《心理学报》审稿意见与作者回应

题目：心理实验数据的联合建模：反应与反应时的混合影响
作者：郭小军, 焦玉月, 柏小云, 罗照盛, 李弘

第一轮

审稿人 1 意见：

在线性混合效应模型的框架下，文章创新地提出了对反应时和作答反应联合建模的方式，并通过模拟研究证明，所提出的混合效应模型的分层模型具有更高的参数估计准确性，第一类错误率和检验力。文章关注了线性混合效应模型在心理学实验数据中的拓展和应用，具有一定的创新价值。下面提出一些疑惑和建议，供作者参考。

意见 1：传统的方法是分别对作答反应和反应时建模，而作者提出了联合建模的方法，该方法相对于传统方法的优势是什么？例如，在教育测量领域，可以对反应时和作答反应联合建模，这是因为反应时和作答反应的背后是速度和能力这两个潜变量。联合建模可以通过多模态的信息，使得潜变量的估计值更准确。而在实验心理学领域，研究者关注的是外显的反应时、作答正确率(显变量)，就是要看不同实验处理是否造成其差异，而不是背后某潜变量的差异。所以作者应当着重阐述联合建模的优势。

回应：感谢审稿专家的意见。

(1)心理测量与实验建模的一般性目的都在于理清因变量的影响机制。无论是心理实验还是教育测量，都是通过外显行为来推断被试的心理过程与认知机制。在心理实验中，将反应时的影响因素分为被试随机截距和斜率、刺激项目随机截距和斜率，以及固定效应(总体均值和处理效应)，进而通过构建线性混合效应模型进行分析。教育测量将项目准确率的影响因素分为被试潜在特质、项目难度和区分度，进而构建了项目反应理论模型(比如两参数 logistic 模型)。心理实验目标是使自变量(不同处理水平)估计更纯粹，而教育测量是使被试特质估计更纯粹。为了使心理实验联合建模的内容描述更符合心理实验研究者习惯，文中从反应与反应时显变量角度对各联合模型的构建进行了重新描述。

(2)文中提出的联合建模相对于传统单独分析的优势在于：①联合建模能够充分整合被试在心理实验操作中不同信息来源。心理实验常见的因变量为反应与反应时，反应能提供被试在实验操作过程中任务难易程度的信息，而反应时则提供了被试在实验操作过程中的流畅性信息，联合建模可以将两种不同类型的信息进行整合，以更全面评估被试的心理与认知过程。②联合建模有利于心理实验不同数据信息间的互补。通过信息的互补，从而比单独分析提供更高的参数估计精度，更好的模型拟合效果。③联合建模不仅仅关注单个因变量的影响机制(比如反应时或反应)，同时探讨了不同因变量间的关系机制，因此文中在阐述联合建模本身特点时，还对联合建模的优势进行了说明。

总之，联合建模为心理实验不同数据类型的整合提供了一种强有力的分析框架，能更全面地揭示被试的心理和认知过程，以及提供更高的参数精度和解释力。

意见 2：研究的核心是提出了混合效应模型的分层模型 HM-MEM，该模型假设“被试在刺激项目上的正确操作假设为被试在实验任务上的认知能力与加工时间是充足的；对于错误操作反应与反应时，被试可能对实验任务的认知能力充足，而任务加工时间不充分或时间限制进而导致操作错误(Miller et al., 2018; Ranger et al., 2015; Rouder et al., 2015)，也可能因为被试的认知能力不足，此时被试在实验任务上的加工时间是做无用功(Lee et al., 2015)，因此被试

的操作错误原因是认知能力与加工速度不足的混合影响构成。”此外，还存在两种模型及其相应的假设，即 HM-MEM-D，假设反应时和作答反应独立，M-LMEM，假设作答只受加工速度的影响。在实际情况下，如何明确作答反应过程符合哪种假设从而选择正确的模型？选择了某一种模型是否就接受了某一种强假设，而如果被试具有异质性，不同被试的反应原理(假设)不同，又应当如何处理？

回应：感谢审稿专家的意见。

(1)在实际应用中，对于明确作答过程符合哪种假设，进而选择正确的模型，可以通过理论与数据驱动共同进行确定。在心理实验情景上，混合 MEM 理论上更适用于既有难度又有时间压力的实验任务；独立 MEM 更适用于有难度但时间完全充裕的实验任务；而速度 MEM 则更适用于非常简单但有时间压力的实验任务。同时，为了验证理论假设，通过实际数据驱动，基于模型拟合指数(比如 AIC、BIC 或 WAIC 和 LOO)进行比较，模型拟合越佳，说明数据更倾向该模型假设。因此，可以结合于理论和数据驱动以确定实验操作反应过程的假设。

