

# 《心理学报》审稿意见与作者回应

题目：分类精确性指数 Entropy 在潜剖面分析中的表现：一项蒙特卡罗模拟研究

作者：王孟成 邓俏文 毕向阳 叶浩生 杨文登

---

## 第一轮

### 审稿人 1 意见：

本文作者使用 Monte Carlo 仿真法对 Entropy 及其变式指标受多个因素影响的情况进行探索。在当前潜剖面/潜类别研究应用不断增多的情况下，本研究有着较强的应用价值。论文整体写作规范，方法恰当。但同时~~在~~在写作上还存在如下不足之处，期望能加以改进。

**意见 1：**文章结构，论文一共分了 8 个小节，这样的结构在《心理学报》发表的论文中并不常见，与 Lubke 和 Muthén (2007)的论文较为相近；实际上，可以把第 1 至第 4 小节合并在一起，写个长的引言和综述部分；第 7 至 9 小节也可以考虑合并调整。

**回应：**感谢评审人提出的建议。我们已根据您的建议对文中的 8 个小节进行合并与调整，最后是分为 4 个小节(引言、模拟研究、结果、结论与讨论)。

**意见 2：**文中简介“潜剖面分析”时，过于简单，一些关键点交代不清。如第一句话为“潜剖面分析主要用于处理连续数据（见图 1）。”，但并未对图 1 进行任何解释，包括其中的符号，让并非对潜剖面分析原理十分了解的读者很难明白；又如“上式简化为：”，但下面公式的左侧完全发生了改变，不知道这里的上式指得是什么；“第 k 个剖面的第 i 个指标的方差可以分解为：”，但是下面的公式看起来是对第 i 个指标方差的分解，而不是第 k 个剖面。

**回应：**感谢评审人细心的指正，非常抱歉我们在文章中没有交代清楚，这里的表达方式给您的阅读带来不便。确实，文中没有对图 1 进行解释。您提到的“上式简化为：”问题也确实是没有交代清楚。对于“第 k 个剖面的第 i 个指标的方差可以分解为：”的问题，我们觉得可能是表述不当让您产生了误解。

鉴于以上三个问题都是针对潜剖面分析这一小节的，为了便于读者的理解，我们已经在文中进行了整体的修正与补充。

**意见 3：**第 4 小节中，“Lubke 和 Muthén (2007)的模拟...准确率超过 90%。”一段，所说的并非他们研究的不足，而是在介绍其研究，不应该放在本部分；另外，这段话，前后表达的意思不一致，前面说其研究考察了分类精确性的影响因素，但后面介绍的是分类标准临界值。

**回应：**非常感谢您的建议和意见。“Lubke 和 Muthén (2007)的模拟...准确率超过 90%。”这句话确实是对 Lubke 和 Muthén (2007)主要内容的简单介绍，目的是为了引出他们研究的局限性，并对他们研究的局限性和不足之处进行扩展，探索在他们研究之外的其他因素的条

件下，哪些因素会影响分类精确性，以及 Lubke 和 Muthén (2007)发现的分类标准的临界值是否依然适用其他的条件。

另外，您提到的第二个问题是对的。这句话前后之间确实没有表达清楚，据此我们已经进行了修改。

**意见 4:** 文中 3.3 节用较大篇幅描述了潜类别间距和类别内方差的影响，但论文并非教科书，不一定要使用大量图表展现这两个因素如何影响，而应把更多篇幅放在自己的研究问题本身。

**回应:** 非常感谢您的反馈和意见。我们已经将这部分内容进行了压缩，去掉了图。

**意见 5:** 仿真实验中，作者只使用了 3 个潜类别和 5 个潜类别的结果；但是，其主要对比的 Lubke 和 Muthén(2007)研究中为 2 个潜类别，为了更好地对比和分析其中规律，应补充 2 个潜类别的结果。

**回应:** 感谢评审人的意见和建议。在进行研究设计的时候，我们也对此问题进行过认真的讨论。鉴于我们的研究目的是考查分类精确性指数 Entropy 及其相关变式(CLC, ICL\_BIC, 校正的 ICL\_BIC)受样本量、潜类别数目、类别距离和指标个数及其组合影响的情况。我们考虑的主要因素与 Lubke 和 Muthén(2007)的研究是不同的,他们的研究主要是考查样本量、潜类别距离、协变量和模型复杂性等因素对因子混合模型(factor mixture model)的参数估计和分类精确性的影响。所以在我们的模型中，设计潜类别数为 2 的情况并不能验证 Lubke 和 Muthén(2007)的研究结果，而且这也不是我们的研究目的所在。

