

《心理学报》审稿意见与作者回应

题目：青少年早期亲社会行为倾向、内化问题和外化问题发展级联的个体内分析

作者：段文婷 孙启武 王铭 吴才智 陈真珍

第一轮

审稿人 1 意见：

该论文采用了为期三年的面板数据，探讨了初中生亲社会行为倾向、内化问题和外化问题的发展级联，并将其结果与传统的交叉滞后面板模型进行比较，相比传统数据分析方法在区分个体内效应和个体间效应有一定的优势。研究结果为青少年心理健康的促进提供了支持。研究设计为三年的追踪研究，研究设计合理，工作量大，研究采用问卷法收集数据，数据较客观，但来源单一，统计分析严谨。

意见 1：

在预测青少年问题行为的发展上作者从亲社会行为倾向的角度出发分析内外化问题的发展，理论依据有待加强。作者在前言中指出“从概念上讲，亲社会行为的增加会减少外化问题（例如，攻击性）；同时，由于亲社会行为需要较好的情绪调节能力以及对积极情绪的促进作用（例如，助人为乐），亲社会行为也会减少内化问题（例如，抑郁）。”这里也提到“情绪调节能力”“积极情绪”是其中的主要因素。那么，作为主要变量的亲社会行为与问题行为是否同时受到了这些因素的影响而出现了相关关系和预测关系呢？在青少年阶段，导致个体出现内外化问题的关键因素是个体的亲社会行为吗？作者没有考虑这个阶段的同伴关系、亲子关系、学业压力以及父母行为的原因是什么？尽管作者在数据分析上系统全面，但理论假设模型的建立需要谨慎推导。

回应：

谢谢专家意见。我们理解专家的意见为本研究在阐述理论模型的充分性上还不够清晰，理论依据有待加强。我们尝试指出过去的文献在亲社会行为、内化问题和外化问题的发展级联研究上，对发展的非连续问题重视不够，并增加了文献，以说明青少年早期心理病理的发展也具有非连续性。同时，增加了图 1，以说明心理病理发展连续性和非连续性的含义。这对之后的模型选择提供了部分理论依据。因为 RI-CLPM 假设连续性变化，而 GCLM 能处理非连续变化及其后续影响。具体修改见 1.4 节。

就专家提出的具体问题，分别回答如下：

1.1 这里也提到“情绪调节能力”“积极情绪”是其中的主要因素。作为主要变量的亲社会行为与问题行为是否同时受到了这些因素的影响而出现了相关关系和预测关系呢？

回应：是的，二个变量因为有共同因素的影响，能使本来没有的关系，变为有相关关系或预测关系，即混淆变量所造成的虚假效用。在研究策略上，一是尽可能构建完备的理论模型，将主要的影响因子纳入模型（即控制协变量的方法）；二是采用随机分组，以平衡这些共同因素的在不同条件下的混淆作用。由于

在心理学的研究中，很多场合无法像随机对照实验进行随机分组，所以，传统的策略就是要尽可能构建完备的理论模型，以消除共同因素的影响。在本研究的情形中，从理论上，我们推测“情绪调节能力”是亲社会行为、内化问题和外化问题的共同影响因素，因此，模型中未纳入“情绪调节能力”就有可能造成亲社会行为、内化问题和外化问题之间虚假的相关关系或预测关系。

这正是传统分析模型的缺点。理由一是在观察数据中，我们无法让所有的潜在相关因素都纳入某一项具体的研究，因为研究者很有可能并不知道还有哪些混淆因素没有纳入模型；二是实践上有些潜在因素无法直接测量，或者测量有难度、成本高昂（例如，基因、家庭环境）等。个体内分析方法在一定程度上克服了传统分析模型的不足。其逻辑是将难以测量或者未测量的变量，例如，基因、家庭环境等等时不变变量（就是在追踪研究设计中稳定不变的变量，如性别，人格特质，基因组等）分解为个体内水平效应（组内）和个体间水平（组间）效应。组间效用不仅包含了“情绪调节能力”和“积极情绪”（这是个体间的稳定差异）的影响，还包含了诸如基因、家庭环境、父母受教育程度等未纳入模型因素的影响（即使有虚假效用，也体现在组间效用当中）。组内效用则不再受到这些混淆因素的影响。或者换句话说，变量在个体内的关系可以视为以个体自身为对照，随时间变化的相互预测关系。在某种程度上，这就是一种排除了各种混淆变量影响的因果关系。

RI-CLPM 和 GCLM 都可以进行个体内分析，但在假设条件、适用的研究问题以及组间变量对组内变量关系的影响等方面也有重要不同。在对评审专家 2 第 11 点的回答中，有关于二者异同的详细说明。这里就不再赘述了。

目前，有关观察数据的因果推论建模这部分的相关文章和讨论在心理学领域还比较少，但在经济学、社会学等相关学科有大量讨论。心理学仍然有不少领域主要的研究手段仍然是观测数据，如何进行因果推论就成了心理学研究者迫切希望解决的问题。这里，我们又列举了一些我们所依据的关键参考文献。

总而言之，目前所使用的建模方法以及结论主要指个体内效用，个体间混淆变量对个体内效用的影响较小。

参考文献：

- Allison, P. D. (2019). Asymmetric fixed-effects models for panel data. *Socius: Sociological Research for a Dynamic World*, 5, 1–12. <https://doi.org/10.1177/2378023119826441>
- Falkenström, F., Finkel, S., Sandell, R., Rubel, J. A., & Holmqvist, R. (2017). Dynamic models of individual change in psychotherapy process research. *Journal of Consulting and Clinical Psychology*, 85(6), 537–549. <https://doi.org/10.1037/ccp0000203>
- Firebaugh, G., Warner, C., & Massoglia, M. (2013). Fixed effects, random effects, and hybrid models for causal analysis. In L. S. Morgan (Ed.), *Handbook of Causal Analysis for Social Research* (pp. 113–132). Springer Netherlands. https://doi.org/10.1007/978-94-007-6094-3_7
- McNeish, D., & Kelley, K. (2019). Fixed effects models versus mixed effects models for clustered data: Reviewing the approaches, disentangling the differences, and making recommendations. *Psychological Methods*, 24(1), 20–35. <https://doi.org/10.1037/met0000182>
- Mulder, J. D., & Hamaker, E. L. (2020). Three extensions of the random intercept cross-lagged panel model. *Structural Equation Modeling: A Multidisciplinary Journal*, 1-11. <https://doi.org/10.1080/10705511.2020.1784738>

1.2 在青少年阶段，导致个体出现内外化问题的关键因素是个体的亲社会行为吗？作者没有考虑这个阶段的同伴关系、亲子关系、学业压力以及父母行为的原因是什么？

回应：我们设想，应该是多因素共同作用导致个体出现内外化问题。如上回答，理由一是个体内分析以个体自身为对照，虽然在青少年早期同伴关系、亲子关系、学业压力以及父母行为都是非常重要的变量，但并不影响模型建构及个体内效用分析。除此之外，应该和具体的研究目的有关。同伴关系、亲子关系、学业压力以及父母行为已有大量研究，不是本研究的目标。我们试图从青少年心理健康促进角度，采取积极视角，探讨亲社会行为、内化问题和外化问题发展级联。在此基础上，有可能在学校或班级水平发展出某种整体干预方案，其理念与我国社会传统“助人为乐”理念契合，是心理健康促进的可为之处，有一定潜力。

意见 2：

前言部分的文献梳理多集中在纵向数据的统计分析方法上，而对变量与变量间的关系的综述较为缺乏。

回应：

不知我们对评审专家的意见理解是否准确。我们梳理了亲社会行为、内化问题和外化问题的一些主要概念、理论和研究结论。这些概念和理论的往往有不同出发点，因此，将它们统一在一个清晰的理论框架中，具有一定的难度。在本轮修改中，结合个体内分析模型选择的问题，我们以心理病理发展的连续性和非连续性作为研究问题提出的切入点。另外，本研究之所以要讨论纵向数据的统计分析方法，主要的考虑是个体内分析的设计逻辑（例如，如何较排除混淆变量的影响，为什么有可能得到因果关系等）可能对不少读者而言不太熟悉，需要适当介绍一些背景知识；并且，模型的选择也以青少年早期的发展特点和模型的假设条件为依据。具体可参考 1.1 回应以及评审专家 2 第 11 个问题的回应。