(2)选择某一种模型意味着接受了该模型的强假设。模型假设定义了模型对数据的基本理解，以及对数据的解释与预测。这些假设有助于简化现实世界的复杂性，使问题可通过数学和统计方法来解决。当选择一种模型时，对于数据的解释与理解都是基于模型假设的前提进行。模型有些假设在现实中难以满足，强假设就可能导致模型的适应性较低。因此，在接受强假设时，也必须认识到模型在数据应用上的局限性。

(3)如果被试具有异质性，不同被试的反应原理(假设)不同，可以采用以下处理方法：①模型本身具有一定的鲁棒性或稳健性。一个鲁棒性较高的模型能够在不同情况下保持较好的测量精度和稳定性，不会因为个别异常值或者违反模型假设的情况而导致整体测量结果的严重偏差。因此，对于较少的具有不同反应原理的被试，对模型的应用不会有太大的影响。同时，模型对数据的拟合较好，本身就意味着相对其他模型，当前的模型是一个相对更合理的模型，而不是一个完全的拟合问题。②基于被试的拟合统计量进行处理。可以如同教育测量模型一样，开发被试拟合统计量，以此筛选异质性被试，进而应用其他假设模型进行分析或者直接剔除。

意见 3：模拟研究中基于反应与反应时数据符合加工速度和认知能力混合作用的特点生成数据，因此满足这一前提假设的 HM-MEM 肯定表现更好。建议基于其他假设生成数据，考察 HM-MEM 的普适性。

回应：感谢审稿专家的建议，我们已经新增模拟研究 2，以速度 MEM 为基准模型的展开模拟研究，旨在探究混合 MEM 和独立 MEM 对简约模型的包容性及其性能表现。具体见 4.2 节内容。

意见 4：文章提出的联合模型和其他已有的联合模型，如 Loeys 等人(2011)的模型的区别与联系，独特的优势是什么？

回应：感谢审稿专家的意见，我们已经在文中补充了相应的区别。

(1)Loeys 等人(2011)的模型是基于 van der Linden(2007)的联合建模思想构建的。相对独立 MEM，Loeys 等人的模型随机效应结构忽视了被试与处理的交互作用(或被试随机斜率)，并不能较好地刻画重复测量实验设计的结构特征。相对速度 MEM 和混合 MEM，Loeys 等人的模型假设心理实验反应与反应时是独立的，而速度 MEM 假设实验任务只受反应时影响，混合 MEM 则假设实验任务的反应与反应时是混合影响。在联系上，Loeys 等人(2011)的建模假设与独立 MEM 是相同的。同时，当实验任务的反应和反应时相互独立时，混合 MEM 模型简化为独立 MEM；当实验任务只受到反应时的影响时，混合 MEM 则简化为速

度 MEM。

(2)与其他已知联合模型相比，文中提出的联合模型是基于混合效应模型的基础上构建的。LMEM 和 GLMEM 在心理实验反应与反应时数据的应用受到研究者关注，并成为当前心理实验数据分析的新趋势，但是现有的研究常将 LMEM 和 GLMEM 单独应用于心理实验反应时与反应的数据分析，割裂了不同数据间的关系。文中通过综合分析反应与反应时的理论关系，基于 LMEM 和 GLMEM 构建了三个联合模型。三个联合模型，既充分利用了 LMEM 和 GLMEM 对心理实验数据的应用优势，又整合了被试的不同信息(数据)来源，能更全面准确地分析实验设计的特点和评估实验效果。

意见 5：文章提出的三种模型，混合效应模型的分层模型，混合效应模型的独立分层模型和加工速度影响的混合的线性混合效应模型，其中文命名无法很好地体现模型假设，建议改为混合 MEM、独立 MEM、速度 MEM。

回应：谢谢审稿专家的建议，已经在文中进行了相应的修改。

意见 6：文章在介绍模型时，例如“2.1 反应时模型 LMEM”，需首先说明实验设计，如，被试内实验设计，刺激在不同实验水平重复。

回应：谢谢审稿专家的建议，已在文中进行了修改。

接下来，将基于单因素两水平重复测量实验设计介绍包含所有随机效应结构的 LMEM 和 GLMEM。在这种实验设计中，自变量包含两个水平，被试接受所有水平处理，同时刺激项目与实验处理水平是交叉关系。比如在命名汉字启动效应中，启动字和目标字分别为语音相关、语义相关和无关，每个实验条件都使用了相同的启动字。基于此，GLMEM 和 LMEM 的理论随机效应结构包含了所有随机效应(Judd et al., 2017)。

