布”，“观察信息矩阵”等，为了使一般读者，特别是想进一步深入研究的读者能看懂，建议作者给出解释。

回应：谢谢专家的建议。对于这些术语进行解释确实会有助于读者更好的理解。按照专家的建议，我们在论文中给出了一些专业术语的解释。

意见 11：写作不认真，出现很多病句，格式，数学符号不统一，对同一概念前后说法不一致等问题，让审稿人读起来很烦躁。请作者统一术语，一个数学符号就表示一个含义，一个概念就用一个词语表示。

回应：感谢审稿专家的耐心与细致。在论文的修改中，我们对于术语，数学符号以及概念的使用，进行了仔细的核查与修改。再次感谢您的建议。

意见 12：其他修改意见请参加原文批注。

回应：感谢专家的负责与严谨。根据您在原文中的批注，我们进行了相应的修改，具体的修

改部分，我们在原文中已经标红。

建议：大修后重审。

第二轮

审稿人 1 意见：

经过第一轮修改，本文有了较大的改进，我已经能够基本认可本文提出的方法。但是，作者仍然需要对以下两个基本问题作出进一步的说明。

回应：感谢审稿专家对我们研究的肯定，我们在阅读审稿意见时，发现您所提出的问题非常专业，我们同样发现您在审稿过程中非常的严谨与细致，对您的在审稿过程中的辛劳，我们表示感谢与敬意。根据您的意见我们对论文进行了相应的修改，新修改的部分使用蓝色字体显示。

意见 1：本文没有对 Li & Wang 的文章进行重复试验（replication studies），而是直接引用了原作者的结果。我建议作者需要针对他们的方法进行重复的试验，这也是科学研究的重要环节之一，具有很多的重要意义。

回应：感谢审稿专家的意见。EM 算法与 MCMC 算法分属两种不同的计算框架，各有其独特的优缺点。EM 算法的特点是耗时非常短，但是需要较为复杂的公式推导。MCMC 算法的特点与 EM 正好相反，它耗时非常长，但是实施起来较为容易。本研究中判断所构建的统计量是否符合理论分布的评价指标与 Li & Wang 研究中的评价指标均为一类错误控制率，即如果研究中所构建的 Wald 统计量服从卡方分布的话，那么，实际观察到的一类错误控制率应该近似的符合预先设定的显著性水平。这可以被认为是一个绝对的评价指标，因此可以根据这一绝对评价指标对研究结果进行评价，而无需再根据一个相对的标准去进行判断不同研究的优劣。Li & Wang 的研究结果已经显示，其在 MCMC 中所构建的 Wald 统计量服从卡方分布，同样的，本研究结果也显示，在 EM 算法中，我们所构建的 Wald 统计量也是服从卡方分布。我们的研究结果本身以及 Li & Wang 的研究都已经分别证明了各自的研究结果都是较为理想可靠的，因此我们没有选择对于 Li & Wang 文章中的试验进行重复。另外，在统计中，为了获得较为可靠的一类错误率，需要对每种实验条件组合进行多次重复，一般是 500 或者是 1000 次，MCMC 算法每次运算的耗时在 5 小时左右，如果我们同样按照 EM 算法下对一类错误控制率研究中的 18 种实验条件组合重复进行 1000 次，总耗时最低为 90000 小时，这也是本研究中没有对 Li & Wang 研究进行重复试验的另一个原因。以上两点是我们没有选择对于 Li & Wang 的文章进行重复试验的原因，希望可以得到专家的谅解与认同。

意见 2: 我目前的基本理解是, 作者针对 Hou 等人文章中不妥当的参数标准误方法进行了改进, 因此本文的基本创新点已经成立。但是作者一再强调所谓“把所有的自由估计参数”纳入计算 (包括结构参数), 这一点令我十分困惑。

1) 什么是结构参数? 请作者以 LCDM 框架下的 DINA 等价形式来具体说明。如果 Templin 等人的文章提及, 务必请提供具体的页码。目前我的理解是, 结构参数应该是用来解释考生总体的一些变量, 例如性别、教育等等人口学变量或者其他的背景变量, 或者是描述总体分布涉及的参数 (考生总体分布的均值、方差等等?)。

回应: 感谢专家这一宝贵的问题, 您的问题对于我们论文的写作非常具有启发性。根据您的
问题, 我们已经在正文中增加了如下表述: “为满足这一约束, 本研究参考 Rupp, Templin 和 Henson (2010)所使用的概念, 设 $\boldsymbol{\eta} = (\eta_1, \dots, \eta_L)'$ 为模型的结构参数 (structural parameters), 用以描述任一受测者来自特定属性掌握模式的概率, ”。并且增加了这一参考文献: Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: theory, methods, and applications*. New York, NY: Guilford.

