

《心理科学进展》 审稿意见与作者回应

题目：密集追踪研究中测验信度的估计：多层结构和动态特性的视角

作者：罗晓慧，刘红云

第一轮

审稿人 1 意见：

该文详细介绍了基于多层验证性因子分析（MCFA）、动态因子分析（DFA）和动态结构方程模型（DSEM）的信度估计方法，在实证数据中进行演示与比较，并对其特点和适用情境进行讨论。该文具有一定的理论和现实意义。但该文还存在如下瑕疵，与作者商榷。

回应：感谢您对本文的认可，也非常感谢您提出的宝贵意见与建议！以下是针对各条意见与建议的回复，相应的修改已在文章中用黄色底纹标出。

意见 1：该文是以单维测量结构为例，但实际应用中，多维测量结构的量表也很多，那么，密集追踪研究中，多维测量结构如何进行测量信度的估计呢？还请作者给予说明。或者将题目具体为“密集追踪研究中单维测验信度的估计”。

回应：非常感谢您指出这一点。本文提到的基于多层验证性因子分析、动态因子分析和动态结构方程模型的三种信度估计方法都适用于多维测量结构的情况。原文中整个测验的个体间信度/个体特定信度的公式(即公式 5、7、11、19、22)可用于计算多维测量结构中某个维度的个体间信度/个体特定信度。我们在修改文稿的第 2、3 和 4 部分进行了相应的补充说明，并将“整个测验”的表述改为“各个维度”，以便于读者理解并运用于多维测量结构的情况。此外，为了便于应用研究者运用本文提到的三种方法估计密集追踪测验的信度，我们在修改文稿中为三种方法应用于多维构念的情况提供了相关参考文献：基于多层验证性因子分析的多维构念的信度估计可参见 Di Sarno 等人(2020)、Neubauer 等人(2022)和 Wright 等人(2017)的研究；基于动态因子分析的多维构念的信度估计可参见 Fuller-Tyszkiewicz 等人(2017)的研究；基于动态结构方程模型的多维构念的信度估计可参见 Xiao 等人(2023)的研究。

意见 2：在“6.1 三种方法的比较分析”中，能否请作者将“本文对不同方法的适用情境提出建议”整理出一个流程图，便于应用呢？

回应：感谢您的宝贵建议。我们已在修改稿中增加密集追踪研究中测验信度估计方法的选择策略流程图(见修改稿的图 4)，以便应用研究者选择合适的信度估计方法。

意见 3：在“6.2 实践应用中的建议”中，建议作者增加一个内容，即应用研究者在密集追踪研究中，应该如何报告信度估计。因为该文的示例中，以动态结构方程模型的信度估计为例，就有每个题目的个体间信度，每个题目的个体内信度，整个量表的个体间信度，整个量表的个体内信度。在密集追踪研究中，哪些信度是必须报告的，怎么报告才符合规范，哪些是选

择报告的？请给出建议。

回应：非常感谢您指出这一点，我们完全同意信度结果的报告对于应用研究者的重要性。我们已在修改稿中增加“6.2.3 信度结果的报告”，具体说明密集追踪研究中，采用各种信度估计方法的研究者必须和可选择性报告的信度，以及如何规范地报告。

意见 4：请作者模仿将方程 4 的信度解释为“个体内信度为该题由潜在状态因子解释的变异与该题状态成分的变异之比”这样，能否给出方程 10、11、18、19、21、22 所表示的信度的含义，便于读者理解。

回应：感谢您的建议。在修改稿中，我们根据您的建议，模仿公式 4 的信度解释表述，对公式 10、11、18、19、21、22 所表示的信度的含义进行了解释，详见修改稿中对相应公式的解释。

意见 5：在表 2，作者给出了个体特定信度的均值和中位数，但是作者在示例部分明确说“后验分布的中位数为个体内信度的点估计”，比较表 1 的个体内信度的点估计和表 2 的个体特定信度的“均值”几乎一样，和“中位数”差别较大，这是为什么？请作者核查或者解释原因。