意见 7：介绍 HM-MEM 具体模型的部分，需要更加详细，阐述公式每个部分的含义，以及为什么使用生存函数，累积分布函数等，让读者清晰了解该方法的逻辑。

回应：谢谢审稿专家的建议，已在文中进行了相应解释。

$f(T_{ijk})$为T_{ijk}的密度函数(如式 1)，$P(Y_{ijk})$为 GLMEM 的正确概率(式 4)，两者的乘积反映了被试在实验任务上的认知能力和反应时是充足的，也就是操作正确的反应与反应时信息。对于操作错误反应与反应时信息由两部分构成，其一是认知能力不足，通过 GLMEM 的错误率$Q(Y_{ijk}) = 1 - P(Y_{ijk})$描述，表明被试缺乏相应的认知能力；其二是认知能力充足而反应时不足，认知能力充足通过 GLMEM 的正确率$P(Y_{ijk})$描述，而反应时不足通过T_{ijk}的生存函数$S(T_{ijk}) = 1 - \Phi(T_{ijk})$表示，其中$\Phi(T_{ijk})$为$T_{ijk}$的累积分布函数，$S(T_{ijk})$表示在时间点$T_{ijk}$时，被试仍未正确操作的可能性或概率(Miller et al., 2018)，两者乘积构成认知能力充足而反应时不足的信息。由于无法明确区分被试是认知能力不足还是反应时不足导致操作错误，因此通过这两部分信息之和构成操作错误反应的信息（Lee et al., 2015）。最终构成混合 MEM，如式 7 所示。

意见 8：“3.1 实验过程描述”中应当明确说明自变量、因变量和实验设计。

回应：谢谢审稿专家的批评与指正，我们已经在文中进行了修改。

实验设计：本实验采用了单因素两水平重复测量实验设计。自变量为任务类型，分别为相容任务和不相容任务；因变量为被试在每个刺激项目上的反应与反应时。

意见 9：“3.2 随机效应结构探究”中，刺激是否包含随机斜率，是与实验设计有关，而不仅仅是模型选择解决的问题。因此，应当先根据实验设计选择合适备选模型，再根据拟合指标进行模型选择。

回应：谢谢审稿专家的意见。

当前对 MEM 的随机效应结构探讨主要局限于以下方面：①脱离实验设计，以模拟探究随机效应结构的选择与影响(Martínez-Huertas et al., 2022)；②缺乏实证检验，基于理论分析实验设计的随机效应结构(Judd et al., 2012, 2017)；③局限重复测量实验设计，探讨重复测量实验设计的随机效应结构特点(Brown, 2021; Lee, 2018)。综上，目前并没有研究对不同实验设计的随机效应结构进行系统的模拟与实证探讨。因此，文中基于遍历原则，从保守角度对实验数据的所有可能随机效应结构进行拟合比较，最终的随机效应结构符合理论假设。

意见 10：模拟研究部分应当明确列出评价指标及其标准。

回应：谢谢审稿专家的意见，文中已经列出了评价指标。

对于参数模拟结果采用平均相对偏差 Bias 进行评价，Bias 的值越接近 0 意味参数的估计值与模拟值越接近，结果越佳。

$$Bias(\hat{\xi}) = \frac{\sum_{r=1}^{R}(\hat{\xi} - \xi)}{R}$$

其中$\hat{\xi}$和ξ分别为估计值和模拟值，R 为重复次数。

意见 11：图 1 的结果建议换成偏差或者相对偏差。

回应：感谢审稿专家的建议，图 1 的结果已经转换为相对偏差 Bias，具体见正文。

……

审稿人 2 意见：

文章结构整体比较合理，也有一定的创新性。不过也有很多需要修改完善之处，具体如下：

意见 1：就实验研究中，联合建模的意义和必要性，以及应用情境的阐述需要进一步加强。

回应：谢谢审稿专家的建议，已经在文中进行了补充。

(1)联合建模的意义和必要性：相对反应与反应时数据独立分开建模分析，联合建模在模拟和实证数据上能提供更为合理的结果(Suh, 2010)，更为准确的参数估计(Bolsinova & Tijmstra, 2018; Loeys et al., 2011)和更佳的模型拟合效果(Man et al., 2019; Loeys et al., 2011)。同时，恰当的心理实验数据分析方法能有效的缓解心理实验研究的可重复危机(Yarkoni, 2020)。因此，探究更为科学的心理实验联合建模方法就显得特别重要。

(2)联合建模在心理实验的应用情景：文中基于心理实验数据的联合建模取得了一些有意义的发现，但是联合建模在心理实验的实际应用仍然需要注意一些问题。首先，对联合建模的反应时假设，文中的 LMEM 假设反应时服从正态分布，但是在一些心理实验研究中，有研究者采用 lognormal、weibull(Loeys et al., 2011)和 wald(Miller et al., 2018)等分布，甚至采用半参数的比例风险模型(Loeys et al., 2014)。为了选择恰当的反应时模型，不同反应时模型需要通过相同的参数估计方法，同时采用模型拟合指数对不同反应时分布假设进行比较以