下面, 我们参考 Rupp, Templin 和 Henson (2010)的著作, 对本文中所使用的“结构参数”这一概念进行说明。

在对模型进行 MMLE-EM 参数估计时, 研究者需要构建关于受测者作答反应的边际似然函数, 令 $P_j(\boldsymbol{\alpha}_l)$ 表示 LCDM 框架中具有属性掌握模式 $\boldsymbol{\alpha}_l$ 的受测者答对项目 j 的概率, 设第 i 个受测者的反应向量为 \mathbf{X}_i , 其边际似然函数可以用公式表达如下,

$$L(\mathbf{X}_i) = \sum_{l=1}^L p(\boldsymbol{\alpha}_l) \prod_{j=1}^J P_j(\boldsymbol{\alpha}_l)^{X_{ij}} [1 - P_j(\boldsymbol{\alpha}_l)]^{1-X_{ij}}$$

这个公式与 Rupp, Templin 和 Henson (2010)著作中第 170 页公式 (8.1) 是等同的。

$$P(X_r = x_r) = \sum_{c=1}^C \underbrace{\nu_c}_{\text{Structural}} \prod_{i=1}^I \underbrace{\pi_{ic}^{x_{ir}} (1 - \pi_{ic})^{1-x_{ir}}}_{\text{Measurement}}$$

因为所有可能属性掌握模式概率 $p(\boldsymbol{\alpha}_l)$ 之和为 1, 本研究中采用了如下公式,

$$p(\boldsymbol{\alpha}_l) = \frac{\exp(\eta_l)}{\sum_{l=1}^L \exp(\eta_l)}$$

这等同于, Rupp 等人著作第 202 页公式 (9.4)

$$v_c = \frac{\exp(\mu_c)}{\sum_{c=1}^C \exp(\mu_c)}$$

与本研究中所使用符号不同的是，他们将属性掌握模式概率 v_c ，称作是结构参数。根据其公式 9.4，可以看出属性掌握模式概率 v_c 是依赖于参数 μ_c 的，所以我们认为将 μ_c （即本研究中的 η_l ）称作结构参数更贴切一些，因此在本研究中我们将 η 称为结构参数。

关于这一参数的具体描述，专家可以参考《Diagnostic measurement: theory, methods, and applications》一书中的第 170，175，200 以及 202 页。

2) 目前我的理解是，本文的方法提出应该与所谓的结构参数没有关系，只是使用了不同的标准误估计方法（观察信息函数取代了期望信息函数）而已。

回应：在认知诊断模型中，不仅仅是项目参数需要从受测者样本的观察数据中估计获得，结构参数同样也是需要从观察数据中获得，即认知诊断模型的参数应该包括项目参数以及结构参数两类。既然如此，那么在估计信息矩阵的时候就应该包含结构参数，所以本文中的信息矩阵估计方法中包含结构参数。de la Torre (2009) 研究中所使用的信息矩阵的估计方法是仅仅包含项目参数（没有包含结构参数）的经验交叉相乘信息矩阵估计方法，不是期望信息矩阵估计方法。

意见 3：排除这个所谓的结构参数问题（我也是一直被这个所谓的结构参数困扰，所以请作者务必用最直观的例子以及相关的文献告诉我，到底什么是结构参数），我需要指出的是 de la Torre 提出的 EM 算法本身并没有错，给出的公式也是正确的，只是用期望信息函数来估计标准误是一种很不好的方法。

回应：同意审稿专家关于 de la Torre 提出的 EM 算法所获得的参数估计结果是没有错的看法，他的研究确实具有重大的价值。遗憾的是，他研究中所提出的信息矩阵计算方法不是很好。