所以很感谢评审人提出这个问题。针对您的意见我们补充了潜类别数为 2 的模拟。考虑到我们模拟设计的因素和水平比较多，这里我们只对部分因素和水平进行补充设计，结果呈现在附表 3 和附表 4 中。

从附表 3 中可以发现，潜类别数为 2 时：

- 1) 类别距离相同的条件下，Entropy 和分类精确率随指标数的增多而增大，错误率随之减小。
- 2) 指标数相同的条件下，Entropy 随类别距离的增大而增大；类别距离为 3 时变化情况更加明显。

从附表 4 中可以发现：指标数为 4 时，类别距离相同的条件下，Entropy 随类别数的增多而增大；但精确率随类别数的增多呈减小的趋势。

这样的结果和潜类别数为 3 或 5 时同等条件下的结果基本一致。另外，这样的结果也表明 Entropy、分类错误率和精确率受类别数的影响，且不容易给出 Entropy 的临界值。

**意见 6:** 实验中，分别使用了 4 个、8 个、12 个、20 个指标的方案，然后从表 1 看，作者给所有指标的均值都设为相同；如果设为不同，会产生结果呢？

**回应:** 衷心感谢您提出的问题与建议。鉴于在一个模拟研究中考虑过多的因素，不利于

研究结果很好地呈现在有限篇幅内，我们才决定除主要研究因素(潜类别数、指标数、类别距离、样本量)之外的因素都设定为固定值(另外，我们之前的一篇模拟研究，审稿专家觉得我们考虑的因素太多，多次建议我们少考虑一些变量)。不过以后的研究可以考虑设定不同的指标均值来探究分类精确性的变化情况。

**意见 7:** 在大样本条件下，发现 Entropy 值和错误率没有随指标数增多而变好，这是否与作者选择重复实验 100 次有关，如果再增大一些，是否可能有更为稳定的结果。

**回应:** 评审专家的这个问题非常专业。重复次数确实可能对模拟研究产生重要的影响。考虑到我们研究涉及的因素和模型复杂程度所导致程序运行的时间问题，以及参考前人模拟研究中的重复次数(e.g., Lubke & Neale, 2006; Tein, Cox, & Cham, 2013)，我们最后将每种组合设定为重复 100 次。

另外，根据您的建议，我们也进行了重复次数为 500 的实验设计，同样是只对部分结果进行重新设计，结果呈现在附表 5(以样本量=1000 为例)。

从附表 5 中可以发现：

1. 类别数=3，指标数=4 时，Entropy 随重复次数的增多而减小，但其对应的分类错误率和精确率基本保持不变。
2. 类别数=3，指标数=20 或类别数=5，指标数=4 时：
  - 2.1 类别距离为 0.5 时，Entropy 随重复次数的增多而增大，但精确率变化不大。
  - 2.2 类别距离为 1.2 或 3 时，Entropy 随重复次数的增多反而有减小的变化趋势，但精确率依然变化不大。

也就是说，增大重复次数时，Entropy、错误率和精确率受到类别数、类别距离和指标数的共同影响。这和重复次数为 100 时的结果是基本一致的，即 Entropy 随类别数、类别距离、指标数的变化而变化，很难确定唯一的临界值。另外，这些结果也表明分类精确率非常稳健不容易受重复次数的影响。