意见 3：

文献综述与问题提出之间逻辑性有待加强。作者引用“Padilla-Walker, Carlo, & Nielson, 2015”文献指出亲社会行为的实施对象不同其社会行为的适应意义也会有所不同，但在自己的研究设计中并没有控制同伴/友谊关系这一变量。

回应：

见对第 1、2 个问题的回应，在个体内分析中，我们没有采用控制协变量的研究策略。为使论述完整，我们在讨论中增添了一句，以使论述完整：“例如，可从亲社会行为的对象不同其社会行为的适应意义不同入手，检查对陌生人、熟人和同伴的亲社会行为行为及其对内化问题和外化问题的影响(Padilla-Walker et al., 2015)” (p.18)。

意见 4：

内外化问题的测量上作者采用“抑郁-焦虑-压力”自评量表和“攻击性”问卷，焦虑抑郁和攻击性仅是内外化问题的某个方面，因此通过这两个问卷能否有效测量“内化”和“外化”问题行为值得商榷。

回应：

不知道我们对专家的意见理解是否准确。我们同意焦虑抑郁和攻击性仅是内外化问题的某个方面。我们用焦虑抑郁和攻击性作为反映内化问题和外化问题的指标。这个问题的准确表述不是有效性的问题，而是指标是否全面反映了内化问题和外化问题的“域”或“因子”。我们并没有试图全面测量“内化”和“外化”问题（这在研究中也会有可行性上的困难，在追踪研究中尤其如此）。或许没有争议的说法是“青少年早期亲社会行为、心理困扰与攻击性的个体内分析”。在文献中，有大量研究采用了类似的逻辑，即使用某个指标或部分指标来测量内化问题和外化问题。与专家商榷：经过考虑后，我们仍然采用现在的说法，以“抑郁-焦虑-压力”自评量表和“攻击性”问卷反映内化问题和外化问题。但在讨论中，将此点作为一个局限提出：

“本研究对内化问题和外化问题的测量，没有采用他评指标，而全部采用了自评指标，数据来源较为单一，指标也较为单一，这对研究结论会有潜在影响”（p.18）

参考文献：

Memmott-Elison, M. K., Holmgren, H. G., Padilla-Walker, L. M., & Hawkins, A. J. (2020). Associations between prosocial behavior, externalizing behaviors, and internalizing symptoms during adolescence: A meta-analysis. *Journal of Adolescence*, 80, 98–114. <https://doi.org/https://doi.org/10.1016/j.adolescence.2020.01.012>

意见 5：

分析模型的建立是否控制了人口学变量（如，儿童年龄，父母受教育程度等）

回应：

在缺失值分析中，我们考虑了父母教育程度是否为随机缺失。在个体内分析中，在个体内水平变量间的关系不受人口学变量的影响。具体见对第 1 个问题的回应。

意见 6：

讨论中对统计模型异同的分析个体内和个体间变量上不同上做了比较，而缺乏从青少年发展理论对角度分析；结果发现了性别差异，但结果分析讨论不足。建议作者挖掘追踪数据、重新梳理变量间关系，加强研究假设的理论基础，建立合理的假设模型，在此基础上进行统计分析和讨论。

回应：

感谢专家的评审意见！我们对讨论部分进行重新梳理，主要修改体现在两个方面。一是从青少年早期心理病理发展的连续性和非连续性的角度进行分析研究结果；二是对为何采信 GCLM 结果进行了说明；三是讨论了性别差异问题。新增添的部分都高亮标示了。

.....

审稿人 2 意见：

本研究采用三种交叉滞后模型考察了亲社会行为、外化问题和内化问题的发展级联。由于这三种变量之间的关系在文献中已有很多讨论，本文的价值很大程度上在于方法的示范，特别是用一组数据放在三个相互关联的纵向模型中，是否会产生一致的结论。然而在解释模型参数的意义、阐述模型的差别，以及模型选择的依据方面还有很大提升的空间。

意见 1:

第 4 页，介绍发展级联这个概念时，需要解释“域”、“水平”、“系统”是指什么，可考虑举例说明。

回应:

谢谢！已经在文中举例说明，并高亮标示了。

意见 2:

第 6 页，“亲社会行为的对象不同……”这一段，没有解释清楚为什么“帮助陌生人意味着社会情境性因素没有产生作用”，而“对朋友的亲社会行为……与情境性因素关系更为密切”。

回应:

谢谢！我们举例补充说明，并高亮标示了。此外，为使本段表述准确，我们将“社会情境因素”改为了“社会影响因素”，从而与 Lai 等(2015)的研究结论表述一致。

参考文献:

Lai, F. H. Y., Siu, A. M. H., & Shek, D. T. L. (2015). Individual and social predictors of prosocial behavior among Chinese adolescents in Hong Kong. *Frontiers in Pediatrics*, 3, 39. <https://doi.org/10.3389/fped.2015.00039>

意见 3:

第 9 页，前者 and 后者指什么，表达不清。图 1 中有三个模型，作者指的是哪一个？

回应:

谢谢专家提醒。我们就此已经修改为“在图 1c 中， $\mu_{t-2}^{(x)}$ 可预测 x_{t-2} ，这是变量 x 在 t-2 时点波动值影响自回归效用；而 $\mu_{t-2}^{(x)}$ 预测 x_{t-2} 并通过 x_{t-2} 而预测 x_{t-1} 和 y_t ，这是 x 在 t-2 时点的波动值通过影响 x_{t-1} 而间接影响 y_t 的交叉滞后效用。”

意见 4:

表 2 相关表，用斜体表示不显著的相关系数不清晰，可以考虑用粗体标示显著相关，这样全文标示符号保持一致。

回应:

谢谢专家提醒。我们按照专家建议修改，使表 2 更加清晰易读;也与表 3 结果表示方法一致。

意见 5: 等值性检验的结论没有报告，这部分内容应该在文章中报告。另外，等值性检验如果增加了等值约束的模型与前一个模型差别显著，那么说明前一个模型更好，不需要再继续再比有更多等值约束的模型了。

回应:

谢谢专家提醒。我们按照专家建议，在 2.4.3 节报告了等值性检验结果 (p.12)。表 1 中删除了内化问题 M4.1,但保留了 M3.1。理由是我们想知道究竟不等值的时间点究竟在哪里，这能为我们后续讨论青少年发展的连续性和非连续性，以及为什么使用 GCLM 提供部分证据。

意见 6:

第 13 页，作者报告的 CLPM 的自由度和 RI-CLPM 与 GCLM 的自由度一致，都是 13，这是不可能的。CLPM 模型没有因子载荷，它的参数应该少于后两种模型，因此它的自由度应该大于后两种模型。这里请作者再检查一下是否弄错了。

回应:

感谢审稿专家的仔细审阅。我们又检查了模型设定，没有发现模型设定上的错误。在三个模型中，参数都是 41 个，自由度为 13。当然，我们也可以使用 RI-CLPM 模型设定语法，固定其中 η 相关的效用，就可以得到和 CLPM 相同的结果。以 CLPM 和 RI-CLPM 为例，在直观上是否可以这样理解：RI-CLPM 只是把随机截距分解出去了并固定相应参数，残差部分模型估计与 CLPM 是一样的。

模型设定的不一致，会导致自由度不同，这样模型间也就难以比较了，比如截距是否固定，是否考虑自回归效用、交叉滞后效用和残差项的平稳随机过程等 (stationary assumption)。我们附上 CLPM 和 RI-CLPM 语法，供专家审阅。其中 PSO 表示亲社会行为，EXT 表示外化问题，INT 表示内化问题。