结构参数的作用，主要是体现在模型的参数估计过程中，在专家的问题 1) 中，我们已经给出了相关参考文献的具体页码。用一种比较直观的说法就是（这一说法可能不是特别准确，但我实在是想不出还有什么更好的直观的描述它的方法了），它类似于在估计 IRT 参数时，以往研究者所设定的受测者能力值服从均值为 0，标准差为 1 的正态分布。只不过在估计 IRT 参数时，这两个值是固定的，不需要估计。在认知诊断中，这

些类似的结构参数是需要估计的。

意见 4: 我通读了 Li&Wang 的论文，但是那篇论文并没有涉及结构参数，所以我很困惑本文的模拟研究设计是如何做到与他们的文章完全一致的。

回应: 感谢专家在审稿过程中的专业、严谨与细致，我们对您付出的努力表示敬意。

Li&Wang 的研究中确实没有提及到结构参数。这主要是由于 MCMC 算法本身的性质所决定的，在 MCMC 中根据模型的后验分布可以在生成链上获得关于项目参数的估计值的多个样本点，对于项目参数的这些样本点求均值可以获得项目参数的估计值。模型中的每一个项目参数都会有多个样本点与其对应，根据这些样本点可以计算出项目参数估计值之间的方差-协方差矩阵。就是说，在 MCMC 中，项目参数的方差-协方差矩阵的计算是通过多个样本点进行的，不需要结构参数的参与。我们认为，以上应该是 Li&Wang 论文中没有涉及结构参数的原因。

另外，作者发现了现存 R 软件包 CDM 中的错误，请在文章中清楚详细地描述这个错误，这个是对研究者与使用者都很有意义的贡献。

回应: 感谢专家的建议。阅读 R 软件包 CDM 的源代码，可以发现 din 函数以及 gdina 函数（具体错误参见 CDM 4.6-0 版本中源代码 gdina 函数的第 1182 行，即使在 2012-07-20 作者曾经进行过修订）对于项目参数标准误的估计方法都是 de la Torre 所提出的，即使是代码完全正确，根据本研究的结论他们对于项目参数标准误的估计仍然是不准确的。因此我们建议研究者与使用者用本研究中所使用的观察信息矩阵的估计方法来估计参数的标准误。

根据您的建议，我们在本文的结论部分增加了如下表述“另外，我们建议认知诊断模型的研究者与使用者，当采用 EM 算法进行参数估计时，在确认认知诊断模型正确设定后，使用本研究中所使用观察信息矩阵的方法计算项目参数的标准误。”。再次感谢您的建议。

审稿人 2 意见:

经过第一轮修改后，论文对使用的相关术语、数学符号等进行了界定和统一，使得文章读起来更加顺畅。作者按照第一轮两位审稿人意见，做了相应修改，使得论文质量有了一定提升。

但修改稿中仍存在一些问題:

回应: 谢谢专家对我们工作的肯定。感谢专家对我们论文的细致审阅，您专业性的建议与意

见极大地提升了本文的可读性。从审稿意见中,我们发现,您在审稿中是非常细致与严谨的,对此我们表示敬意。根据您的建议与意见,我们对论文进行了相应的修改,新修改的部分使用的是蓝色字体表示。

意见 1: 作者将 eta 定义为模型的结构参数,但不管是从 LCDM 或者是由 LCDM 缩减后的 DINA 模型,我都没有看到 eta 的存在,映射函数 h 中的参数为 lambda 参数,但我理解作者所说的 eta 应该不是 lambda,而且按照作者对信息矩阵的修正,eta 应该是包含在 $\hat{\Sigma}_{Rj}$ 和 $\hat{\Sigma}_{Fj}$ 中的,请作者首先给出 eta 的具体含义,必要时请给出计算方法。其次,像 Li 等(2015)一文中的公式 23,也建议作者给出 $\hat{\Sigma}_{Rj}$ (或 $\hat{\Sigma}_{Fj}$) 的表达式,让读者一目了然。

回应: 感谢专家的建议。关于 η 参数,我们已经在给审稿专家 1 的第一个问题的 1) 中进行了详细的说明,并且给出了参考文献的详细页码。

η 参数参与到 $\hat{\Sigma}_{Rj}$ 与 $\hat{\Sigma}_{Fj}$ 的运算中了,但是 $\hat{\Sigma}_{Rj}$ 与 $\hat{\Sigma}_{Fj}$ 中并没有包含 η , 具体说明如下:

在认知诊断模型的信息矩阵中,由于非对角元素不为零,所以需要通过对完整的信息矩阵求逆的方法获得方差-协方差矩阵,完整信息矩阵中需要包含结构参数 η , 因此我们说它参与到了 $\hat{\Sigma}_{Rj}$ 与 $\hat{\Sigma}_{Fj}$ 的运算中。当获得了方差-协方差矩阵后,将项目 j 所对应的方差-协方差部分取出,便可获得 $\hat{\Sigma}_{Rj}$ 与 $\hat{\Sigma}_{Fj}$, $\hat{\Sigma}_{Rj}$ 与 $\hat{\Sigma}_{Fj}$ 是项目 j 在对照组与目标组中所对应的方差-协方差矩阵,所以说是 $\hat{\Sigma}_{Rj}$ 与 $\hat{\Sigma}_{Fj}$ 中并没有包含 η 。综上所述,我们说 η 参数参与了 $\hat{\Sigma}_{Rj}$ 与 $\hat{\Sigma}_{Fj}$ 的运算,但是 $\hat{\Sigma}_{Rj}$ 与 $\hat{\Sigma}_{Fj}$ 中并没有包含 η 。

根据您的建议,在正文中,我们已经给出了项目 j 的渐近方差—协方差矩阵 $\hat{\Sigma}_j$ 的表达式。增加了相关表述,项目 j 的渐近方差—协方差矩阵 $\hat{\Sigma}_j$ 可以表达如下,

$$\hat{\Sigma}_j = \begin{pmatrix} \hat{\sigma}_{\lambda_{j,0}}^2 & \hat{\sigma}_{\lambda_{j,0},\lambda_{j,K_j(1,\dots,K_j)}} \\ \hat{\sigma}_{\lambda_{j,0},\lambda_{j,K_j(1,\dots,K_j)}} & \hat{\sigma}_{\lambda_{j,K_j(1,\dots,K_j)}}^2 \end{pmatrix}$$

意见 2: 请给出本文公式 (9) 中 Δ'_0 的表达式

回应: 感谢专家的建议。我们已经将原文修改为“ Δ'_0 是观察到的受测者反应模式关于模型参数 γ 的一阶偏导 $\frac{\partial \pi_0}{\partial \gamma'}$ ”。

意见 3: 表 1 中的 Q 矩阵可以参照王卓然等 (2014) 一文中的呈现方式,作者的呈现方式太

占空间了。

回应：本研究中有 30 项目，由于空间的限制无法采用王卓然等的呈现方式（她们的研究中有 25 个项目）。

意见 4：议作者在评价指标中就给出一类错误和统计检验力的计算方法，而不要放到结果呈现中给出。

回应：多谢，已修改。见正文 3.2 评价指标中蓝色字体部分。

意见 5：作者在 4.1 部分说道：“一个非常有意思的现象是在本研究的一些实验条件下（见表 2）Wald 统计量一类错误率有细微的保守而 Li 等人研究结果中的 LCDM-DIF 以及 Wald 统计量在某些实验条件中一类错误率却有稍许膨胀”。请作者指出具体是哪些条件，并尝试分析为何会出现该现象。

回应：感谢专家的建设性意见。我们已经在原文中增加了相关解释：[从公式\(8\)中可以发现 Wald 统计量的准确性,依赖于模型参数估计值的准确性。当受测者数量较少\(如 \$N = 500\$ 时\)或者是模型中的“噪音”过大时\(如项目的猜测与滑动参数均为 0.3 时\),模型参数估计值的准确性会受到相对较大的影响,因此,在本研究的 \$N = 500\$ 以及 \$g_j = s_j = 0.3\$ 这两种条件下 Wald 统计量一类错误率有细微的保守。](#)

意见 6：从表 4 至表 7 结果可以看出， $k=3$ 的时候统计检验力均要比 $K=1$ 和 $K=2$ 时高，这是一个很有意思的现象，请作者尝试分析一下原因。

回应：在 DINA 模型中，存在两种类型的噪音，分别对应于猜测参数 g_j 与滑动参数 s_j 。猜测参数指的是受测者在没有全部掌握项目所要求的全部属性时，可以依靠猜测答对这一项目，滑动参数指的是即使受测者掌握了项目所要求的全部属性，仍有可能答错这一项目。当滑动参数所占的比重较大时(如 $K=1$ 和 $K=2$)，这一噪音所给参数估计的准确性造成影响就越大，因此， $k=3$ 的时候统计检验力均要比 $K=1$ 和 $K=2$ 时高。