**意见 8:** 文中所有公式应加编号，以便阅读。

**回应:** 感谢评审人提出的建议。我们已对所有公式进行编号。

**意见 9:** 3.2 小节公式和段落缩进均存在问题。

**回应:** 感谢您的细心指正。我们已经对这些公式和段落的问题进行修正。

**意见 10:** 脚注 1 所表述的内容不清晰。

**回应:** 非常感谢您的细心指正。我们已经对脚注 1 的内容进行重新修改。

**意见 11:** 图 4 和图 5 中的图例均没有显示完整。

**回应:** 感谢评审人的指正。如果您这里的意思是图例的缺失，例如图 4 中的两个图，只有一个图给出了图例，另一个图没有给出图例。那很抱歉我们在图中没有描述清楚，给您的阅读带来了不便。因为图 4 或图 5 中的两个图的图例是一样的，都是有 6 种情况，也都用相

同的颜色和线段表示，所以在每一个图中，我们都只给出了一个图例。

为了让读者更容易理解，我们对文中内容进行了如下补充(红色部分为新增加的):

图4  $k = 3$  指标数= 4(左图)和指标数= 20(右图)时 Entropy 和分类错误率随类别距离 (MD)和样本量的变化情况 (两图的图例相同)

图5  $k = 5$  指标数= 4(左图)和指标数=20(右图)时 Entropy 和分类错误率随类别距离 (MD)和样本量的变化情况(两图的图例相同)

如果我们理解错了您的意思，还请您进一步指正。

#### 审稿人 2 意见:

作者在论文中讨论了分类精确性指数 Entropy 受样本量，潜在类别数目等因素的影响情况，并进行了模拟实验，本文写作较好，有一定实际意义，同意发表。

回应: 感谢肯定!

---

## 第二轮

尊敬的审稿专家，感谢您审阅本文，您专业而明锐的反馈，让本文更加完善、严谨，感谢您辛勤的付出。我们已就您提出的意见和建议进行了逐一回复，如果我们在理解您的意思上存在错误之处，还请您进一步指正。

#### 审稿人 1 意见:

论文经过修改后，质量得到了进一步提高。

希望对第 4 部分，结论与讨论部分进行一些修改，目前 4.2 节内容中，四条结论后面又跟有讨论，并非完全是“本研究的结论”。建议参考其他论文的讨论部分加以修改。

回应: 感谢您的审阅。我们已参考相关论文对第 4 部分进行了重新布局，具体做了如下调整:(1)把结论和建议单独出来作为文章的第 5 部分。这样使得第 4 部分的讨论更加独立、紧凑。(2)把第 4 部分的顺序进行了调整，把原第一条结论放到最后，这样使得讨论部分层层递进，由单个影响因素到多个影响因素。

#### 审稿人 2 意见:

意见 1: 本文对 LPA 的分类精确度的影响因素进行模拟研究，从研究意义上，分类精确性是混合模型评估的重要参考标准，然而这个标准通常仅作为“门槛”标准（即只需要达到.8 或.9 以上），通常使用信息指标 BIC、AIC、BLRT 等进行模型比较。那么，对 entropy

的影响因素进行研究，对实际的应用有何启示，需要作者进一步思考。

回应：我们非常同意审稿专家的观点即混合模型（mixture model）的比较和选择通常采用信息指标 BIC、AIC、BLRT。然而本研究的主要目的是评价分类精确性指标 entropy 的影响因素，而我们模拟和分析的模型也是真模型，所以不涉及模型选择的问题。另外，先前的研究，如 Peugh 和 Fan(2013)以及 Tein, Coxe 和 Cham(2013)已经对 entropy 用作模型选择时的表现进行过系统模拟。

针对审稿专家提出的“对 entropy 的影响因素进行研究，对实际的应用有何启示”我们认为审稿专家的建议非常好，为此我们在“4.2 给应用研究者的建议”部分增加了 entropy 的影响因素对实际的应用的启示的相关内容。

**意见 2：**研究中的模拟实验根据马氏距离，但是这种参数设计未能贴切实际的潜在剖面分析结果，若数据之间的变异只在于量化的差异，那为何不直接使用聚类分析？这是该文最为重要的问题。

回应：不可否认，审稿专家的疑问具有普遍性。已有不少研究对两种方法进行了比较，结果发现潜类别分析优于聚类分析（e. g., Magidson & Vermunt, 2001）。聚类分析和潜在剖面分析功能类似，但两种方法之间存在本质差别。前者不存在分类误差而后者是基于模型的分类型方法，允许存在分类误差，正因为如此才有分类精确性指标 entropy。

此外，模拟实验采用马氏距离作为类别间距的标准也是先前绝大多数研究采用的方式。

Magidson, J., & Vermunt, J. K. (2001). Latent class models for clustering: a comparison with k-means. *Canadian Journal of Marketing Research*, 20(3), 36-43.