回应：感谢您的提问。表 1 中个体内信度的点估计与表 2 中个体特定信度的均值或中位数的含义并不相同，数值大小不可比较。表 1 的点估计结果为**个体内信度的后验分布**(本文实例中由 200 个个体内信度的可信值组成，详见下述)**的中位数**，表 2 的均值和中位数结果基于**个体特定信度的分布**(本文实例中由 252 个个体特定信度组成，252 为被试量)得到。

(1) 表 1 中某题目或测验的个体内信度的点估计及 95% 贝叶斯可信区间的计算过程

第一步，通过 Mplus 中 SAVEDATA 部分的语句保存得到模型中随机参数的贝叶斯后验分布(具体指相关参数后验分布的 200 次迭代结果，即 200 个可信值(plausible values))。

(a) 某题目或测验的**个体内信度的后验分布**的计算示意图

ID	P ₁			...			P _k			Rel _i ^(w)		
	1	...	200	1	...	200	1	...	200
1												
2												
...												
251												
252												
	-	-	-	-	-	-	-	-	-	均值	...	均值

个体内信度的后验分布

(b) 某题目或测验的**个体特定信度的分布**的计算示意图

ID	P ₁			...			P _k			Rel _i ^(w)		
	1	...	200	1	...	200	1	...	200
1												
2												
...												
251												
252												

ID=1 个体的个体特定信度的后验分布

→

ID=1 个体的个体特定信度点估计(中位数)

...

...

ID=252 个体的个体特定信度的后验分布

→

ID=252 个体的个体特定信度点估计(中位数)

个体特定信度的分布

图 1 某题目或测验的个体内信度的后验分布和个体特定信度的分布的计算示意图

注：ID=1~252 为本文实例中的 252 个个体，P₁~P_k 为 Mplus 中存出的个体特定信度计算所需的参数，Rel_i^(w) 为个体 i 整个测验的个体特定信度，以此为例说明个体特定信度和个体内信度的相关计算过程。

第二步,对于每次迭代(1~200)的相关参数,根据相应公式(如公式 22 和 23)计算某题目或测验的个体特定信度(本文实例中,252 名个体在每个题目和整个测验都各有 200 个个体特定信度,见图 1b)。这一步的计算具体见 OSF 开放资源平台中“3_动态结构方程模型”文件夹的 R 代码的 28~80 行。

第三步,对题目和测验个体特定信度的每次迭代结果的所有个体求均值,得到 200 个个体内信度(即 200 个个体内信度的可信值),组成该题目或测验的**个体内信度的后验分布**(见图 1a),后验分布的中位数为该题目或测验的个体内信度的点估计,2.5%和 97.5%分位数分别为该题目或测验的个体内信度的 95%贝叶斯可信区间的上、下限。

(2) 表 2 中某题目或测验的个体特定信度的分布的计算过程

此时计算的前两步与上述相同,第三步中,不对所有个体求均值,而是基于每个个体得到**该个体的个体特定信度的后验分布**(见图 1b),每个后验分布由 200 个个体特定信度的可信值组成。对每个个体特定信度的后验分布,取其中位数为该个体特定信度的点估计。最后,综合 252 个个体的个体特定信度的点估计得到表 2 中的分布及其描述性统计指标(如中位数、均值等)。

意见 6: 表 2 所示的动态结构方程模型的个体特定信度似乎就是表 1 的个体内信度,请统一名称或给予明确的说明,否则名称不同,读者会误认为是两种不同的信度。

回应: 非常感谢审稿专家指出这一点,这让我们意识到原文对“个体内信度”和“个体特定信度”的区别及其计算过程的说明仍不够清晰。原文在“4 整合多层结构和动态特性的信度估计方法”介绍各个题目和整个测验的个体特定信度的计算方法后提到,“此外,通过整合所有个体在各个题目和整个测验的个体特定信度可以分别得到各个题目和整个测验的个体内信度,即描述个体内水平信度的整体指标(具体计算方法见本文的实证示例)。”随后,在“5.1 实证数据与分析方法”具体介绍了个体特定信度的计算过程,在“5.2 结果与讨论”的第二句简要介绍了个体内信度的计算过程。