这是 CLPM 模型语法模型设定部分:

```
DATA:      FILE IS cas_update.txt;
VARIABLE:  NAMES ARE id gender int1-int3
           ext1-ext3 pso1-pso3;
           Usevariable =ext1-ext3 int1-int3 pso1-pso3;
           Missing are all (-999);
analysis:  ESTIMATOR = MLR;
           !MODEL = NOCOVARIANCES;

MODEL:
  !AR effect
  PSO2 ON PSO1;  PSO3 ON PSO2;  EXT2 ON EXT1;  EXT3 ON EXT2;  INT2 ON INT1;
INT3 ON INT2;
  !Lagged-effect
  EXT2 ON INT1 PSO1;  EXT3 ON INT2 PSO2;  INT2 ON EXT1 PSO1;  INT3 ON EXT2 PSO2;
  !co-movements
  PSO1 with EXT1;  PSO1 with INT1;  INT1 with EXT1;
  INT2 with EXT2;  PSO2 with EXT2;  PSO2 with INT2;
  INT3 with EXT3;  PSO3 with EXT3;  PSO2 with INT3;
```

这是 RI-CLPM 语法模型设定部分:

```
DATA:      FILE IS cas_male_female.txt;
VARIABLE:  NAMES ARE id gender int1-int3
           ext1-ext3 pso1-pso3;
```

```

Usevariable =ext1-ext3 int1-int3 pso1-pso3;
Missing are all (-999);
analysis: !type = general ;
          !INFORMATION=EXPECTED;
          !iterations = 1000;
          ESTIMATOR = MLR;
          MODEL = NOCOVARIANCES;
MODEL:
! Create random intercepts for each variable
!unit effects
EXT BY ext1@1 ext2@1 ext3@1;
int BY int1@1 int2@1 int3@1;
PSO BY pso1@1 pso2@1 pso3@1;
! Estimate the means of the random intercepts and
! fix all other means/intercepts in the model to zero
      [EXT*];
      [int*];[PSO*];
      [ext1@0];
      [int1@0];[pso1@0];
      [ext2-ext3@0]!model 1,unconstrained
      [int2-int3@0]!model 1,unconstrained
      [pso2-pso3@0]!model 1,unconstrained
! Create person-mean centered variables
cint1 by int1@1; cint2 by int2@1; cint3 by int3@1;
cpso1 by pso1@1; cpso2 by pso2@1; cpso3 by pso3@1;
cext1 by ext1@1; cext2 by ext2@1; cext3 by ext3@1;
! Constrain the measurement error variance to zero
int1@0; int2@0; int3@0;ext1@0;ext2@0;ext3@0; pso1@0;
pso2@0; pso3@0;
!autoregressive (AR factor)
cint2 on cint1 ;!(ar1);
cint3 on cint2 ;!(ar1);
cpso2 on cpso1 ;!(ar2);
cpso3 on cpso2 ;!(ar2);
cext2 on cext1 ;!(ar3);

```

```

cext3 on cext2 ;!(ar3);

!cross-lagged effects
cext2 on cpso1 ;! (lag1);
cext3 on cpso2 ;! (lag1); !prosocal predict ext
cint2 on cpso1 ;! (lag2);
cint3 on cpso2 ;! (lag2); !prosocal predict int

cext2 on cint1 ;! (lag3);
cext3 on cint2 ;! (lag3); !int predict ext
cint2 on cext1 ;! (lag4);
cint3 on cext2 ;! (lag4); !ext predict int

!relations at the 1st wave
cint1 with cext1 cpso1;
cext1 with cpso1;

!residual covariance(co-movements)
cpso2 with cint2; cpso2 with cext2;!year 2015
cext2 with cint2;

cpso3 with cint3; cpso3 with cext3;!year 2016
cext3 with cint3;

!unit effects correlation
EXT with int; EXT with PSO;
int with PSO;

!restriction
EXT with cpso1@0 cint1@0 cext1@0;
int with cpso1@0 cint1@0 cext1@0;
PSO with cpso1@0 cint1@0 cext1@0;

```

意见 7:

模型的 lambda 没有报告，RI-CLPM 和 GCLM 的载荷是否一致？如果非常不一致，那么两个模型抽取出的组间差异 eta 还具有可比性吗？Usami（2020）这篇文章中讨论了 GCLM 的 eta 和 RI-CLPM 的 eta 的意

义并不一致。建议作者在介绍三种模型时（第 8 页）准确严谨地介绍每个模型中参数代表的意义，特别是 GCLM。即使它和 RI-CLPM 的参数有相似之处，甚至用了同样的符号，它们的意义并不相同。

Usami, S. (2020). On the differences between General Cross-Lagged Panel Model and Random-Intercept Cross-Lagged Panel Model: Interpretation of cross-lagged parameters and model choice. *Structural Equation Modeling: A Multidisciplinary Journal*, 28(3), 1–14. <https://doi.org/10.1080/10705511.2020.1821690>

回应：

感谢专家的仔细审阅和宝贵意见。因所述内容较多，请详见后面我们对第 11 个问题的回应。的确，在初稿中，我们没有说明 eta 的意义，两个模型的 eta 意义也不完全相同。我们在修改稿中大幅度修改了模型选择的考虑过程（2.4.1 节）。要点概括如下：第一，我们的焦点在个体内水平变量的关系，正如文题所强调的那样，我们不关心个体间差异的比较；第二，RI-CLPM 和 GCLM 源自不同的学科的建模传统不同。第三，将青少年心理病理发展的连续性和非连续性纳入模型选择的考虑，也就是说，不完全依靠统计模型，而是依据心理学的理论选择合适的模型。第四，简要地介绍了 GCLM 模型的背景。

意见 8：

研究中采用的 GCLM 模型没有包含通常的 MA 和 CLMA 效应，这样做是出于什么考虑呢？

回应：

谢谢专家提醒。在我们的理解中，考虑到某种干预措施不会一直实施，也就是干预措施往往不会是一个时间序列数据，所以 GCLM 的一个优势是，在追踪数据中，如果采取了某项干预措施（例如某项影响经济发展的政策，或者某种自然情境下的干预项目），则可通过 MA 和 CLMA 项检查其效用，以及效用的作用路径。在心理学的干预研究中，如果也采用了追踪设计，则可通过 MA 和 CLMA 项检查其效用，以及效用的作用路径。也就是说，如果有某种干预的话，根据不同的研究目标，GCLM 中则可有针对性地纳入 MA 和 CLMA 项，以检验某些特定的研究假设。本研究中，是一个自然调查的追踪设计，没有干预措施，所以没有采用 MA 和 CLMA 项。另外，我们也很难找到可靠依据，例如，为什么初二学生会会有某个特殊的波动值。此外，在 GCLM 中，如果包含所有的 MA 和 CLMA 项，是一个过拟合模型，因为参数比自由度多。

意见 9：

在三种模型中选择 GCLM 的理由没有阐述，就直接在发展联级检验中直接采用了 GCLM 模型。需要说明为什么采信 GCLM，放弃另外两种模型。

回应：

感谢专家的仔细审阅。请详见我们对第 7 点和第 11 点的回应，以及对审稿专家 1 的第 1-3 点的回应。另外，我们更新了讨论部分，以更清晰说明采信 GCLM 的依据。更新部分已经高亮标示了。

意见 10：

研究用 GCLM 模型考察几个变量在个体内的关系，得到了与假设相反的结论（内化问题和外化问题的交叉滞后效应出现显著负值），作者认为是个体内关系和个体间关系的不同导致的。然而研究并未讨论变

量之间的个体间关系。建议作者在结果中报告个体间的关系，然后就可以和个体内的关系进行比较和讨论了。比如介绍表 2（相关表）结果时就是一个很好的讨论个体间关系的地方。

回应：

感谢专家的建议！由于本文主要目标在于进行个体内分析，并且在 RI-CLPM 和 GCLM 中，eta 项具有不同的含义，因此，无法直接比较。为避免歧义，我们删除了讨论部分“个体内关系和个体间关系的不同”的相关表述，因为我们并没有讨论变量在组间水平的关系。

意见 11：

GCLM 得到的自回归效应和交叉滞后效应不能看作是个体内的关系，因为模型中的效应是从变量的观测值出发指向另一个变量的观测值，而观测值包含了组内和组间的信息。GCLM 区分了共变效应的组内和组间部分，但是未区分自回归和交叉滞后效应的组内和组间部分。如果纯考察组内关系，应该看 RI-CLPM，而不是 GCLM。