**意见 3：**审稿人认为潜在剖面分析的一大用途是反映类别之间质化的差异，而非仅仅存在量化差异。因此，建议增加与实际数据相符的模型参数，具体可以参考某篇实证研究展开模拟研究。

回应：本研究模拟所采用的参数是结合了先前相关研究(e.g., Peugh & Fan, 2013; Tein, Coxe & Cham, 2013)和实际应用情境。参考某篇实证研究开展模拟，我们认为单个研究不存在普遍性，会造成模拟结果缺乏普遍意义。

**意见 4：**此外，该文结论四中“要得到好的分类精确性，且不用担心样本量的影响，类别距离最好达到 3”。这一结论在实际应用中是否可行，在模型建立中无须考虑样本量对模型拟合与分类统计功效的影响？

回应：感谢您的问题。“要得到好的分类精确性，且不用担心样本量的影响，类别距离最好达到 3”，这个结论是根据模拟结果得到的。模拟结果表明，不管样本量如何，类别距离为 3 时的分类精确性都很高。至于这一结论在实际应用中是否可行？我们认为在满足所设

模拟条件的基础上是可行的。

**意见 5:** 另外，实际数据中的类别差异仅仅是指标之间单一的距离？当各指标在类别之间的差异有所不同时（例如指标 1 中 A 类的比 B 类高，而指标 2 中 A 比 B 低），也需要同样的类别距离才能达到理想的分类精确性吗？

回应：我们设置的模拟条件是相对简化的，即类别间所有题目存在相同的距离，但在实际中这个距离可能是不相等的。但当“各指标在类别之间的差异有所不同时（例如指标 1 中 A 类的比 B 类高，而指标 2 中 A 比 B 低）”，我们认为并不会影响结果，因为单个的条目对类别组内的个体来说是相同的，类别分组也正是采用这种方式进行的。

当然还存在另外一个问题即不同类别距离的条目混合在一块时的情况如何？这需要另外的模拟研究去探究。

**意见 6:** 本文的一项结论“样本量越大，entropy 越小”与前人的研究与理论相悖：样本量大时，分类的统计功效更大，对类别的区分度也更大。如 Lubke, Neale(2006) 研究发现“Increasing the within-class sample size results in an increasing proportion correct model selection when considering AIC and aLRT”。当类别的 separation 越小，越需要增加类别内的样本量，以有足够的 power 以正确区别出该潜类别。请再仔细检查研究结果，若依然相悖，需要找充足的文献解释说明。此外，建议作者进一步补充阅读其他文献，例如 Dias, 2004; Lukociene, Varriale, and Vermunt, 2010，以基于这些研究成果全面综合地分析类别 separation、样本量与 entropy 之间的关系。

回应：针对这个问题，我们不认为我们的结果和前人的研究与理论相悖，因为本文的研究问题前人并没有直接涉及。Lubke 和 Neale(2006)的研究结果是针对样本量对模型选择的影响，并未涉及对分类精确性的影响，而我们研究是样本量对分类精确性指标 entropy 的影响。Lukočienė, Varriale, & Vermunt (2010)是针对多水平数据进行潜类别分析时如何确定合理的类别个数的研究，而我们的研究并不涉及模型选择（类别数目的保留）。Dias 的研究是针对潜类别分析时如何设定类别距离。

谢谢审稿人提供的文献和建议。上述文献我们已经阅读，之所以没有引用是因为与我们的研究内容无直接关系。

Dias J. G. (2004). *Controlling the Level of Separation of Components in Monte Carlo Studies of Latent Class Models*. In: Banks D., McMorris F.R., Arabie P., Gaul W. (eds) *Classification, Clustering, and Data Mining Applications. Studies in Classification, Data Analysis, and Knowledge Organisation*. Springer, Berlin, Heidelberg

Lubke, G., & Neale, M. C. (2006). Distinguishing between latent classes and continuous factors: Resolution by maximum likelihood? *Multivariate Behavioral Research*, 41, 499–532.