意见 7：作者在讨论部分，有些用于过于绝对。例如：“研究者发现 Wald 统计量在检验 DIF 时，有着许多其他统计量所不具备的优点，然而当前研究中对于 Wald 统计量在检验 DIF 时的一类错误率的表现，存在明显的结论冲突。”这句话的阐述欠妥当。不同研究者采用的研究方法不一样，而作者也只是采用了 EM 算法，如果这么推论，作者的结论也是和其他研究

冲突的。作者并没有实现所谓的统一，所以请斟酌。例如：“本研究采用 Hou 等人以及 Li 等人研究中所使用的同等条件通过模拟发现，这些差异主要是由于 Wald 统计量计算方法的差异引起的”。该结论同样欠妥。Li 的研究就没有该问题，他们的研究是由于算法不同导致的结果，并非 wald。例如：“因此，我们认为本研究提出的改进的 Wald 统计量的计算方法解决了 DIF 研究中一直困扰研究者的 Wald 统计量在检验 DIF 时的一类错误率的表现这一重要问题”。是解决了利用 EM 算法进行 wald 检验的问题，并非 MCMC，该结论下的有点大。例如：“因此，我们认为在 Hou 等人(2014)以及王卓然等人(2014)研究中所产生的一类错误膨胀的问题，是由于不恰当的信息矩阵估计方法而引起的”。请注明是其中的 wald 检验方法，并不包括其他方法。这么说会引起读者的误解。例如：“由于本研究关注的重点在于，提出一个恰当的 Wald 统计量的计算方式，用以准确有效地来检验认知诊断测验中可能存在的 DIF 项目，澄清以往研究中所用 Wald 统计量计算方式所产生的令人困惑的结果。”该结论同样太大，请注明是使用 EM 算法时，请重新组织语言。如果作者真是想澄清以往研究中所用 Wald 统计量计算方式所产生的令人困惑的结果，应该将修正后的 wald 放到 Li 等的研究中，用 MCMC 做一遍才知道结果。

回应：感谢专家的建设性意见。

“研究者发现 Wald 统计量在检验 DIF 时，有着许多其他统计量所不具备的优点，然而当前研究中对于 Wald 统计量在检验 DIF 时的一类错误率的表现，存在明显的结论冲突。”一句中，我们想表达的意思应该是：“**先前**研究者发现 Wald 统计量在检验 DIF 时，有着许多其他统计量所不具备的优点，然而**前人**研究中对于 Wald 统计量在检验 DIF 时的一类错误率的表现，存在明显的结论冲突。”。我们已经对原先的相关表述进行了修改。

“本研究采用 Hou 等人以及 Li 等人研究中所使用的同等条件通过模拟发现，这些差异主要是由于 Wald 统计量计算方法的差异引起的”一句中并没有“Li 的研究有欠妥”的意思。在前一句中，我们已经明确表达了，“Hou 等人(2014)以及王卓然等人(2014)的模拟研究发现 Wald 统计量会存在一类错误控制率膨胀的问题，Li 等人(2015)的模拟研究却发现，其研究中所用的 LCDM-DIF 以及 Wald 统计量在使用 MCMC 计算时有着良好的一类错误控制率。”。我们研究所关注的问题是，为什么同样采用 Wald 统计量的方法，研究结论却有冲突，我们已经明确说明 Li 等人(2015)的模拟研究却发现，其研究中所用的 LCDM-DIF 以及 Wald 统计量在使用 MCMC 计算时有着良好的一类错误控制率。

“因此，我们认为本研究提出的改进的 Wald 统计量的计算方法解决了 DIF 研究中一直困扰研究者的 Wald 统计量在检验 DIF 时的一类错误率的表现这一重要问题”所要表达是，

以往研究中所采用 EM 算法获得项目参数后所采用的 de la Torre 提出的信息矩阵计算方法获得的 Wald 统计量的一类错误控制率与通过 MCMC 计算所获得的一类错误控制率的结论不同。为表达精确起见，我们已经对原文进行了修改“因此，我们认为本研究提出的改进的 Wald 统计量的计算方法解决了 DIF 研究中一直困扰研究者的 Wald 统计量在检验 DIF 时的一类错误率的表现不同这一重要问题”。