在此基础上,我们在修改文稿的“5.1 实证数据与分析方法”,进一步明确“个体内信度”和“个体特定信度”这两个信度指标存在区别,并进一步说明了两个信度指标的计算过程。希望修改后的文稿有助于审稿专家和读者更好地理解两者的含义和估计方式。

意见 7: “值得说明的是,在基于动态因子分析和基于动态结构方程模型计算信度时,某些个体的某些迭代结果中潜在(状态)因子方差的估计值可能为负。为了排除这些有问题的迭代结果对信度估计的影响,我们将相应迭代中的个体特定信度替换为缺失值,即不纳入最终对信度的计算”,这种做法的依据是什么?请作者给出参考文献。便于读者以后在遇到类似问题时,采用文中的这种做法。

回应: 感谢您的宝贵建议。我们参考了 Xiao 等人(2023)的做法(见该文中的注释 2),我们也在修改文稿中补充了这一参考信息,以便其他研究者参考处理此类问题。

意见 8: “在随机效应部分,个体内水平的因子载荷、随机测量误差的自然对数、自回归效应和动态误差的自然对数都被分解为固定部分和随机部分”,请给出相应的 4 个方程,这样用文字配合相应的方程,读者更容易理解。

回应：感谢您的宝贵建议。我们在修改文稿中给出了相应的4个方程(详见方程18~21)。

意见 9：在方程15、16前注明这是个体内或者层1方程，方程17前注明这是个体间或者层2方程。这是个惯例，在多水平条件下，注明方程是属于哪一水平或者哪一层。

回应：感谢您的宝贵建议。我们已在修改文稿中标注方程15和16为个体内水平的方程，方程17为个体间水平的方程。

意见 10：在“随机测量误差方差的自然对数、和动态误差方差的自然对数都被分解为固定部分和随机部分”中，请作者解释为什么随机测量误差方差和动态误差方差要取自然对数，目的是什么？便于读者理解。

回应：感谢您的宝贵建议。对随机测量误差方差和动态误差方差取自然对数，主要是为了确保估计得到的每个个体的随机测量误差方差和动态误差方差均为正值。此外，对这些误差方差取自然对数还有助于基于多元正态分布，考察这些误差方差的随机对数与其它参数(如个体均值或自回归效应)的随机效应的相关关系(Hamaker et al., 2018)。我们已将上述说明补充在修改文稿中。

意见 11：图1~图3是作者的原创，还是有出处？如果有出处，请注明出处。

回应：感谢您的提问。图1、图2和图3分别基于原文中给出的多层验证性因子分析模型、动态因子模型和动态结构方程模型的定义方程画出。图1和图3的作图方式分别参考了Geldhof等人(2014)补充材料中的图1和Xiao等人(2023)正文中的图1，这些参考信息已在修改文稿中图1和图3的注释中说明。

意见 12：因为是密集追踪数据可被看成是时间点嵌套于个体的两层嵌套结构，因此图1一般是个体内在下(层1在下面)，个体间在上(层2在上面)，更便于理解。

回应：感谢您的宝贵建议。我们已按照您的建议对图1进行了修改，以便于读者理解。

意见 13：在“整合多层结构和动态特性的信度估计方法”部分，请明确指出基于DSEM的信度估计方法是源于哪篇文献？便于读者进一步深入阅读。

回应：感谢您的宝贵建议。基于DSEM的信度估计方法基于Xiao等人(2023)的文章，这篇文献已补充在修改文稿的相应位置。

意见 14：“首先，动态因子分析混淆了观测分数的状态成分和特质成分，这可能会导致个体特定信度的估计结果有偏差。”请解释观测分数的状态成分和特质成分是指什么？便于读者理解。