Lukočienė, O. Varriale, R., & Vermunt, J. K. (2010). The simultaneous decision(s) about the number of lower- and higher-level classes in multilevel latent class analysis. *Sociological Methodology*, 40(1), 247–283.

**意见 7:** 文中提到“小样本下，指标数越多，entropy 结果越好”。而在实际应用中指标数越多，所需估计的参数越多，模型越难拟合，entropy 也难以满足.8 或.9 以上。对于这种情况，如何解决？4.2 建议中“小样本时，可以通过增加指标数来提高分类精确性”，例如  $N=50$  时，指标数 20 个，若存在 5 个类别，固定类别方差的情况下也至少有  $20*5+20+4=124$  个参数，仅有 50 个样本能对如此复杂的模型进行参数估计吗？建议作者再确认模型估计的收敛结果，或有无其他诸如 boundary solution 或过度拟合的问题。

回应：感谢您的问题。我们已经对全部小样本的模拟结果进行了重新检查，并未发现结果有任何异常。同时我们也尝试了对 17 个指标的实际数据进行两个小样本( $n = 41$  和  $n = 70$ )分析( $C=3$  和  $C=5$ )，所得 entropy 结果和本文模拟结果类似，3 个类别  $n = 41$  时的 entropy = .998,  $n = 70$  时 entropy = .918.

**意见 8:** Entropy 作为分类精确性的指标只是潜在剖面模型在模型选择中需要参考的指标之一，另外还需要参考模型拟合度和简洁性，通常就是在精确性和类别数目之间权衡。若在评价分类精确性时还需要考虑类别数目，这样则将重复考虑类别数目的问题，并且很可能对类别数目的判定时有所偏差。

回应：我们非常同意您的观点。本研究只考虑了拟合真模型或正确模型时 Entropy 受各种因素影响的情况，至于拟合虚假模型时的情况本文并没有考虑，当然这也是一个非常重要并值得将来进一步研究的课题。再次感谢您的建议。

**意见 9:** 鉴于以上研究设计、结果与解释、研究意义等问题，希望作者重新考虑模拟研究中的参数设定，并且结合前人文献对研究结果有更综合的解释，再提出更有说服力的研究结论。

回应：感谢您审阅本文。可能是由于表述不清让您产生了疑问，为此我们认真地对您提出的上述疑问和建议做了细致的回应，请您再次审阅。如果我们的理解有误，还请您进一步指正。

---

### 第三轮

论文经过修改后，对该研究的实际意义、研究设计等方面有了更清晰的阐述，同时也对审稿人的疑问做了较为详尽的回应，消除了审稿人大部分的疑虑。以下还有一些疑问需要作者再进一步思考与修改。

回应：感谢您的肯定和积极反馈，我们会认真考虑您的意见并尽最大努力给予修改和补充。当然，如果存在错误之处还请您进一步指正。

**意见 1:** 关于意见 6 的回应，审稿人认为作者还需综合考虑 entropy、分类准确性及其与其他因素之间的关系，对研究结果作更充分的理论解释。虽然所提及的几篇文献并没有直接研究 entropy，但模型选择的准确性恰恰是和 entropy 有很大关系：许多研究用 entropy 来反映潜在类别变量与外显变量的关系强度，也有用于表示，也是决定正确选择模型的重要因素（如 Lukociene, & Vermunt, 2010）。同时，Lukociene, Varriale 和 Vermunt（2010）发现 entropy 受样本量的正性影响。换一个角度，若按本研究的结果，样本越小，entropy 越大，更容易正确选择类别数目，这与前人研究结果（如 Lubke, Neale, 2006）“样本量越大，模型正确选择率越高”相悖。既然本研究发现了与前人研究有所不同的结果规律，审稿人认为有必要综合阐述其原因，为以后的模拟研究找到突破口。

Lukociene, O., Vermunt, J. K (2010). Determining the number of components in mixture models for hierarchical data. In: Fink, A., Lausen, B., Seidel, W. and Ultsch, A. (eds.), *Advances in data analysis, data handling and business intelligence*, 241-250. Springer: Berlin-Heidelberg.

