

《心理学报》审稿意见与作者回应

题目：预测视角下双因子模型与高阶模型的模拟比较

作者：徐霜雪 俞宗火 董圣鸿 李月梅

主编初审

去掉那些数学推导部分，送审。

回应：非常感谢主编的意见。针对数学推导部分，已将其从正文中删掉，本文将其放在附录部分。

第一轮

审稿人1意见：

意见1：该文的主要结论是，在效标为潜变量时，高阶因子模型表现优于双因子模型。但是在引言中，作者并没有论述为什么要将效标分为显变量和潜变量分别进行模拟研究。1) 在用预测效度进行模型结构效度的评价中，有没有将效标设为潜变量的先例？请作者补充。2) 将效标分为显变量和潜变量分别进行模拟研究，作者是否根据文献，是否会预期到会有结果的不同？这样将效标分为显变量和潜变量进行研究就有了较为充分的理据。

回应：非常感谢评审专家提出的宝贵意见和建议。根据您的建议，我们已对引文部分做了修改。将效标分为显变量和潜变量分别进行模拟研究的理论依据部分，我们在引言中补充内容如下（见文中引言的标蓝部分）：

作为衡量测验有效性的外在标准，效标既可以是工资绩效等外显变量，也可能是积极情绪、自尊等内潜变量(Chen, Hayes, Carver, Laurenceau, & Zhang, 2012)。对人心理的测量具有间接性的特点，所以，大多心理学变量都是内潜变量。如果这些内潜变量的得分仅仅是各个题目的得分的简单相加的话，会存在很多缺陷(顾红磊, 温忠麟, & 方杰, 2014)。所以, Reise等人(2013)用效标潜变量来评估用单维的方式来测量多维的心理学构念带来的结构系数偏差的大小。Muthén等人(1987)认为，结构系数偏差更可能发生在潜变量模型中。因此，我们预期，相对于效标变量为显变量的情况，当效标变量为潜变量时，结构系数偏差可能会更大。

意见2：研究者在进行应用研究的过程中，使用双因子模型或高阶模型进行分析时，本文的结论对应用研究者的意义在哪里呢？请作者进一步在讨论中阐述本研究的实际意义。

回应：感谢评审专家提出的意见。为了更清楚表明本文结论对应用研究这的意义，我们在文中讨论部分进行了如下的补充（见文末讨论与展望中标蓝部分）：

以往有关高阶因子模型和双模型的研究主要集中在模型拟合比较方面，而科学研究与应用的目的在于对世界上的各种现象进行描述、解释、预测和控制。本研究显示，相对于双因子模型而言，高阶因子模型能对效标变量做出更加精准的预测，这有助于应用研究者对其所研究的现象进行更好的预测和控制。因此，本研究的实践意义就在于研究者可以根据本研究

的结论更客观地评价两个模型的优劣,并最终根据研究目的与模型适配性来选用准确的模型,这样才能更准确地揭示变量之间的因果关系。

意见 3: 模拟研究中, 每种实验条件重复 100 次, 这个是否存在依据? 请作者补充。

回应: 感谢审稿专家提出的意见。一般来说, 重复实验次数越多, 实验误差越低, 实验精确度越高。从数学角度来说, 适宜重复次数应保证使误差项自由度至少为 12(管宇 & 杨琪瑜, 2005)。本研究中使用的是大样本, 要估计的残差项(即误差项)远小于样本量, 所以误差项的自由度远远大于 12, 100 次的重复实验次数可以达到估计参数的精确度。实际上, 在心理测量学中, 模拟研究的重复测验次数并没有一个严格的标准, 结合以往研究对重复测验次数取值来看, 100 次已可达到模拟实验所要求的精确度, 所以本研究出于软件运行速度的考虑, 选择 100 次重复实验次数。

审稿人 2 意见:

本文通过 Monte Carlo 模拟数据, 在大样本条件下 ($N = 1000$), 比较了当效标分别为外显变量和潜变量时, 双因子模型和高阶因子模型的不同因子负荷条件下预测精确度的差异。本文的局限在于:

意见 1: 本文的理论框架不甚明晰。以往的研究表明, 两模型虽然在分数解释上有所分歧, 然而两者在个别拟合指标上差异即使显著, 差别也很小; 而且两模型都能达到良好的拟合指数标准。本文作者指出, 从测量角度考虑, 预测效度是很重要, 于是要进一步探明两模型在不同条件下预测精确度的差异。然而, 从统计学的角度来说, 作者基于什么考虑认为应对两模型做出比较?