回应：

不知我们对审稿专家的意见理解是否准确，这个问题涉及 GCLM 与 RI-CLPM 之间的区别与联系。GCLM 确实用组间水平变量矫正了组内水平变量的关系，但组内水平变量关系仍然是个体内分析。我们尝试就此进行说明，看能否与评审专家达成共识。

传统 CLPM 的一个不足是混淆了个体间差异和个体内差异。一个自然的想法是，如同在多层线性模型中所做的那样，对变量进行去中心化处理，例如减去个体 n 次重复测量的均值。去中心化后，变量间的关系就是个体内的关系了。但是，这种处理方法没有采用潜变量的思想。个体 n 次重复测量的均值可以视为一个潜变量。在重复测量次数较小的情况下（本研究情形为三次）容易犯所谓 Nickel 偏差，即用观察值均值去中心化会使得自回归效用估计值偏小。这就是 RI-CLPM 的最初想法。用多层结构方程模型的方法可以部分克服 Nickel 偏差，例如 RI-CLPM 和 DSEM（Asparouhov & Muthén, 2019）等。以 RI-CLPM 为例，个体间的差异用潜变量 η 表示，其交叉滞后关系体现在潜变量 μ 的相互关系中，以表示在个体内水平变量的相互关系。

在表面上看，GCLM 采用了看起来不同的方法，但共同之处都在于将个体间的差异分离出去，在个体内水平变量间的交叉滞后效用代表了某种因果关系。不同之处在于：在 GCLM 中，变量的自回归效用和变量间的交叉滞后效用体现在显变量上，而不在潜变量上（DSEM 的个体内效用也体现在显变量上）。GCLM 是 Allison（2017）模型的一种复杂形式，而 Allison（2017）的模型来源自所谓固定效用模型（fixed-effect model）。这里的固定效用（常见于经济学领域）与心理学研究中常见的多层线性模型的固定效用所指不同（McNeish & Kelly, 2019）。所以，我们先从固定效用模型说起。

对时间序列数据而言，固定效用模型一个最简单的形式可以写作：

$$Y_{i,t} = \mu + \beta X_{i,t-1} + \alpha_i + \varepsilon_{i,t} \quad (1)$$

其中， μ 为截距， β 为回归系数， α_i 为个体间的稳定差异， $\varepsilon_{i,t}$ 为测量误差，该误差与方程右边的项无关。

这个方程的 α_i 项即可视为包含了所有个体间的稳定差异。通过这样的方法，即能以个体自身为控制，在时间序列数据中，在个体内水平探讨变量间的关系。常见的估计方法有均值去中心化，一阶差分法（first difference），或构建虚拟变量的方法等，这样就保留了个体内变量的变化值；为服务不同的研究目标，针对不同的时间序列数据残差序列关系的假设，常见的估计方法有最小二乘法(OLS)，广义最小二乘法(GLS)，最大似然法（ML）等。这个模型也可以写成结构方程形式。用结构方程可以用下图表示（这里只包含了 X 对 Y 的影响）：

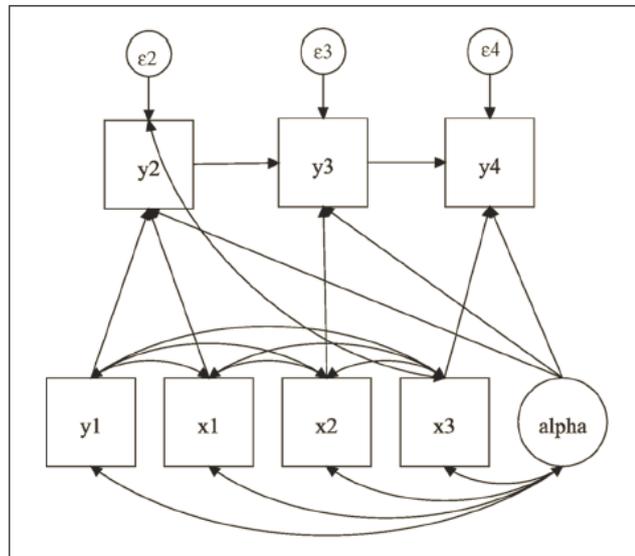


Figure 1. Path diagram for 4-period dynamic panel model.

（图 1：引自 Allison(2017), p.6）

有两点值得注意。一是在这个模型中（见公式 1）， α （即个体间所有的稳定差异）对 y 有影响，并且会因此修正 x 对 y 的交叉滞后效用，以及 y 的自回归效用。由于 α_i 项包含了所有个体间的稳定差异，在此模型中， x 对 y 的交叉滞后效用和 y 的自回归效用都是个体内水平的效用。二是这个模型（公式 1），只是假设误差项与公式右边其它项无关，但并没有假设 α_i 项与公式的其它项无关（即考虑到所谓外生性问题的影响, exogeneity assumption，以及初始条件 x_1, y_1 的影响）。在上图中的体现，即 α 与 y 的初始水平 y_1 和 x_1 - x_4 都有相关，并“影响”（调整）之后的自回归效用和交叉滞后效用。

GCLM（包括 Allison 的模型）与 RI-CLPM 的重要不同在于假设更少，或许也更符合本研究的实际情况。GCLM 可以处理不稳定的系统。如果系统是不稳定的（我们设想在青少年早期急剧变化的时期就是如此），则用 RI-CLPM 进行处理则不符合实际情况，容易带来模型估计偏差。

就系统的稳定性概念，用时间序列的自回归效用（当然，通常测量的时点较大，比如 $n > 20$ ）说明如下：

$$X_t = \alpha X_{t-1} + \varepsilon_t \quad (2)$$

其中 t 是实数，表示测量的时间点， ε_t 为测量误差。从初值 X_0 出发， α 越小，初值影响减小越快。当 $|\alpha|$ 接近 1，初值和前面的 ε_{t-j} 影响减小越慢。只要 $|\alpha| < 1$ ，序列最终可以稳定下来。这时系统是**稳定的**。但如果 $|\alpha| > 1$ ，则系统是**非稳定的**。这时，就没有均值回归(mean reversion)效用了。

(可参考：

https://www.math.pku.edu.cn/teachers/lidf/course/atsa/atsanotes/html/_atsanotes/atsa-arstation.html)。

这种讨论在心理学中较为少见。如果我们认为青少年早期的发展是既有连续性，也有非连续性(图 2)，用 RI-CLPM 分析数据，就违反了模型的假设。在此种情形下，判断依据是青少年发展的理论而不是统计模型本身。

在本研究中，如果假设三个变量的追踪数据所构成的“系统”是稳定的，并且假设在个体内水平变量间的关系不受个体间水平时不变变量的影响，则可以用 RI-CLPM 进行处理。RI-CLPM 的优势将随机截距解释为稳定的特质变量，并将随机截距分解后所剩的残差解释为个体在这个时点的变化值，的确简洁明了，易于解释。GCLM 也可以处理稳定的系统，只不过自回归效用和交叉滞后效用为个体间水平变量 (η) 矫正了(即在图 1 中 α 的影响)，但自回归效用和交叉滞后效用仍然为个体内水平效应¹，虽然并不像 RI-CLPM 那样直观。

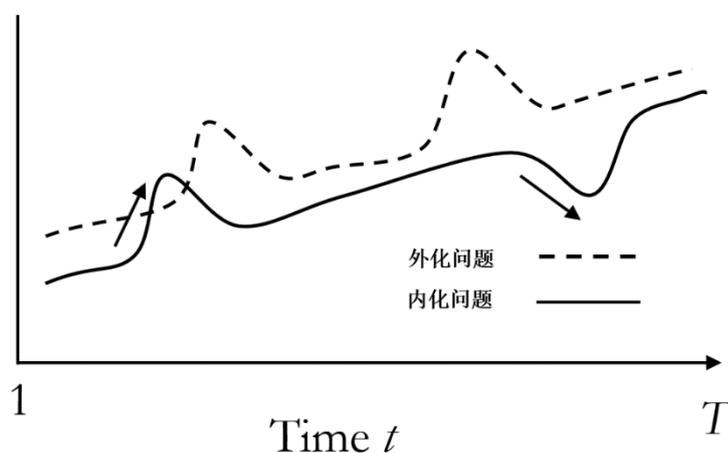


图 2：青少年心理病理发展的连续性和非连续性

说明：纵轴为某种发展指标，例如内化问题和外化问题；横轴为时间。

参考文献：

Firebaugh, G., Warner, C., & Massoglia, M. (2013). Fixed effects, random effects, and hybrid models for causal analysis. In L. S.