按照您的建议我们已经将“因此，我们认为在 Hou 等人(2014)以及王卓然等人(2014)研究中所产生的一类错误膨胀的问题，是由于不恰当的信息矩阵估计方法而引起的”。一句修改为“因此，我们认为在 Hou 等人(2014)以及王卓然等人(2014)研究中所产生的 Wald 统计量一类错误膨胀的问题，是由于不恰当的信息矩阵估计方法而引起的”。以避免引起误解。

我们已经将“由于本研究关注的重点在于，提出一个恰当的 Wald 统计量的计算方式，用以准确有效地来检验认知诊断测验中可能存在的 DIF 项目，澄清以往研究中所用 Wald 统计量计算方式所产生的令人困惑的结果。”一句修改为“由于本研究关注的重点在于，在 EM 算法框架下提出一个恰当的 Wald 统计量的计算方式，用以准确有效地来检验认知诊断测验中可能存在的 DIF 项目，澄清以往研究中所用 de la Torre (2009, 2011)所提出的信息矩阵方法计算 Wald 统计量时所产生的令人困惑的结果。”。多谢您的建议。

意见 8: 作者说道：“首先，样本大小对于 Wald 统计量有重要影响，因此，后续研究中可以使用本研究中所用 Wald 统计量考察这一因素对于 DIF 的影响”。本研究中作者已经考察样本容量的影响了，为何还作为今后可以的研究方向之一呢？

回应: 感谢专家的问题。Wald 统计量依赖于参数估计的准确性，参数估计准确性又受到样本容量大小的影响。在本研究，为了方便以及清晰的与以往研究结果进行比较，我们只采用了以往研究中所使用的两个样本容量大小 $N = 500, 1000$ 。我们认为究竟在多大的样本下，才能有好的的一类错误控制率以及统计检验力是值得后验研究者进行研究的问题之一。另外，尽管我们发现当 $N = 1000$ 时，在 DINA 模型下 Wald 统计量有着好的一类错误控制率及统计检验力，那么在其他模型中（如饱和的 LCDM 或者是 DINO, C-RUM 等等），这一样本容量大小是否有同样的结论，依然是值得后续研究者研究的问题。因为，在本研究中我们所采用的是一个广义的框架，所以我们认为“后续研究中可以使用本研究中所用 Wald 统计量考察这一因素对于 DIF 的影响”。

意见 9: 论文自检报告第 15 条要求作者删除基金号，作者填写的“是”，但我仍然从系统中

看到了基金号，请问这是什么原因？

回应：根据我们的理解，第 15 条指的应该是我们所提交给审稿系统的 word 文档中不含有“作者与单位、基金号、英文摘要等作者信息。”，所以填写的“是”。提交论文初稿时，按照投稿系统要求，需要填写基金号。至于外审专家二为什么能够在审稿系统中看到基金号，我们猜想这应该是投稿系统中的“bug”吧，我们已经将这一情况电话反馈给编辑部老师。

第三轮

审稿人 1 意见：

经过两轮的修改，文章已经有了较大的改进，基本思路已经非常清晰，对所有的问题已经有了较好的解释。在一些对文字与公式的小修改的基础上，可以发表。

目前一个需要修改的问题是公式 10 与 11。首先，公式 10 出现了一个笔误， $\frac{p_0}{\pi_0^2}$ 中分母应该没有平方，也请参照文章的第一版。第二，公式 10 与 11 写的形式不一样，10 采用了向量的方式，而 11 采用了加和符号，请统一。第三，公式 10 中的一阶导用了一个专门的符号 Δ'_0 ，但是公式 11 中没有用，请统一。

回应：感谢审稿专家的建设性意见。为了便于读者阅读及对原理的把握，我们重新对公式 10 与 11 的表达形式进行了统一。我们对于公式 10 与 11 统一采用了向量或矩阵形式，对符号也进行了统一，删除了一些冗余的符号，对文字表达也进行了重新的修改，以求清晰与简洁。再次感谢专家的严谨与细致。

审稿人 2 意见：

本文基本按照二位审稿人要求修改完毕，已符合发表要求，建议发表。

回应：感谢专家在审稿过程中的严谨与辛劳。