回应：感谢您的宝贵建议。个体观测分数的特质成分是指个体的某一构念在多次观测中的一般水平，状态成分是指个体的某一构念的某次观测相对其一般水平的偏离(Ringwald et al., 2022)。我们在修改文稿的相应位置补充了对特质成分和状态成分的含义的解释。

意见 15：作者说密集追踪研究的信度估计主要是用MCFA、DFA、DSEM法求信度。请

分别给出应用 MCFA、DFA、DSEM 的密集追踪研究的信度估计的实证文章的例子，便于读者进一步深入阅读。

回应：感谢您的宝贵建议。我们已在修改文稿的第 2 部分的第一段和第二段、第 3 部分的第二段和第 4 部分的第二段分别补充了用 MCFA、DFA 和 DSEM 估计信度的实证文献的例子，以便于读者进一步阅读与参考。

.....
审稿人 2 意见：

近年来密集追踪研究在心理学等领域受到了越来越多的关注，如何更好地估计密集追踪情境下的信度大小是研究者一直关心的重要问题。本文从密集追踪数据的多层结构和动态性两大特点出发，介绍了基于多层验证性因子分析、基于动态因子分析和基于动态结构方程模型的三种信度估计方法，并在实证数据中对这三种信度估计方法进行演示与比较，为应用研究者进行密集追踪研究下的信度估计提供了新的思路，对于国内心理学研究者了解各种方法的特点和适用情况有很好意义，但文中还有一些问题需要改进。

回应：感谢您对本文的肯定与认可，也非常感谢您提出的宝贵意见与建议！以下是针对各条意见与建议的回复，相应的修改已在文章中用黄色底纹标出。

意见 1：在第二部分“聚焦多层结构的信度估计方法”中，作者提到“这一方法假设各个题目的因子载荷和残差方差对所有个体都相等，进而得到所有个体的测量信度都相等”，在多层验证性因子分析模型框架下，如何定义个体的测量信度？另外，作者也提到“基于多层验证性因子分析的信度估计方法没有考虑密集追踪数据中连续观测点之间的时间依赖性”，此表述不太准确，因为多层模型本身就能考虑同一个体内多次测量之间的相关，只是没有考虑时间的先后顺序而已。

回应：感谢您的宝贵意见。对于您提到的第一个问题，原文的表述确实不太恰当，容易引起误解或疑惑。我们在修改文稿中将这句话调整为“这一方法假设各个题目的因子载荷和残差方差对所有个体都相等，故只能得到对个体内水平信度的一个整体评估”。对于您提到的第二个问题，我们非常同意原文的表述不够准确，我们在修改文稿中将原文的“时间依赖性”调整为“时序关系”。

意见 2：在第三部分“聚焦动态特性的信度估计方法”中，作者提到公式(9)“描述了当前时间点的潜在因子水平对下一个时间点的潜在因子水平的影响”，应该是前一个时间点对当前时间点的影

回应：感谢您的宝贵意见。表述为“前一个时间点的潜在因子水平对当前时间点的潜在因子水平的影响”确实更符合公式(9)的内容，我们已在文稿相应位置进行了修改。

意见 3：在第四部分“整合多层结构和动态特性的信度估计方法”中，定义个体间水平的信度时，下标是否有误？

回应：非常感谢您的细心审阅。非常抱歉原文定义个体间水平的信度的下标时有疏漏，我们已将其修改为：“题目 j 的个体间信度($Rel_j^{(B)}$)和各个维度的个体间信度($Rel^{(B)}$)”，详见修改