¹就此，我们也与 GCLM 提出者 Mike Zyphur 通信。我们对 GCLM 的理解与 Zyphur 的解释是相近的，Zyphur 与 Usami(2020) 的看法并不完全相同(通信时间，2021 年 5 月 11 日)。我们猜想，可能的原因是 Hamaker 和她的学生 Usami 主要是心理学背景，而 Allison 和 Zyphur 主要是经济学背景，因考虑的研究对象不同，采用了不同的建模方法。现在，在观察数据的因果推论文献中，研究者提出心理学和经济学的建模方法需要相互学习，取长补短(McNeish & Kelly, 2019)。此外，RI-CLPM 在发展研究应用中的适合性仍然在激烈的争论当中，不过，就此的详细讨论已经超出了本文的范围了。

Morgan (Ed.), *Handbook of Causal Analysis for Social Research* (pp. 113–132). Springer Netherlands.
https://doi.org/10.1007/978-94-007-6094-3_7

Allison, P. D., Williams, R., & Moral-Benito, E. (2017). Maximum Likelihood for cross-lagged panel models with fixed effects. *Socius*, 3, 2378023117710578. <https://doi.org/10.1177/2378023117710578>

Asparouhov, T., & Muthén, B. (2019). Comparison of models for the analysis of intensive longitudinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(2), 275–297. <https://doi.org/10.1080/10705511.2019.1626733>

McNeish, D., & Kelley, K. (2019). Fixed effects models versus mixed effects models for clustered data: Reviewing the approaches, disentangling the differences, and making recommendations. *Psychological Methods*, 24(1), 20–35. <https://doi.org/10.1037/met0000182>

孙启武, 吴才智, 于丽霞, 王巍欣, 沈国成. (2021). 阅读进度反馈信息对工作同盟和咨询效果的影响. *心理学报*, 53(4), 349–361.

第二轮

审稿人 1 意见:

作者回答了前面提出的问题, 但有一些问题还需作者认真思考。

回应:

感谢专家进一步提升修改的建议。在第一轮和第二轮审稿中, 专家老师都对本文的文献部分不够满意。本轮修改我们又重新梳理了文献。在梳理过程中, 我们对本研究本身也有更多新的认识。这个过程对提升论文的质量帮助非常大。十分感谢!

需要说明两点: 首先, 亲社会行为倾向、内化问题和外化问题的发展级联本身已经有大量相关研究 (Memmott-Elison et al., 2020), 其理论建构大多数仍然为“解释性的”, 而非“可检验的”, 这是本领域文献中的一个现状。在我们的理解中, 对亲社会行为倾向、内化问题和外化问题的发展级联建构可检验的理论模型本身非常复杂, 是一项非常有挑战性的工作, 目前尚且不是本研究的目标。并且, 经过认真思考和讨论, 我们认为本研究的数据不足以检查这些假设。同时, 我们的想法与专家意见一致, 这方面的理论建构和检验确实还很不够。

其次, 我们的想法是, 文献部分着重介绍的内容要围绕本研究的目标。本研究目标一是在中国早期青少年群体中检验亲社会行为倾向、内化问题和外化问题的发展级联。这实际上是一个重复研究, 我们的研究目标在于检验之前文献中所观察到的亲社会行为倾向、内化问题和外化问题关联性及其性别差异在中国早期青少年群体中是否也能得到数据的支持。本研究的目标之二, 也是本研究的优势所在, 是采用个体内的分析方法来检验亲社会行为倾向、内化问题和外化问题的发展级联。个体内分析有许多不同的模型, 我们采用最近受到较多关注的 RI-CLPM 和 GCLM。这两个模型具有不同的假设, 但在实际应用中常常被忽视。因为它们是新方法, 所以容易忽视采用模型所需满足的条件, 从而容易给一些研究者造成误解甚至误用。

在本轮修改中, 我们补充了连续性问题简要的理论说明, 以期待能加强模型选择和研究问题之间的内在逻辑性。上述考虑如有不足、不妥之处, 还请专家老师再予指正, 甚感!

参考文献:

Memmott-Elison, M. K., Holmgren, H. G., Padilla-Walker, L. M., & Hawkins, A. J. (2020). Associations between prosocial behavior, externalizing behaviors, and internalizing symptoms during adolescence: A meta-analysis. *Journal of Adolescence*, 80, 98–114. <https://doi.org/10.1016/j.adolescence.2020.01.012>

意见 1:

作者在摘要部分提到的结论是“内化问题和外化问题的相互影响具有性别差异”，且仅此一条。在前言中对性别差异的文献综述和理论假设推导薄弱。另外，作者在前言中叙述了亲社会行为与内化问题和外化问题关系的影响因素，但在研究问题和假设中并没有相关的内容。

回应:

感谢评审专家的意见。我们做了进一步的修改，具体说明如下:

首先，关于“内化问题和外化问题的相互影响具有性别差异”这一研究结论。本文的主旨是采用个体内分析方法探讨初中生亲社会行为倾向、内化问题和外化问题的发展级联及可能的因果关系，并比较不同分析方法对结果的可能影响。在投稿时，考虑到中文摘要的字数限制（200 字以内），之前的摘要确实没能精炼地表述出本文的主要研究结果与结论，现做了修改与调整。如下:

认识亲社会行为倾向、内化问题和外化问题的发展级联对青少年心理健康促进具有重要价值。采用个体内分析方法分析了初中生三年的追踪数据，并与传统的交叉滞后面板模型进行了比较。结果发现：（1）亲社会行为倾向与内化问题和外化问题具有中等程度共变关系，其公开、依从、情绪因子可负向预测下一年内化问题；（2）内化问题与外化问题间存在较强共变关系，具有中等或较弱的相互预测关系；（3）男生 T2 的外化问题能负向预测 T3 的内化问题，女生 T2 的内化问题能正向预测 T3 的外化问题；（4）个体内分析方法得到的研究结果与传统交叉滞后分析方法不同。研究表明：亲社会行为倾向可降低青少年内化问题，青少年心理健康促进方案要考虑内化问题与外化问题的相互影响及其性别差异；个体内分析方法在青少年心理病理发展研究中具有重要应用价值。

第二，关于性别差异的文献综述与理论假设推导。根据专家建议，我们在综述中对亲社会行为倾向对内化问题和外化问题的影响因素部分做了修改，增加了性别因素的影响，并增加了性别相关的研究假设。详细见正文 1.3 节以及 1.5 节研究假设部分。在理论假设推导方面，我们简要介绍了心理病理发展级联的连续性和非连续性问题的理论说明，详细见正文 1.4 节。

第三，关于亲社会行为与内化问题和外化问题的影响因素。我们认为这部分内容确实会使得论证过程不够聚焦，离开我们想要讨论的主要问题。为此，我们删除了这部分无关内容，增加了性别相关因素的影响，也同时删除了讨论部分的相关语句，以使得文章更为连贯。

意见 2:

作者在文中描述的有关发展的连续性和非连续性问题似乎和前人探讨心理发展的连续性和非连续性（阶段性）问题的角度不同。如图 1 中两条代表发展趋势的线出现波峰和波谷，这可能说明的是非线性的

发展趋势，而不是发展的“非连续性”。因此，作者提到的“从青少年早期心理病理发展的连续性和非连续性的角度进行分析研究结果”有待商榷。

回应：

感谢评审专家的意见。为避免前言部分过于冗长，避免读者的误解，我们删除了图 1 及相关说明。同时，我们增加了对心理病理发展的连续性和非连续性问题的理论解释。对此，说明如下：