回应: 感谢审稿专家提出的问题。双因子模型一般是作为高阶模型的竞争模型提出(Chen et al., 2012; Chen, Jing, Hayes, & Lee, 2013; Chen, West, & Sousa, 2006; Rindskopf & Rose, 1988; Yung, Thissen, & Mcleod, 1999)。单纯从数理统计角度分析, 双因子模型与高阶因子模型具有嵌套关系, 后者嵌套于前者(Yung et al., 1999), 在满足比例约束的条件下二者是等价的(Schmid & Leiman, 1957)。双因子模型的优点主要体现在探讨局部因子的作用时, 可以通过局部因子的负荷直接判断其作用大小 (Chen et al., 2006; Reise, Moore, & Haviland, 2010); 相对于高阶因子模型而言, 双因子模型也有其不足: 模型更加复杂, 需要估计的参数更多, 在小样本条件下, 有时需要提供初始值才能收敛(Chen et al., 2006; 顾红磊 et al., 2014)。正是因为相对于高阶模型来说, 双因子模型有这些不足, 而以往的研究更多表明双因子模型比高阶因子模型更好, 忽视了它的不足。所以我们认为, 有必要从更多的方面(如, 本文的预测效度视角)对二者进行比较研究。

意见 2: 本文采用的数据是依据双因子模型的参数值, 模拟而来的。拟合模型采用了双因子和高阶因子模型。结论是, 即便双因子模型是正确的模型, 当效标为潜变量时, 该模型比起高阶模型, 仍然产生出带有较大偏差的结构系数。这个结论很令人意外。同时, 由于真实的模型往往是未知的, 读者会感到好奇, 如果模实验数据是依据高阶模型的参数值模拟而来的, 是否会得出同样的结论?

回应: 非常感谢审稿专家提出宝贵的意见。原文可能存在一些表述不清的地方, 我们已在文中 3.1 部分进行了修正(见蓝色字体部分)。修正为:

本文是用 Monte Carlo 模拟的方法, 按照双因子模型下相应的参数值, 导出方差-协方差矩阵, 由矩阵出发产生数据。按照 Schmid-Leiman(1957) 转换方法可将双因子模型的参数值和高阶因子模型的参数相对应起来, 此时方差-协方差矩阵也相当于是由高阶因子模型相应

参数值而导出。其中，双因子模型中全局因子的负荷是高阶因子模型中低阶因子负荷与高阶因子负荷的乘积，而局部因子的负荷则为低阶因子负荷与低阶因子残差的平方根（也可转化为低阶因子的残差在低阶因子上的负荷）之积。

所以，研究中的数据并不是单纯由双因子模型的参数值模拟而来，其中也隐含了高阶因子模型的一套参数值，两者是对应的。此外，本文的结论也验证了这一点，即如果数据产生是由双因子模型而来，那么双因子模型的预测效度应该好于高阶因子模型的预测效度，然而本文的结果显示高阶模型的预测效度要好于双因子模型。

意见 3: 本文仅考虑了样本数为 1000 的大样本情况。据文献，在 $N=1000$ 时，有足够 Power 来检测两模型拟合的差别，而当样本偏小时（此种情况在心理学实验中通常出现），两模型常常得出类似的结果。读者更感兴趣的是，在小样本条件下（例如， $N=200$ 以及 $N=500$ ），本文的结论是否依然成立。

回应: 感谢审稿专家提出的意见。当 $N=200$ 以及 500 时，结果显示两个模型的结构系数偏差都相对更大，但高阶因子模型的预测效度总体上要好于双因子模型，这个趋势和 $N=1000$ 时的结果是一样的。这可能是本研究中模型较为复杂造成的。