确定最佳的模型，从而优化联合建模的效果。其次，文中仅针对常用因变量反应与反应时进行了联合建模分析，但是联合建模的框架并不受此局限，它可以整合更多的不同心理实验数据集，比如眼动(詹沛达，2022)、鼠标轨迹(Liang et al., 2023)以及脑电(Visalli et al., 2024)等数据，这些数据提供了被试不同实验操作的信息来源，有助于从多方面更深入和更全面地解释被试的心理认知过程。同时，对不同心理实验数据进行联合建模时，需要对不同数据间的理论关系进行深入分析，这是进行联合建模的基础和前提。最后，探明实验数据的随机效应结构是混合效应模型应用的基础，现有的混合效应模型的随机效应结构主要通过似然比检验方法进行确定(Barr et al., 2013; Mart ńez-Huertas et al., 2022; Matuschek et al., 2017)。似然比检验基于近似卡方分布的零假设检验方法进行推断，同时对卡方分布的近似依赖大样本特性。但是，零假设检验方法存在明显的局限性而常被诟病(Hoijtink et al., 2019)。相比传统的假设检验方法，贝叶斯因子(Bayes Factor)具有诸多优势。它不依赖于大样本假设，还能够揭示备择假设与虚无假设成立可能性的高低，因此心理学界出现了以贝叶斯因子分析取代传统假设检验的呼声（胡传鹏 等; 2018）。心理实验研究往往都是小样本设计，并不非常吻合似然比检验方法的假设。随着心理实验设计的复杂度增加(Park et al., 2020)，文中遍历所有随机效应结构的方法并不是一种高效的方法，因此有必要从贝叶斯因子角度去探究 LMEM 和 GLMEM 的随机效应结构的确定。

意见 2：文章的数据和代码，请上传到指定的科学数据银行。
回应：谢谢审稿专家的建议，我们已经上传到科学数据银行。

文中 IAT 实验数据与三个联合模型的 stan 代码可以通过网址 https://www.scidb.cn/anonymous/NmJVeklu 自行下载。

意见 3：2.3 节最后一段请重新表述，调整结构，将 HM-MEM-D 的条件和(10)放在一起，将 M-LMEM 的条件和(11)放在一起。另外概述三个模型之间的关系和各自的适用条件。
回应：谢谢审稿专家的建议。

(1)文中已经重新表述了 2.3 节最后一段。

(2)文中已经将 HM-MEM-D(独立 MEM)的条件和式(10)放在一起，将 M-LMEM(速度 MEM)的条件和式(11)放在一起。

(3)三个模型之间的关系和适用条件：在心理实验中，混合 MEM、独立 MEM 和速度 MEM 三个联合模型之间存在一定关系。混合 MEM 反映了实验任务中反应和反应时之间的混合影响；当实验任务的反应和反应时相互独立时，混合 MEM 模型简化为独立 MEM；当实验任务只受到反应时的影响时，混合 MEM 则简化为速度 MEM。因此，在心理实验的应用情景中，混合 MEM 适用于既有难度又有时间压力的实验任务；独立 MEM 适用于有难度但时间完全充裕的实验任务；而速度 MEM 则适用于非常简单但有时间压力的实验任务。

意见 4：表 1 的结果是采用贝叶斯估计吗？如果是，在贝叶斯分析中，AIC 和 BIC 的惩罚项有时并不容易得到；这不仅仅是计算估计参数的数量，除非所有先验都是完全非信息的，而从附录 1 中来看这里的情况并非如此。任何非平坦先验都会导致信息池化，从而导致有效参数个数(即贝叶斯分析中通常使用的惩罚项)小于估计参数数。Gelman 等人在"Bayesian Data Analysis"一书中对此进行了讨论。这就是为什么许多贝叶斯预测性能指标估算的是有效参数个数(如 DIC)，而不仅仅是估算参数个数的原因。因此，建议作者报告 DIC 和其他拟合指数。
回应：谢谢审稿专家的意见。表 1 中，LMEM 和 GLMEM 在文中采用 R 语言 lmerTest 软件包(Kuznetsova et al., 2017)自带的 REML(Restricted Maximum Likelihood method)估计方法进