回应：感谢您提供额外的文献，由于这篇文献是图书的章节所以我们之前并没有检索到。这篇文章主要探讨多水平混合模型时模型选择即类别数目的保留，考查的是 AIC, BIC 等信息指标。

在 LCA 中条件概率的大小反应了观测指标和潜变量(潜类别变量)之间关系强度(Collins, & Lanza, 2010)。之所以采用 Entropy 作为 class separation 的指标是因为 Entropy 基于个体后验归属概率，而后验归属概率又是基于条件概率得来的。采用 Entropy 作为 class separation 的指标是比较“曲折”的途径。

Lukociene 等的上述文章采用不同的条件概率作为 class separation 的参数，并给出不同条件概率所对应的 Entropy，然而文章并未采用 Entropy 作为模型选择的指标。

“换一个角度，若按本研究的结果，样本越小，entropy 越大，更容易正确选择类别数目，这与前人研究结果（如 Lubke, Neale, 2006）“样本量越大，模型正确选择率越高”相悖。”

针对这一意见解释如下：我们的研究所得结果是“样本越小，entropy 越大”并没有得到“更容易正确选择类别数目”的结论，因为我们拟合的模型是真模型，不存在模型选择的问题（即正确选择类别数目）。因为我们没有检验 entropy 在模型选择上的表现，所以结果无法与前人研究（如 Lubke & Neale, 2006）“样本量越大，模型正确选择率越高”的结论做比较。另外，Peugh 和 Fan (2013)考查了 entropy 及其变式在模型选择时的表现。他们的结果发现，相



比于大样本量 ( $N = 3000$ )，小样本量 ( $N = 300$ ) 时 entropy 正确选择模型的概率更高(p632, table5)。

**意见 2:** 与意见 6 相关，文中 2.2 根据预实验提出研究假设，正如作者所提及，模拟研究的条件有限，不能无条件地推广，那么预实验用了哪些条件？似乎文中没有报告。另一方面，审稿人认为，数据模拟研究的假设最好基于数理统计的原理思想，或者至少基于前人研究结果推断，仅从某预实验或某些条件的模拟结果提出假设，恐怕无法全面把握其数理以便找到关键规律。

回应：感谢您的提醒，我们已在文章中做了如下脚注：

预实验所设条件除了本文报告的条件外，还设定了类别内指标相关 ( $r = 0.25$ )。由于设定指标相关对 Entropy 值并不会产生影响所以文中并未报告。

谢谢您的建议，我们也非常认同您的观点，当然我们所选择的研究条件也是在前人研究的基础上选取的，具体已在文中阐述。

**意见 3:** 关于意见 7，同理，在 17 个指标下有  $17*5+17+4=106$  个参数，所估计的参数比样本量多，似乎不符合模型识别的条件 (Collins, & Lanza, 2010)。或者希望作者能根据文献或从数理上支持说明文中的这一结果。

Collins, L. M. and Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*, Hoboken: Wiley.

回应：关于模型识别，混合模型和结构方程模型类似，都要求满足最基本的前提条件即已知信息多于未知信息(Collins & Lanza, 2010, p92)<sup>1</sup>。根据邱皓政(2008)提供的模型识别时待估参数和自由度的公式 (公式 3.17 和 3.18, p36)，影响模型识别的关键是条目量，假设存在三个观测指标 (每个指标 5 个选项) 可以形成  $5*5*5$  个单元格 (列联表) 即可提供数据信息点 (初始自由度是  $125-1=124$ )，而需要估计的参数有  $T-1$  个 (假设  $T=5$ ) 潜在类别概率和  $[(5-1)*3]*5$  个条件概率，所以此时的自由度是 60 ( $124-64$ )。自由度大于 0，所以模型可以识别。