此外，文中还存在着些表述不清楚的地方。

意见 4: 由于两模型的路径并不完全一一对应。那么，作者是如何计算出高阶模型的结构参数偏差的？文中需要添加一实例加以说明。

回应: 非常感谢审稿专家提出的问题。高阶模型的结构系数偏差的计算是本研究的一个特殊点，正如专家所说，两模型的路径不是完全一一对应的。关于如何计算出高阶模型的结构参数偏差，我们已在文中补充了实例（见文中 2.3 标蓝部分），如下文：

结合高阶因子模型与双因子模型的预测图可知，双因子模型的结构系数偏差的计算可直接通过全局因子与局部因子对效标变量的预测系数估计值与真值的偏差百分比计算出，同样地，高阶因子模型中的高阶因子的结构系数偏差也可直接通过其对效标变量的预测系数估计值与真值的偏差百分比计算出，然而，低阶因子的结构系数偏差不可以直接通过低阶因子对效标变量的预测系数计算，但是低阶因子的结构系数可转化为低阶因子的残差对效标的预测系数估计值，从而计算出其与真值的偏差百分比。

在进行模型指定的时候，直接用低阶因子的残差去预测效标变量，即残差回归方法（residual regression model）(Chen et al., 2012)。这种方法所蕴含的原理是由 Gustafsson 等人 (1993) 提出的，即将低阶因子的方差固定为 0，低阶因子的残差对低阶因子的负荷固定为 1，此时，与双因子模型中各局部因子的效应相对应，低阶因子的效应即可转化为低阶因子的残差在低阶因子上的负荷估计值与低阶因子在各观测变量上的负荷估计值的乘积，低阶因子的残差转化成因子的形式，对效标变量进行预测，其回归系数即为高阶因子模型中低阶因子对效标变量的结构系数，其结构系数偏差是直接计算残差对效标变量的预测系数估计值与真值的偏差百分比。例如，当全局因子为 0.7，局部因子为 0.7 时，高阶因子模型中低阶因子的残差对低阶因子的负荷估计值为 0.752（某次实验时的模型估计值），低阶因子在观测变量 Y1 上的负荷估计值为 0.989，高阶因子在低阶因子上的负荷估计值为 0.659，此时，低阶因子的残差对观测变量 Y1 的效应为 $0.752 \times 0.989 = 0.7437$ （相当于双因子模型中局部因子负荷的估计值，约为 0.7），此时，低阶因子的残差就可以直接对效标变量进行预测，且所估计出的预测系数可以直接计算出结构系数偏差的大小。

意见 5: 效标的数据是如何模拟出来的？使用了哪些参数？

回应: 感谢审稿专家提出的问题。研究一中效标观测变量的数据产生方法,是通过设定全局因子和 3 个局部因子对效标观测变量的回归系数为 0.3,且效标观测变量为服从均值为 0,方差为 1 的标准正态分布。这样就可以建构模型与效标变量之间的方差-协方差矩阵,从而产生数据。

研究二中的效标潜变量,是由 3 个观测变量组成,这 3 个观测变量在潜变量上的负荷都设为 0.7(设置方法是根据 Reise 等人(2013)设置效标测量模型的原则且基于计算简便原则而来),指定全局因子和 3 个局部因子对效标潜变量的回归系数都为 0.3,这 3 个观测变量也都服从均值为 0,方差为 1 的标准正态分布,同时,效标潜变量的方差设为 1,由此可建构模型变量与效标变量之间的方差-协方差矩阵,从而产生数据;文章研究一中研究方法部分以及研究二中的研究方法部分对效标变量的产生进行了相应的说明。

其他问题:

意见 6: 从表一来看,当模拟参数中全局因子负荷为 0.7,局部因子负荷为 0.6 时,模型拟合指数都是高于其他大多数因子负荷。作者怎么看待模型拟合好,然而结构系数偏差反而大的现象呢?