行探究，并不是贝叶斯估计方法。我们已经进一步修改表述，使得内容更为清晰。

对于 IAT 实验数据的随机效应结构的探究，采用混合效应模型最常用的 R 语言 lmerTest 软件包(Kuznetsova et al., 2017)进行分析，其默认的参数估计方法为 REML(Restricted Maximum Likelihood method)。lmerTest 软件包能够提供模型拟合指数 AIC 和 BIC 来评价不同随机效应结构的模型拟合效果，模型拟合指数越小，模型拟合越好，随机效应结构越符合实验数据特点。

意见 5：如果生成数据的模型是简单模型，用复杂模型估计的结果如何？

回应：谢谢审稿专家的建议。我们已经新增模拟研究 2，以速度 MEM 为基准模型的展开模拟研究，旨在探究混合 MEM 和独立 MEM 对简约模型的包容性及其性能表现。具体见 4.2 节内容。

意见 6：语言表述有待进一步打磨，整体上感觉文章的可读性较差。例如，文章缩写比较多，读起来不顺畅；有些叙述比较重复，例如公式(2)和(3)的表述与文字部分重复等，”……还要乘以错误反应时的密度函数……”类似口语化的叙述需要检查修改。

回应：谢谢审稿专家的批评与指正。

(1)文中已经减少了模型的相应缩写，比如将 HM-MEM、HM-MEM-D 和 M-LMEM 的缩写修改为混合 MEM、独立 MEM 和速度 MEM。

(2)删除了公式与文字重叠的内容。

(3)对文中内容的表述已经进一步多次调整和修改。

第二轮

审稿人 1 意见：

感谢作者所做的回复和修改，在第一轮修改中，作者已经较好的回应了我关注的几个问题，文章质量有一定提高。这一稿中仍有一些地方值得商榷，现列举如下：

意见 1：对这个研究我最大的关注点是，作者应当加强对联合建模优势的充分论证。尤其在实证研究的结果系数解释上，没有体现联合建模的特点，相比传统方法有什么好处？得到的结果能够提供什么信息？

回应：感谢审稿专家的建议。

为了更好地突出联合建模的优势，我们在实证研究和模拟研究中分别增加了反应时和反应数据分开建模的比较，即分开 MEM。分开 MEM 是将线性混合效应模型(LMEM)和广义线性混合效应模型(GLMEM)分别应用于实验反应时与反应数据，而不考虑反应与反应时关系。研究结果发现：

(1)实证研究：在模型拟合指数上，相比分开 MEM，独立 MEM 的 WAIC 和 LOO 略小，而速度 MEM 和混合 MEM 明显更优。在独立 MEM 和分开 MEM 的各参数估计上，两者的参数估计结果略有差异；但是在标准误上，独立 MEM 的各参数的标准误普遍比分开 MEM 的标准误更小，说明联合建模有利于提高参数估计的稳定性。

(2)模拟研究：在不同基准模型的模拟研究中，分开 MEM 比独立 MEM 的各参数具有更大的参数估计偏差，精度更低，且具有相对更高的第 I 类错误率。在各联合模型中，混合 MEM 比其他联合模型能更好地识别不同模拟条件的参数，并且具有较佳的第 I 类错误率和统计检验力。

综合上述内容，心理实验数据的联合建模比分开建模更有优势。

意见 2：摘要部分，作者花大量篇幅介绍了线性混合效应模型在实验数据处理中的优势，这并不是本文关注的重点。在应用该模型的基础上，本研究提出的是联合建模的方法，这才是应当突出强调的地方。

回应：谢谢审稿专家的批评指正。我们已经重新表述了摘要部分。

混合效应模型(Mixed-Effects Models, MEM)作为当前心理实验数据分析的新趋势，能够将被试和刺激项目同时作为随机变量，有效地分析实验效应和相关的被试(或刺激项目)差异，从而避免了传统方差分析的随机效应固定化问题。通常，心理实验的反应与反应时数据会分开进行描述与分析，这不利于充分利用与整合被试的不同数据信息。现有的心理测量与认知过程模型已经尝试对反应与反应时数据进行联合分析，为心理实验数据的联合建模提供了启示。基于此，文中构建了混合 MEM、独立 MEM 和速度 MEM 三个联合模型，并与反应和反应时数据的分开建模(即分开 MEM)进行比较。在 IAT 实验数据分析中，分开 MEM 在数据拟合与参数估计上均不如独立 MEM，而混合 MEM 的模型拟合指数优于独立 MEM 和速度 MEM。在模拟研究中，分别以混合 MEM 和速度 MEM 为基准模型展开不同模拟比较。模拟结果显示，分开 MEM 参数估计的相对偏差普遍大于独立 MEM，且具有较高的第 I 类错误率；而混合 MEM 比其他联合模型能更好地识别不同模拟情景的参数，并且具有较佳的第 I 类错误率和统计检验力。因此，在心理实验中，联合建模方法比分开建模具有更大优势。此外，实验任务的反应和反应时更可能存在复杂的混合影响关系。