或者从 LCA 最根本的逻辑来解释这个问题。LCA 的目的是通过个体在观测指标上的作答形态 (response pattern) 将其分成不同的类别组 (class)。假如有 20 个题目，被试只有 10 人，理论上来说此时最多可以分成 10 个类别即每人在 20 个条目上的作答形态都不一样。类似因素分析，10 个题目可以分成 10 个因子。更极端的例子，假如只有 1 个被试回答了上述 20 个条目，那么类别就只有一个而且类别内就 1 个人。当然在实践中不可能存在这样的情况。

---

<sup>1</sup> Collins & Lanza (2010, p92)只是概括的讨论了 LCA 的识别问题。

$$\text{总待估参数} = (T-1) + T(I-1) + T(J-1) + T(K-1) = (I + J + K - 2)T - 1 \quad (3.17)$$

$$df = (JK-1) - [(I + J + K - 2)T - 1] \quad (3.18)$$

式中  $T$  表示潜类别变量， $IJK$  表示 3 个观测指标。

邱皓政(2008). *潜在类别模型的原理与技术*.北京:教育科学出版社.

**意见 4:** 文中 P17 2.4 段落的第四行“平均所有值”，这是指类别概率的所有值还是什么？  
回应：这句话完整的表述是 另外，通过平均所有  $q_{c_i, c_i}$  值获得分类错误率。式中的概率是公式 6 的分类错误率。可能您看到的版本公式漏掉了（我们提交的版本是有的）。

再次感谢您的积极反馈和建设性的意见。

---

#### 第四轮

作者对审稿人的大部分疑问作了较为清晰的回应，但针对意见 3，还需作者进一步思考：首先，这是一个潜在剖面模型，而非潜在类别模型，二者的模型公式不同，所涉及参数（LPA 主要是条件均值、潜在类别概率、方差）也不同，而邱皓政（2008）仅涉及潜在类别模型的参数个数计算。

其次，Collins & Lanza (2010) P92 提到，larger sample size means there is more info in the data.这也是决定模型拟合的关键因素之一。一般而言，对样本量的需求随模型的复杂程度而有所提高，即参数越多，越需要更多的样本量。

再次，根据作者所举的例子，若只有 1 人做 20 题，确实理论上只会会有一个类别，但是这个情境实际上能进行模型的参数估计吗？

综上，若作者认为这个模拟条件的模型识别没有问题，请提供充足的 LPA 参考文献或数理推导。

回应：感谢您严谨的治学态度。您的问题让我们对 LPA 有了更加深入的理解。通过查阅相关文献我们发现最近几年发表的论文和教科书(e.g., Collins & Lanza, 2010; Peugh & Fan, 2013; Sterba, 2013)均没有对潜在剖面模型识别问题的介绍和探讨，然而更早一点的专著已对识别问题进行了深入的分析。Titterington et al. (1985)指出除了有限混合均匀分布(mixture of uniform densities)以外，其他有限混合连续分布(finite mixtures of continuous densities)都是可以识别的。本研究模拟的 LPA 的指标是标准正态分布，所以所设定模型属于有限混合连续分布，因此模型是可以识别的。类似的介绍也见(McLachlan & Peel, 2000, p26-28)。

另外, Collins & Lanza (2010)提到的 larger sample size means there is more info in the data. 我们对这句话的理解是这样的: 样本量越大, 样本包含的信息越丰富, 模型估计的结果越可靠, 即统计功效越高, 所以说这句话并不是针对模型识别的。当然不可否认, 在实践中会出现理论上可以识别但无法收敛的情况, 针对这种情况增加样本量是可能的解决途径。本研究拟合的是真模型并未出现不收敛的情况。

非常感谢您的审阅, 您严谨的治学态度不仅让本文更加完善也加深了我们对混合模型的认识和理解。

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York, NY: Wiley.

Titterton, D. M., Smith, A. F. M., and Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*. New York: Wiley.