后的文稿。

意见 4: 实证研究部分，基于动态因子分析模型和动态结构方程模型的信度估计都使用了贝叶斯估计方法，迭代次数 10000 次时，模型是否已经收敛？另外，贝叶斯方法中要设定未知参数的先验分布，理论上能避免方差估计为负的情况，但作者强调说“在基于动态因子分析和基于动态结构方程模型计算信度时，某些个体的某些迭代结果中潜在(状态)因子方差的估计值可能为负”，这应该如何解释？

回应: 感谢您的提问。对于您提到的第一个问题，当迭代次数为 10000 次时，根据 Hamaker 等人(2018)的建议，我们通过 PSR(potential scale reduction; Asparouhov & Muthén, 2010)和各参数的轨迹图(trace plot)判断模型此时已收敛，我们已在修改文稿中补充说明了这一点。

对于您提到的第二个问题，确实可以通过设置未知参数的先验避免方差为负的情况，但有研究者指出，这样也会带来一个问题，即高估接近 0 的随机变异(Browne & Draper, 2006; Gelman, 2006)。由于动态结构方程模型有较多的随机参数，研究者常常关心各参数是否存在随机效应(即随机效应的大小是否与 0 有显著的差异)。为了更准确地估计参数的随机效应及其是否显著地大于 0，我们遵循 Mplus 官方的建议，对方差类参数采用默认的无信息先验分布(即 Inversed Gamma (-1, 0)，详见 Asparouhov & Muthén; 2021)。此外，对于贝叶斯后验分布中有问题的迭代结果的处理，我们参考了 Xiao 等人(2023)的做法，这一信息也已在修改文稿中进行补充说明。

意见 5: 实证研究部分，作者提到“在每次迭代中对所有个体的某个题目或整个测验的个体特定信度求均值，得到该题目或测验的个体内信度的后验分布”，这句话应该如何理解？

回应: 感谢您的提问。原文的表述可能不够清晰，某个题目或整个测验的个体内信度的后验分布的具体计算过程如下。

第一步，通过 Mplus 中 SAVEDATA 部分的语句保存得到模型中随机参数的贝叶斯后验分布(具体指相关参数后验分布的 200 次迭代结果，即 200 个可信值(plausible values))。

第二步，对于每次迭代(1~200)的相关参数，根据相应公式(如公式 22 和 23)计算某题目或测验的个体特定信度(本文实例中，252 名个体在每个题目和整个测验都各有 200 个个体特定信度)。这一步计算具体见 OSF 开放资源平台中“3_动态结构方程模型”文件夹的 R 代码的 28~80 行。

ID	P ₁			...			P _k			Rel _i ^(w)		
	1	...	200	1	...	200	1	...	200
1												
2												
...												
251												
252												
	-	-	-	-	-	-	-	-	-	均值	...	均值

个体内信度的后验分布

图 2 某题目或测验的个体内信度的后验分布的计算过程示意图

注: ID=1~252 为本文实例中的 252 个个体, P₁~P_k 为 Mplus 中存出的个体特定信度计算所需的参数, Rel_i^(w) 为个体 i 整个测验的个体特定信度, 以此为例说明个体内信度的相关计算过程。

第三步,对题目和测验个体特定信度的每次迭代结果的所有个体求均值,得到200个个体内信度(即200个个体内信度的可信值),组成该题目或测验的个体内信度的后验分布(见图2),后验分布的中位数为该题目或测验的个体内信度的点估计,2.5%和97.5%分位数分别为个体内信度的95%贝叶斯可信区间的上、下限。

我们在修改文稿的“5.1 实证数据与分析方法”,对个体内信度的计算过程补充了更详细的说明,希望有助于审稿专家和读者的理解。

意见 6: 实证研究部分,“在基于动态因子分析的信度估计过程中,有145人的动态因子模型无法拟合(因为该个体估计的方差协方差矩阵不正定等),故信度估计结果仅基于模型拟合的107人的数据”,即三种方法的比较并不是基于同样的样本,这样的比较是否公平?