首先，原图 1（参见第一轮审稿对审稿专家 2 回应中的图 2）的横轴是时间，波峰和波谷表示出现了个体内的某种变化及其时序关系。当然，从时间序列上讲，我们可以采用不同的连续性曲线拟合数据，例如线性增长、二次增长、三次增长等。但这些增长都是连续性的发展，而非连续的发展是指某种时序上的突然变化。在潜变量增长模型中，这有时也被表述为具有不同的潜在类别。以本研究为例，内化问题和外化问题的相互正向预测关系表明的是某种连续性发展，但在某个时间点，内化问题和外化问题却变为了相互负向的预测关系，则是某种非连续性的发展。或许用公式可以更清晰地说明：

$Y_i = f_j + \varepsilon_i$ ，其中 $f_j = f1 \ t_j \leq n, f_j = f2 \ t_j > n$ 。这个公式表示在 j 时间内 Y 可以某个函数 f 表示。当时点小于等于 n 时， Y 可以用函数 $f1$ 描述；当时点大于 n 时， Y 可以用函数 $f2$ 描述。如果 $f1=f2$ ，则是一个连续变化，如果 $f1 \neq f2$ ，则是一个非连续变化。

其次，我们讨论的心理病理发展确实与传统的连续性和阶段性不同。我们在正文 1.4 进行了简要的区分说明。这是儿童青少年心理病理发展级联的基本问题。当然，如何处理这一问题具有许多不同的建模方法，如不同的潜变量增长模型等。例如，分阶段混合增长模型就将这种非连续性分为不同的阶段建模。在本研究中，连续性与非连续性问题并非我们关注的问题，但其为我们在进行模型选择时提供了参考。我们认为这是后续研究值得用力的一个点。

参考文献：

Kohli, N., Hughes, J., Wang, C., Zopluoglu, C., & Davison, M. L. (2015). Fitting a linear-linear piecewise growth mixture model with unknown knots: A comparison of two common approaches to inference. *Psychological Methods, 20*(2), 259–275. <https://doi.org/10.1037/met0000034>

意见 3：

作者在前言和讨论中的叙述是关于亲社会行为，并提到“亲社会行为的对象不同，与内化问题和外化问题关联性也不同”，但后面并没有对其进行探讨和分析。建议作者进一步加强文献综述与研究问题提出之间逻辑关系，加强研究假设的理论基础，建立合理的假设模型，更加严谨地对研究发现进行讨论和推论。

回应：

谢谢专家提出的改进建议。此处“亲社会行为的对象不同”的文献综述确实与本文的主要分析内容关联不大，已经删除。我们增加了与本文主要探讨的调节影响因素年龄、性别有关的文献，调整了相关讨论和结论的表述。通过增加这部分内容，试图增加文献综述与假设基础、讨论及推论之间的逻辑关系。

意见 4：

通过补充“本研究对内化问题和外化问题的测量，没有采用他评指标，而全部采用了自评指标，数据来源较为单一，指标也较为单一，这对研究结论会有潜在影响”（p.18），并没有很好的回答了“研究采用抑郁-焦虑-压力自评量表和攻击性两个问卷是否有效测量了‘内化’和‘外化’问题行为”的问题。

回应：

感谢评审专家的意见。不知我们是否准确理解了专家的意见。以下我们从定义、文献和量表结构三个方面回答，看能否与专家达成共识。我们对内化问题和外化问题的定义也进行了修改。修改部分在正文高亮标出了。

在定义上，我们在正文中给出了内化问题和外化问题的定义(1.1 节第一句话)，即“内化问题通常指与情绪相关的困扰，而外化问题通常指行为的失调(Graber & Sontag, 2009)”。另外，Memmott-Elison 等人（2020）的元分析也给出了类似的定义，他们使用的是“外化行为”和“内化症状”（externalizing behaviors and internalizing symptoms）。行文时我们之所以没有使用“外化行为”和“内化症状”，是考虑到内化问题和外化问题更符合中文文献的表达习惯。其中，外化行为的定义通常为：由于自我调节能力缺乏所致的负面行为（Cole et al., 1996）。例如：攻击性、异常同伴交往、多动症、危险性行为、药物使用和犯罪及其它外化行为。内化症状的定义为：内化症状是与过度自我调节相关的内部（如认知、情感）困难（Cole et al., 1996），包括抑郁、（低）自尊、自我伤害/自杀意念和抑郁等一般内化症状等。据此，我们认为从定义上来看，攻击性问卷符合外化问题的定义；而抑郁-焦虑-压力自评量表符合内化问题的定义。

文献中，以近几年学报已经发表的文章为举例。如下：何晓丽等人（2020）的研究采用了 Achenbach 和 Edelbrock (1987)编制的青少年行为自评量表，全量表共计 112 道题目，分为焦虑/抑郁、退缩、躯体主诉、社交问题、思维问题、注意问题、违纪行为和攻击行为 8 个分量表。郭海英等人（2017）则采用社交焦虑、抑郁和孤独感表示内化问题。刘晓薇等人（2021）使用儿童行为核查表青少年自我报告版中文版测量青少年外化问题行为，量表包含攻击行为和违纪行为两个维度。谢其利等人（2021）采用 Goodman (2001)编制的困难与长处问卷中文版的品行问题(Conduct Problems)和多动—注意不能分问卷，测量学生自我报告的外化问题行为(如，多动、不服从、撒谎、打架等)。常淑敏(2019)采用 Achenbach (1991)编制的儿童行为量表中外化问题分量表的教师报告版测量初中生的外化问题行为，该量表包括攻击行为和违纪行为两个维度。可见，研究者对外化问题和内化问题的测量有多种测量内化问题和外化问题的方法，但在其测量的具体内容上，不同研究之间还是一致的，即抑郁、焦虑等测量的是内化问题，违纪攻击性等测量的是外化问题。本研究采用攻击问卷测量的是外化问题，采用抑郁-焦虑-压力自评量表测量的是内化问题，与以往研究一致。

或许有疑虑的是，DASS-21 中的“压力”是否也是内化问题？在量表结构上，有研究证据表明抑郁-焦虑-压力自评量表测量的是一个共同的心理困扰因子（common mental disorder）（Osman et al., 2012）。在本研究样本中，以量表因子分为指标，表明抑郁、焦虑和压力三个因子测量的是同一个潜变量，可达到弱等值性，即因子结构和因子负荷在三个测量时点相同（详细见正文测量等值性检验部分）。

综上，我们认为抑郁-焦虑-压力自评量表和攻击性两个问卷有效测量了‘内化’和‘外化’问题。若有不足、不妥之处，还请专家老师再予指正。

参考文献：

常淑敏，张丽娅，王玲晓(2019). 发展资源在减少青少年外化问题行为中的累积效应及关系模式. *心理学报*, 51(11):

- 郭海英, 陈丽华, 叶枝, 潘瑾, 林丹华. (2017). 流动儿童同伴侵害的特点及与内化问题的循环作用关系: 一项追踪研究. *心理学报*, 49(3), 336–348.
- 何晓丽, 袁小龙, 胡铭, 周丽晨. (2020). 父母元情绪理念与青少年问题行为: 迷走神经的调节作用. *心理学报*, 52(8), 971–981. doi:10.3724/SP.J.1041.2020.00971
- 刘晓薇, 潘斌, 陈亮, 李腾飞, 纪林芹, 张文新(2021). 受欺负与青少年外化问题关系的“健康环境悖论”: 敌意性归因的中介作用. *心理学报*, 53(2), 170–181. doi:10.3724/SP.J.1041.2021.00170
- 谢其利, 郑惠珍, 江光荣, 任志洪, 樊燕飞, 刘加怀, 张文. (2021). 班主任协商管理行为、师生关系与小学 4~6 年级学生外化问题行为的相互作用关系: 交叉滞后研究. *心理学报*, 53(10), 1120–1132.
- Graber, J. A., & Sontag, L. M. (2009). Internalizing problems during adolescence. In R. M. Lerner, L. Steinberg (Ed.), *Handbook of adolescent psychology: Individual bases of adolescent development, Vol. 1, 3rd ed.* (pp. 642–682). New York: John Wiley & Sons Inc. <https://doi.org/10.1002/9780470479193.adlpsy001020>
- Memmott-Elison, M. K., Holmgren, H. G., Padilla-Walker, L. M., & Hawkins, A. J. (2020). Associations between prosocial behavior, externalizing behaviors, and internalizing symptoms during adolescence: A meta-analysis. *Journal of Adolescence*, 80, 98–114. <https://doi.org/10.1016/j.adolescence.2020.01.012>
- Osman, A., Wong, J. L., Bagge, C. L., Freedenthal, S., Gutierrez, P. M., & Lozano, G. (2012). The Depression Anxiety Stress Scales–21 (DASS–21): Further examination of dimensions, scale reliability, and correlates. *Journal of Clinical Psychology*, 68, 1322–1338. <https://doi.org/10.1002/jclp.21908>

此外, 经考虑以及进一步的审读, 我们还进行了一些小的修改。比如, 将“一般交叉滞后模型”改写为“广义交叉滞后模型”; 删除了 1.2 节原第三段, 因为这部分内容会使文献部分不够聚焦等等。这些修改在正文中都作了标示, 此处不再赘述。

.....