回应: 感谢审稿专家提出的问题。侯杰泰等人(2004)认为:结构方程的应用研究中,检视理论模型与数据的拟合指数,若拟合度高,则说明数据证明理论模型正确的说法,其实不够严谨,严格而言属于谬误。结构方程分析中的验证,其实只是辨别哪些模型有错误(与数据相斥),需要摒弃或修正。模型拟合效果好不一定说明变量之间真正存在因果关系。

意见 7: 第五页:最后一行“目前还很难判断。。。”一句中,删除“出版”二字。

回应: 感谢评审专家提出的意见。已删掉最后一行“目前还很难判断。。。”一句中的“出版”二字。

意见 8: 图 2 中的两个路径图中存在明显的错误。图 Mc 中,箭头应改为从 S2 指向 Criterion。图 Md 存在同样的错误,箭头应改为从 S2 指向 Criterion。另外,这是个高阶模型,而图中并没有画出高阶潜变量。

回应: 非常感谢评审专家的提出本文存在的错误。图 2 中的两个路径图出现的问题确实是在画图时候出现了错误。在文中我们已经进行了修正。

意见 9: 结果:需要汇报在同一实验条件下,100 次重复中拟合失败的数目。

回应: 感谢评审专家提出的意见。已根据您的建议在文中补充报告了同一实验条件下,100 次重复中拟合失败的数目,具体见文中标蓝部分。

意见 10: 目前通常用于结构方程建模的软件还有 LISREL 和 Mplus。这两种软件可以用于高阶模型预测吗?为了普及高阶模型预测的使用,建议作者在附录部分提供 EQS 或其他软件的程序。

回应: 感谢评审专家提出的意见。国外有研究者使用 Mplus 软件用于高阶模型预测的研究,如 Chen 等人(2012)使用 Mplus 软件,对数据用高阶模型及其残差回归模型(residual regression model)来分析,以探察外向性与外部变量之间的关系。所以,从理论上来说 LISREL 和 Mplus 软件都可以用于高阶模型预测,但是本研究出于程序运行的便利性考虑,使用了 EQS 软件,作为一个样例,针对全局因子负荷为 0.7,局部因子负荷为 0.7 的实验条件下效标为观测变量时,两个模型预测效度的程序代码已在附录 2 给出。

第二轮

审稿人 1 意见:

意见 1: 作者对我提出的问题进行了针对性的修改, 我对修改表示满意。

回应: 非常感谢评审专家的肯定与支持。

审稿人 2 意见:

意见 1: 作者已在上次审稿意见的基础上作了仔细的修改, 我没有新的意见。不过, 上次审稿意见中, 审稿专家提出, 由于在心理学样本数通常较小 ($N = 200, N = 500$), 而通常此时模拟数据的结果与真值的偏差较大样本 ($N = 1000$) 更大。因此, 建议作者把 $N = 200$ 及 $N = 500$ 条件下的结果作为文章附录 (Appendix)。

回应: 非常感谢审稿专家的建议。根据专家的意见, 我们对这部分内容做了补充, 放在附录 3 和 4 中。由附录 3 和 4 可以看出, 所有实验条件下 (即 $N = 200$ 及 500 时), 两个模型的拟合良好, 且并无显著差异。但在预测效度方面, 高阶因子模型的预测效度总体上都要好于双因子模型, 这个趋势和 $N = 1000$ 时的结果是一样的, 或者说, 正文中的主要结论, 在 $N = 200$ 及 500 时, 同样是成立的。此外, 当 $N = 200$ 时, 无论效标变量为显变量还是潜变量, 高阶因子模型和双因子模型的结构系数偏差都比 $N = 1000$ 时的更大, 而当 $N = 500$ 时, 高阶因子模型和双因子模型的结构系数偏差比 $N = 200$ 时要小, 但仍大于 $N = 1000$ 时的结构系数偏差。这说明随着样本容量的增大, 模型的结构系数偏差会变小。

编委复审

意见 1: It is now 11000 words, given there are 2 studies, it is alright.

回应: 谢谢编委专家的意见和肯定, 非常感谢。

意见 2: I have polished the English abstract for the authors' consideration.

回应: 非常感谢编委专家的宝贵意见和建议。已根据您对英文摘要的修改进行更正和修饰。