回应: 非常感谢您指出这一点。我们非常同意基于动态因子分析与基于其它方法估计信度时所基于的样本不同,估计得到的信度结果可能不宜比较。我们已在修改文稿的相应位置补充强调这一点,提醒读者谨慎地解读相关结果。

意见 7: 讨论部分,表3中提到三种信度估计方法的耗时,建议作者补充电脑配置参数,让读者可以更好地对此进行评估。

回应: 感谢您的建议。电脑配置参数(包括CPU型号和内存参数)已补充在修改文稿的表3的注释中。

参考文献

- Asparouhov, T., & Muthén, B. (2010). *Bayesian analysis using Mplus: Technical implementation*.
- Asparouhov, T., & Muthén, B. (2021). *Bayesian analysis of latent variable models using Mplus*.
- Browne, W. J., & Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis, 1*, 473–514.
- Di Sarno, M., Zimmermann, J., Madeddu, F., Casini, E., & Di Pierro, R. (2020). Shame behind the corner? A daily diary investigation of pathological narcissism. *Journal of Research in Personality, 85*, 103924.
- Fuller-Tyszkiewicz, M., Hartley-Clark, L., Cummins, R. A., Tomin, A. J., Weinberg, M. K., & Richardson, B. (2017). Using dynamic factor analysis to provide insights into data reliability in experience sampling studies. *Psychological Assessment, 29*(9), 1120–1128.
- Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods, 19*(1), 72–91.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis, 1*(3), 515–534.
- Hamaker, E. L., Asparouhov, T., Brose, A., Schmiedek, F., & Muthén, B. (2018). At the frontiers of modeling intensive longitudinal data: Dynamic structural equation models for the affective measurements from the COGITO study. *Multivariate Behavioral Research, 53*(6), 820–841.
- Neubauer, A. B., Schmidt, A., Schmiedek, F., & Dirk, J. (2022). Dynamic reciprocal relations of achievement goals with daily experiences of academic success and failure: An ambulatory assessment study. *Learning and Instruction, 81*, 101617.

- Ringwald, W. R., Manuck, S. B., Marsland, A. L., & Wright, A. G. (2022). Psychometric evaluation of a Big Five personality state scale for intensive longitudinal studies. *Assessment*, 29(6), 1301–1319.
- Wright, A. G., Stepp, S. D., Scott, L. N., Hallquist, M. N., Beeney, J. E., Lazarus, S. A., & Pilkonis, P. A. (2017). The effect of pathological narcissism on interpersonal and affective processes in social interactions. *Journal of Abnormal Psychology*, 126(7), 898–910.
- Xiao, Y., Wang, P., & Liu, H. (2023). Assessing intra-and inter-individual reliabilities in intensive longitudinal studies: A two-level random dynamic model-based approach. *Psychological Methods*. Advance online publication. <https://doi.org/10.1037/met0000608>
-

第二轮

审稿人 1 意见：修改稿较好地回应了审稿人的意见，建议发表。

回应：感谢您对本文的肯定，您的建议对本文质量的提升有非常大的助益，非常感谢！

审稿人 2 意见：作者基于审稿意见进行了仔细的修改，阐述较为清晰，使得文章质量有了较大的提升。

回应：感谢您对本文的认可，您的建议帮助我们提升了文章的清晰度和可读性，非常感谢！

编委 1 意见：这篇文章根据两位审稿人的意见做了比较好的修改和完善，建议发表。但发表前要对照贵刊的格式要求进行修改，例如，中文参考文献条目，不应当出现“&”的符号。

回应：感谢您对本文的积极反馈和宝贵建议。我们参考期刊格式要求仔细核查了参考文献的相关内容，相应的修改已在文章中用黄色底纹标出。

编委 2 意见：仔细阅读了文章和对审稿专家意见的回复，认为目前的稿件已达到发表水平。

主编意见：根据编委和审稿专家的意见，建议发表。