审稿人 2 意见:

作者很好地回应了我在阅读初稿时的一些疑问, 并作出了相应修改。我对这些修改满意, 但还有一些小的问题罗列如下。总之, 这是一篇值得推荐的文章, 特别是三种方法的实证比较和说明会对很多研究者有借鉴意义。

回应:

十分感谢审稿专家的肯定。专家老师的修改意见对我们进一步提高论文质量帮助很大!

意见 1:

连续性和非连续性问题在第 4 页首次提出后没有给出解释。建议增加一两句话简单解释一下这个概念。我看到对这个概念的具体解释在第 7 页, 但是距离第一次提出概念有些远, 读者带着这个疑问读下来, 认知负荷比较重。

回应:

感谢专家的建议! 这样确实更好!

为更好地说明模型选择过程中的考虑，我们在第一轮修改中引入了连续性和非连续性的概念。确实，在上一轮修改中没有立刻解释这个概念，且容易与传统概念混淆。本轮修改在提出连续性问题后，用举例的方法说明了连续性和非连续性问题的所指。本轮修改新增 1.4 节，新引 Cicchetti 和 Toth(1998)的文章，以期能对此现象做一个理论上解释说明。

另外，相关说明也可见对审稿专家 1 意见 2 的回答。

Cicchetti, D., & Toth, S. L. (1998). The development of depression in children and adolescents. *American Psychologist*, 53(2), 221–241. <https://doi.org/10.1037/0003-066X.53.2.221>

意见 2:

第 9 页有一个笔误，“三个变量之间的发展级联模型与图 1 类似”，应该是“与图 2 类似”。

回应:

谢谢专家的认真审读与意见！我们对此有所疏忽，所有作者对全文的文字表述再次进行仔细核对。此外，考虑到原图 1 容易引起误解以及造成前言部分过分冗长，我们在修改稿中删除了图 1。

意见 3:

第 10 页，第二段第一句话没有主语。“在 RI-CLPM 和 GCLM 模型中，虽然符号相同，但意义并不相同”，应该说明什么符号相同。或许可以改为“某些参数虽然使用的符号相同”。

回应:

谢谢专家的仔细阅读！修改稿中已经更正！

意见 4:

从作者给出的代码看，在模型设置时，作者把图 2 中后两个模型的 λ 均固定为 1 了，这没有问题，也解释了为什么在结果部分后两个模型和第一个简单模型的参数数量是一样的。为了减少读者困惑，建议把图 2 中的 λ 就写作 1，或者在表 3 的说明中说清楚 λ 固定为 1。不然从图 2 看，显然后两个模型的参数多于第一个模型，不应该得到一样的自由度。

回应:

谢谢专家的意见和建议！为减少读者困惑，已经按照专家建议修改，在表 3 中增加了如下说明：“在 RI-CLPM 和 GCLM 模型估计中，图 1 中的 λ 固定为 1 了，所以它们与 CLPM 的自由度相同。”

意见 5:

在这一稿中，作者并没有选择 GCLM 模型，而是把三种模型的结果并排呈现。而由于三种模型得到的结论并不一致，不确定这样做是否会给读者带来困扰。从拟合结果看，GCLM 显然好于 CLPM 和 RI-CLPM，无论是 CFI, RMSEA, AIC、BIC，还是卡方检验。或许作者可以提 GCLM 拟合的更好，但出于严谨，也把另外两个模型的结果展示出来，这样三个模型的结果更加确证，而不同的结果需要谨慎。

回应:

谢谢专家的意见和建议！我们重新修改了相关表述。具体如下(p.18):

此外，所有的模型拟合指标都提示 GCLM 优于 CLPM 和 RI-CLPM。这些结果都支持采信 GCLM 的结果。我们同时呈现了 CLPM 和 RI-CLPM 的结果。这样，三个模型的共同结果可更加确证，而对不同的结果，我们则需要谨慎考虑不同模型的适用条件。

第三轮

主编意见:

作者较好地回复了审稿专家的建议，但目前文中仍存在部分问题如下:

意见 1:

作者在文中假设了内化问题和外化问题的相互预测关系，将亲社会行为作为内化和外化问题的预测变量。但内外化问题同时也是亲社会行为的预测变量，亲社会行为和内外化问题之间也可能存在双向的预测关系。作者是否考虑添加内外化问题对亲社会行为的预测路径？或是更加突出亲社会行为作为预测变量的理论依据？

回应:

非常感谢主编老师的建议！经过认真讨论和仔细考虑，添加内外化问题对亲社会行为的预测路径和突出亲社会行为作为预测变量的理论依据都能使文章更为清晰、完整；并且，添加内外化问题对亲社会行为的预测路径后，会得到与文献一致性更多的结果。与原模型结果的主要不同是 T1 的外化问题可负向预测 T2 的亲社会行为倾向，而亲社会行为倾向可以负向预测内化问题和外化问题。具体如下：

第一、模型的修改过程。我们首先分析了新模型（M1），分别采用 CLPM、RI-CLPM 和 GCLM 分析数据，预测路径包括：内化问题和外化问题的相互预测，亲社会行为预测内化问题和外化问题以及内化问题和外化问题预测亲社会行为。结果显示三个模型与数据的拟合度均在可接受范围，其中 CLPM 拟合指数为：CLPM 拟合指数为 $\chi^2 = 68.16$, $df = 9$, CFI = 0.96, RMSEA = 0.087, SRMR = 0.037, AIC = 12368.75, BIC = 12583.12; RI-CLPM 拟合指数为 $\chi^2 = 36.72$, $df = 9$, CFI = 0.99, RMSEA = 0.060, SRMR = 0.020, AIC = 12325.38, BIC = 12539.76; GCLM 拟合指数为 $\chi^2 = 10.69$, $df = 9$, CFI = 0.99, RMSEA = 0.015, SRMR = 0.018, AIC = 12296.86, BIC = 12511.24。

原模型（M0,即不包括内外化问题对亲社会行为的预测路径）的结果为：CLPM 拟合指数为 $\chi^2 = 73.75$, $df = 13$, CFI = 0.96, RMSEA = 0.073, SRMR = 0.042, AIC = 12366.64, BIC = 12561.96; RI-CLPM 拟合指数为 $\chi^2 = 39.48$, $df = 13$, CFI = 0.99, RMSEA = 0.048, SRMR = 0.020, AIC = 12321.31, BIC = 12516.63; GCLM 拟合指数为 $\chi^2 = 23.93$, $df = 13$, CFI = 0.99, RMSEA = 0.031, SRMR = 0.038, AIC = 12305.68, BIC = 12501.01。

按照 Nylund 等（2007）的建议，嵌套模型比较，如果仅从统计上考虑，可参考 BIC。BIC 值越小，则模型更为优化。从结果上看，加入内外化问题对亲社会行为的预测路径后，CLPM 和 RI-CLPM 和 GCLM 的 BIC 值均有所升高。考虑到内化问题对亲社会行为倾向的预测路径均不显著，我们只保留外化问题对亲社会行为倾向的预测路径(M2)。

结果显示三个模型与数据的拟合度均在可接受范围, 其中: CLPM 拟合指数为 $\chi^2 = 69.79$, $df = 11$, CFI = 0.96, RMSEA = 0.079, SRMR = 0.038, AIC = 12365.95, BIC = 12570.80; RI-CLPM 拟合指数为 $\chi^2 = 36.35$, $df = 11$, CFI = 0.99, RMSEA = 0.052, SRMR = 0.020, AIC = 12321.94, BIC = 12526.79; GCLM 拟合指数为 $\chi^2 = 11.44$, $df = 11$, CFI = 1.00, RMSEA = 0.007, SRMR = 0.018, AIC = 12293.52, BIC = 12498.37。