意见 3：引言部分指出“在心理实验中，实验任务是只受到反应时影响，还是反应与反应时独立影响，亦或者反应与反应时的混合影响，目前尚无研究基于 LMEM 和 GLMEM 在心理实验中探讨这些问题，也没有相应的分析模型与方法”这是作者提出几种不同模型所对应的假设，也不是本研究解决的问题。所以在这里陈述的目的是什么？或者说，通过对实际数据拟合几种模型，并比较拟合指标，能够大致判断其实验任务的机制？这也可以算作该模型的一种应用方向？

回应：谢谢审稿专家的意见，我们已经重新表述此句内容。

(1)内容重新表述为“在心理实验中，刺激项目往往会设置呈现时间，同时任务相对简单。因此，构建符合心理实验特点的联合模型就显得尤为重要。但是，目前尚无研究系统地比较 LMEM 和 GLMEM 在心理实验中反应与反应时数据联合分析的表现。”

(2)这里这样阐述的目的是：目前尚未有研究基于实验数据系统性地比较不同联合模型的表现，也无研究探讨实验任务的机制更倾向哪一种理论假设。在本研究中，通过拟合 IAT 实验数据，发现了混合 MEM 能最佳拟合实验数据，这表明 IAT 实验数据更倾向混合 MEM 的假设，进而可以初步推断 IAT 实验任务的机制可能涉及反应与反应时的混合影响。这相当于对此问题进行了回应和探索，也可以视为混合 MEM 在实际应用中的一种尝试。

(3)不同模型通常基于不同理论假设进行构建，为了验证数据更贴合哪种理论假设，可以通过比较各模型的拟合指标来评估这些理论假设在实际数据中的相对支持程度，从而为相关理论假设提供实证支持。这种通过模型拟合指数比较的方法，常被用于评估和比较不同理论假设模型。具体可以参考以下文献：

Mart ńez-Huertas, J. Á., Olmos, R., & Ferrer, E. (2022). Model selection and model averaging for mixed-effects models with crossed random effects for subjects and items. *Multivariate Behavioral Research, 57*(4), 603–619.

詹沛达. (2022). 引入眼动注视点的联合-交叉负载多模态认知诊断建模. *心理学报, 54*(11), 1416−1423.

意见 4：在实证研究中拟合混合模型之前，首先通过对反应模型和反应时模型分别建模，选择了合适的随机效应模型。这是否是实际中应用该方法的必要步骤？因为模拟研究中也没有

体现这部分内容，直接假设随机效应模型的定义是正确的。

回应：谢谢审稿专家的意见。我们已经在模拟研究部分增加了未探究随机效应模型定义的原因。

(1)增加内容：此外，由于模拟研究旨在探究各模型的整体性能，故假设各模型的随机效应结构是已知的。

(2)对混合效应模型的实际应用是否需要选择合适的随机效应结构，可以分为两种情况：①根据已有文献结果直接确定实验设计的随机效应结构(DeBruine & Barr, 2021; Judd et al., 2012; Loeys et al., 2011)。②基于数据驱动进行探索性分析，选择合适的随机效应结构 (Barr et al., 2008; Baayen et al., 2013)。本文采取了较为保守的策略，既考虑了数据驱动的结果，也参考了已有研究的理论模型，并发现两种方法得到的结果是一致的，这提高了研究结论的可信度。

(3)实证研究与模拟研究步骤可以存在一定差异：实证研究旨在探讨心理实验数据更适用哪种联合建模假设，因此需要先进行建模，再选择合适的随机效应模型。相比之下，模拟研究的目的是系统比较各联合模型的性能，不需要像实证研究那样进行建模和模型选择，而更关注各联合模型的参数估计偏差、第 I 类错误率和统计检验力等性能指标。虽然两种研究方法的具体步骤不尽相同，但它们相辅相成，为彼此提供了重要证据和支撑。因此，模拟研究中并未涉及模型构建和随机效应模型的选择，而是重点关注模型性能评估。

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

DeBruine, L. M., & Barr, D. J. (2021). Understanding mixed-effects models through data simulation. *Advances in Methods and Practices in Psychological Science, 4*(1), 1–15.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: a new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103*(1), 54–69.

Loeys, T., Rosseel, Y., & Baten, K. (2011). A joint modeling approach for reaction time and accuracy in psycholinguistic experiments. *Psychometrika, 76*(3), 487−503.