与原模型(M0)比较, CLPM 和 RI-CLPM 模型的 BIC 值仍然高于原模型的 BIC 值。但 GCLM 的结果优于原模型。我们呈现了新模型 (M1) 的结果, 并基于 M1 做进一步的分析。其理由是: (1) 通常在 GCLM 和 RI-CLPM 建模中, 除非有坚实的理论依据, 才修改模型; (2) 我们也仔细比较了 M1 和 M2, 在预测路径的显著性上, 二者相同。

第二, 增添了亲社会行为作为预测变量的理论依据。虽然有少量文献提示亲社会行为和外化问题之间有相互的预测作用, 但这并非本文的重点。我们更关心亲社会行为的发展对内化问题和外化问题的预测作用。新增添的理论依据概括了三种不同的观点: 一是参与亲社会行为的儿童青少年通常有较好的自我调节能力, 从而不太可能因过度自我调节而出现内化问题, 也不太可能因自我调节不足而出现外化问题 (Graber & Sontag, 2009; Memmott-Elison et al., 2020)。二是从社会信息加工的角度, 研究者认为参与亲社会行为的儿童青少年往往具有向善的归因偏向与社会能力; 并且, 参与亲社会行为会引发他人的积极回应, 从而能巩固儿童青少年积极的内部工作模型。具有积极的内部工作模型的儿童青少年能更充分、更准确加工社会情绪信息 (包括负面的情绪信息), 从而降低内化问题或外化问题 (Cassidy & Shaver, 2016; Flouri & Sarmadi, 2016)。三是从亲社会行为动机过程的角度, 亲社会行为具有不同的类型, 亲社会行为也具有不同的对象; 这些不同使得亲社会行为可能具有不同的动机过程 (Davidov et al., 2016; Paulus, 2018), 比如“关心别人需要什么”, “关心他人的不幸”或“在别人面前做正确的事”这几种亲社会行为动机, 其社会认知过程和自我调节过程并不完全相同。例如, 过度“关心别人的需要” (例如, “听话”) 反而意味着过度的自我调节, 从而带来内化问题。

第三, 由于模型 M1 的 GCLM 结果提示内化问题和外化问题不能相互预测, 我们猜想是因为模型中加入了亲社会行为倾向。从理论上讲, 亲社会行为倾向既能影响内化问题也能影响外化问题; 因此, 我们考虑, 内化问题和外化问题不能相互预测是因为模型中加入了亲社会行为倾向, 从而“稀释”了内化问题和外化问题的相互预测作用。如果 M1 的 GCLM 模型中包括内化问题和外化问题的相互预测作用, 那么外化问题或许能显著预测内化问题。结果提示, 外化问题可以显著负向预测内化问题, 而内化问题不能预测外化问题。这个结果与文献的结果一致。该部分结果见新增添的 3.4 节。

参考文献:

- Cassidy, J., & Shaver, P. R. (2016). *Handbook of Attachment: Theory, Research, and Clinical Applications (Third Edition)*. New York: The Guilford Press.
- Davidov, M., Vaish, A., Knafo-Noam, A., & Hastings, P. D. (2016). The motivational foundations of prosocial behavior from a developmental perspective—evolutionary roots and key psychological mechanisms: Introduction to the special section. *Child Development, 87*(6), 1655–1667. <https://doi.org/10.1111/cdev.12639>
- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural equation modeling: A multidisciplinary Journal, 14*(4), 535–569. <https://doi.org/10.1080/10705510701575396>

意见 2:

目前文中对与假设不一致结果的讨论较为笼统。例如对性别差异的解释，仅说明“与瑞士儿童追踪数据的个体内分析结果一致”，那么这种一致的结果要如何进行解释？或“可能与分析方法有关”，那么分析方法具体是带来了何种影响？；对个体内分析结果的解释为“发展级联的非连续性”，那么非连续性是何种原因造成的？目前较为笼统的解释未能使读者对结果完全信服，建议作者更深入地探讨结果与假设不一致的原因。

回应:

谢谢主编老师的建议！第一，我们加强了对性别差异的解释，解释了在女生群体中亲社会行为倾向不能预测内化问题和外化问题可能原因，解释了在男生群体中内化问题和外化问题不能相互预测的可能原因。第二，我们解释了分析方法为何对统计结果有影响，有何影响。第三，我们弱化了“发展级联的非连续性”解释，将其作为模型选择的依据，删除了原有解释。以上具体见正文高亮部分。

此外，我们还仔细考虑了讨论部分的逻辑，具体做法一是修改了讨论内容的顺序，二是删除了一些关联性不够密切的统计细节，三是依据亲社会行为倾向对内化问题预测的理论观点解释研究结果。

意见 3:

Buss-Warren 攻击性问卷的内部一致性信度似乎偏低，请加以解释。

回应:

谢谢主编老师的建议！在本研究样本中，Buss-Warren 攻击性问卷五个因子 2014~2016 年的内部一致性信度如下表：

因子	2014 年	2015 年	2016 年
身体攻击（8 条目）	0.83	0.84	0.83
言语攻击（5 条目）	0.65	0.62	0.63
敌意（8 条目）	0.78	0.79	0.78
愤怒（7 条目）	0.71	0.73	0.74
间接攻击（6 条目）	0.71	0.74	0.69
总量表	0.92	0.93	0.92

从量表因子分来看，言语攻击因子内部一致性信度偏低。我国 2011 年报告的一项全国大样本调查（年龄 9~18 岁，样本量 23904）发现 Buss-Warren 攻击性问卷的内部一致性信度偏低，尤其以言语攻击为最低（ $\alpha = 0.51$ ），但因其内容效度和预测效度较好，至今仍然被广泛采用。在本研究样本中，总量表的内部一致性信度较高，在个体内分析中，我们采用了内化问题和外化问题的总分进行了分析，因此言语攻击因子内部一致性信度偏低对结果的影响较小。

参考文献:

张萍. (2011). *中国农村 9-18 岁中小学生攻击问卷标准化常模及简版攻击问卷的研制——攻击问卷本土化研究*. 武汉: 华中科技大学, 硕士论文, 未出版.

意见 4:

文中存在一些细节表述不正确的问题，例如 P14 结果描述中，“RI-CLPM 的结果显示仅 T2 的内化问题可以正向预测 T3 的内化问题（INT2→EXT3）”应为“……可以正向预测 T3 的外化问题”。请作者进行仔细检查及修改。

回应:

感谢您的指正！我们已经重新校读，并进行相应的修改。

意见 5:

文中存在一些参考文献格式问题。例如“寇彧 & 张庆鹏, 2006”，中文著者不加“&”，而是用逗号隔开；“高欣 et al., 2013”，中文著者不用 et al.，而用“等”；参考文献列表中的字体和格式也不统一。请作者对照心理学报参考文献要求进行仔细检查及修改。

回应:

参考文献格式问题确实有所疏漏，感谢您的指正！已经按照心理学报对参考文献要求进行了仔细检查及修改。

意见 6:

研究结论的展示方式不符合要求，太过于笼统的描述，而应该具体清晰化，1.2.3.。结论不用再描述意义等无关因素。整体文章的文字表达与逻辑水平要再提高，才符合心理学报的要求。

回应:

谢谢主编老师的建议！我们按照心理学报的要求重写了研究结论，并梳理了全文的文字表述，以期待更加清晰准确。研究结论的具体修改如下：

本研究表明：（1）青少年早期亲社会行为倾向的发展可降低内化问题和外化问题，不同类型的亲社会行为倾向对内化问题和外化问题的预测作用不同。（2）青少年早期亲社会行为倾向、内化问题和外化问题的发展级联具有性别差异。（3）广义交叉滞后面板模型（GCLM）更符合发展级联模型的理论描述，能兼顾青少年早期发展连续性和非连续的特点，适合于分析青少年早期的发展级联问题。