意见 5：模拟研究部分的表述不够规范。应先说生成数据的模型是什么，再说模拟条件，分析方法，评价指标及其标准。目前评价指标只写了 bias，没有后文的第一类错误率，power，覆盖率等。建议作者使用相对偏差 rbias，可以用-0.1 到 0.1 作为标准。

回应：谢谢审稿专家的批评与指正。我们已经重新表述了模拟研究部分。

(1)在模拟研究中，我们介绍模拟条件设置时增加了生成数据的模型，并且将模拟研究内容分为三小节：模拟条件设置、实验数据生成和评价指标。在评价指标中，我们增加了第 I 类错误率和统计检验力 power 的描述，以对应模拟研究的结果内容。

①模拟条件设置的第一句修改为：在模拟研究 1 中，参考实证研究结果和 Loeys 等人(2011)的研究设置，以混合 MEM 为基准模型进行模拟研究。

②实验数据生成的第一句修改为：为了使生成的反应与反应时数据符合混合 MEM 的特点，数据生成过程如下。

③评价指标新增内容修改为：对于各模型的处理效应识别性能，分别采用第 I 类错误率与统计检验力进行评估(温忠麟 等，2019; Judd et al., 2017)。第 I 类错误率指真值为 0 时，估计值显著不等于 0 的概率，一般认为第 I 类错误率越接近真值 0.05 越好，且 0.025-0.075 之间的范围通常被认为是可接受的。统计检验力则指真值不为 0 时，估计值显著不等于 0

的概率。统计检验力越趋近于 1，模型性能越好。

(2)文中采用相对偏差(rbias)来评价参数估计效果，更符合评价指标的内涵，已经修改为 rbias。考虑到心理实验特点，rbias 采用-0.1 到 0.1 的标准可能过于严格，原因如下：

①模型参数估计的 rbias 容易受到被试与刺激项目量的影响。心理实验通常样本量较小(Loeys et al., 2011)，因此 rbias 的范围应该更宽。在心理与教育测量模型中，常基于 500 和 1000 等被试量模拟生成数据(Ranger & Kuhn, 2012)。相比之下，心理实验通常只有 30 到 60 名被试。

②在心理与教育测量中，项目反应理论(IRT)模型的参数范围通常在-3 到 3 之间，而文中的混合效应模型的随机效应与固定效应的范围远远超出这个区间。例如，被试随机截距服从均值为 0，标准差为 100 的正态分布 $N(0,100)$，其 95%的置信区间为[-196,196]，远远大于-3 到 3 的范围。在小样本条件下，这种较大的参数范围会使得参数估计的偏差明显增大。

因此，文中以 0 为中心，采用相对优劣的标准来评价各模型的参数估计效果可能更为恰当。

Loeys, T., Rosseel, Y., & Baten, K. (2011). A joint modeling approach for reaction time and accuracy in psycholinguistic experiments. *Psychometrika*, *76*(3), 487−503.

Ranger, J., & Kuhn, J.-T. (2012). A flexible latent trait model for response times in tests. *Psychometrika*, *77*(1), 31–47.

意见 6：结果部分为什么没有协方差估计的准确性？

回应：谢谢审稿专家的意见。

在模拟研究中，我们呈现了各随机效应的变异以及固定效应的估计准确性。之所以没有呈现各随机效应间相关系数的准确性，主要基于以下考虑：

(1) 模型间的可比性。文中的各联合模型和分开 MEM 主要基于 LMEM 和 GLMEM 构建，因此文中聚焦 LMEM 和 GLMEM 的随机效应与固定效应两类核心参数的估计准确性进行比较。但是，不同模型包含的相关系数类型与数量差异很大(如表 4 所示)，不具有跨模型的可比性。

(2) 相较于分开建模，联合建模同时考虑了心理实验反应与反应时数据的关系，以期提高模型的核心参数估计准确性。尽管联合模型提供额外信息，如果无法有效改善模型的核心参数估计，联合建模的价值也是非常有限的。在已有的心理测量联合建模的模拟研究中，也通常不报告参数间的相关系数的估计准确性，文中也参考已有联合建模的做法(郭小军 等, 2024; Bolsinova & Tijmstra, 2018; Man et al., 2019)。因此本文的重点是比较不同模型在核心参数估计方面的差异。

总之，文中通过重点比较混合效应模型的随机效应与固定效应参数，以了解各模型的性能与差异。

郭小军, 柏小云, 罗照盛. (2024). 作答时间与反应依赖关系建模：基于双因子模型视角. *心理学报*, *56*(3), 352–362.

Bolsinova, M., & Tijmstra, J. (2018). Improving precision of ability estimation: Getting more from response times. *British Journal of Mathematical and Statistical Psychology*, *71*(1), 13–38.

Man, K., Harring, J. R., Jiao, H., & Zhan, P. (2019). Joint modeling of compensatory multidimensional item reaction and response times. *Applied Psychological Measurement*, *43*(8), 639–654.

意见 7：作者增加了模拟研究 2，以速度 MEM 产生数据？为什么不增加以独立 MEM 产生数据的情况？

回应：谢谢审稿专家的建议，我们已经在模拟部分进一步进行了解释。同时，文中以速度

MEM 产生数据，而不是以独立 MEM 产生数据的原因在于：

(1) 从模型复杂度的角度来看，混合 MEM 是最复杂的，独立 MEM 次之，速度 MEM 最简单。因此，研究 1 选择了复杂度最高的混合 MEM 作为基准模型，而研究 2 则选择复杂度最低的速度 MEM，这样可以最大化地模拟结果的差异，更好地反映不同实验情景下的模型性能。

(2) 从不同的建模理论假设来看，速度 MEM 的构建是基于心理实验的认知过程模型假设，而混合 MEM 和独立 MEM 则是根据心理测量建模假设构建。因此，模拟研究 1 以混合 MEM 为基准模型，而模拟研究 2 以速度 MEM 为基准模型分别代表了这两种不同的建模理论依据。

综合上述原因，为了使模拟研究结果更具有代表性，分别以混合 MEM 和速度 MEM 为基准模型生成心理实验数据，以有效地评估各模型的参数估计精度以及处理效应的识别性能。

意见 8：建议作者在讨论部分增加应用建议，例如先选择合适的随机效应，再根据实验任务类型确定合适的假设及其相应的模型，最后拟合模型解释结果。

回应：谢谢审稿专家的建议，我们已经在文中增加了应用建议。

基于上述研究发现，在将混合效应模型的联合建模方法应用于心理实验数据分析时，需要注意以下几个方面：(1)需要先确定数据的随机效应结构。随机效应结构可以基于已有文献从理论出发进行确定，也可以结合不同的指标从数据驱动角度确定最佳的随机效应结构。(2)在合理的理论假设基础上构建联合模型。根据实验任务类型确定不同数据间的理论假设，并基于此构建相应的联合模型。(3)基于模型拟合统计量，拟合模型并解释数据结果。通过比较不同模型拟合指数，选择最优的拟合模型，然后基于该模型解释实验结果。

……

审稿人 2 意见：

作者对上次的修改意见进行了全面细致的回复，较好地回应了稿件存在的问题，也做了较大幅度的修改和完善，整体上文章质量较上次明显提升。

建议进一步检查修改语言表述的准确性和简洁性。有些句子依然存在可读性差的问题；建议明确列出文章得出的主要结论。英文摘要需要进一步修改完善。

回应：谢谢审稿专家的建议。

(1)我们已经进行了交叉式通读全文，并对文中语言的表述进行了相应修改。

(2)我们单独设置了 5.2 研究结论，以突出文章的主要结论。

文中根据心理实验反应与反应时数据的不同关系假设提出了不同联合模型，基于实证与模拟研究结果，得出如下结论：

(1)相较于将心理实验数据分开建模，联合模型能更为有效地整合不同数据信息，从而提高参数估计精度和处理效应的识别性能。在联合建模假设中，实验任务更倾向于受到反应与反应时的混合影响。

(2)实证研究结果表明，混合 MEM 的模型拟合效果优于独立 MEM 和速度 MEM，而分开 MEM 拟合效果最差。同时，独立 MEM 的参数估计稳定性优于分开 MEM。

(3)模拟研究结果表明，在混合 MEM 基准模型上，独立 MEM、速度 MEM 和分开 MEM 的参数估计存在明显偏差，且具有较高的第 I 类错误率和统计检验力。相比之下，混合 MEM 在不同基准模型上均具有较好的参数估计精度、较低的第 I 类错误率和较高的统计检验力。此外，分开 MEM 在不同基准模型上的估计偏差上要略大于独立 MEM，但两者的第 I 类错误率和统计检验力接近。

(3)对英文摘要，我们已经邀请英语专业人士进行修改与润色。

第三轮

审稿人 1 意见：

作者较好的回复了我的所有问题，并在文中有了充分体现。建议发表。

回应：谢谢审稿专家。

编委意见：

建议作者对全文进行精简润色，删除可有可无的句子。对摘要已给出了修改参考，删除了约三分之一字数。

回应：感谢编委专家的意见。参考编委专家的修改模板，我们已经对文章正文进行多轮精简润色，共精简 1000 多字。

主编意见：

本论文借助内隐倾向测验数据，对该数据中的反应和反应时进行联合建模，并比较了该方法相较于分开建模的优势。本论文的研究选题具有创新性，研究框架清晰，研究方法选用合理。