

《心理学报》审稿意见与作者回应

题目：儿童认知发展水平诊断工具 IPDT 的动态化编制及其在低社会经济地位儿童中的应用
作者：张丽锦，暴卿，陈蕾，梁渊

第一轮

审稿人 1 意见：

本研究编制了一套针对表征、分类、规律领域的小学儿童认知发展水平动态测验，检验了此测验的合理性及对推理能力和数学成绩的预测作用，同时以低社会经济地位儿童为主要对象，考察了此测验对儿童潜能评估的准确性。研究所针对的问题有高度的实用价值和社会意义，整体的设计和操作规范，得出的结果具有很强的启示性。在论文写作方面，还存在一些不够明确的地方，建议作者调整。

意见 1：从整体来看，目前比较大的问题是，编制的动态测验在研究二中实际上只起到了划分潜能水平的作用。研究二中所关注的前后测和干预所针对的能力并非研究一编制的测验中所包含的能力（虽然可能有重叠）。但是在写作中，这一方面体现得不够明显，特别是由于文中多次提到动态测验本身即具有干预能力，因此在研究二的阅读中总有模糊感觉其干预本身的成就也是动态测验达成的。讨论 4.1 部分更是如此（这部分提到了数学成就，所以看起来也不是只针对研究二在说）。因此，建议作者在文中更明确地强调这一点。从这个角度看，研究二的重心不放在“干预”上，而更多放在“基于潜能评估”上，可能会使两个研究衔接更紧密。此外，作者同时需要简单阐述为何选择推理和数学两个方面来检验新编制的动态测验的敏感性，几种能力之间是什么关系。

回应：十分感谢审稿专家提出的问题，尤其是对修改方向给予的建设性意见。专家的问题引发了我们对研究核心问题的再思考：（1）本研究二基于研究一所改编的 IPDT 动态测验对儿童的认知潜能进行评估和细致分类后，进一步通过“前测-干预-后测”范式来考查五年级儿童经推理认知干预后在认知与数学表现上的进步。在维果斯基关于“最近发展区”的概念（即，儿童独立完成的水平和在成人或有能力的同伴帮助下完成的水平，即现有水平与潜在水平之间的差距）与测量中，“前测-干预-后测”范式是典型的动态测验范式，以干预后的进步、提示量和/或迁移量作为考查潜在发展水平（或“最近发展区宽度”，或“认知改变能力”）的指标。也就是说，推理认知干预及效果的考查在动态测验的研究者们看来，也是一个动态干预研究。不过本研究的这个干预过程是非标准化的，所以在 3.2.3 程序中的“（4）推理认

知干预”中没有逐级干预和计分的内容；相应地，在 4.3 结果报告中也只有后测分和差异分（增长分）两类动态测验分数的描述。当然，本文在这方面确实存在对动态测验相关范式和程序表述不清而引起歧义的问题，结合第二位审稿专家的建议（见审稿专家 2 的第 1 个问题），我们将标题改为“儿童认知发展水平诊断工具 IPDT 的动态化编制及其在低社会经济地位儿童中的应用”。这样是否符合读者对动态测验一般意义的认识和理解？请审稿专家指正并赐教。同时，我们对研究二的内容阐述也进行了相应修改，将“对潜能的评估”部分进行了更明确而详尽的阐述（详见研究二的研究目的和结果部分）。（2）关于 IPDT 与推理认知和数学成绩的关系问题，IPDT 是 Furth（1970）对皮亚杰临床访谈法使用的一系列实物认知测验任务的图形纸笔化改编，是评估儿童认知发展的诊断性工具；而瑞文图形推理测验也是对一般认知能力的测查（张厚粲，王晓平，1989），这两种认知评估工具均为图形形式，均不受文化、教育水平、语言和受测人数的限制（张厚粲 等，1989；方富熹，盖笑松，张丽锦，2005）。并且研究者认为，IPDT 可以帮助教师对推理困难的学生进行甄别，从而进行更有针对性的教育（Patterson & Milakofsky, 1980）。此外，以往研究证实，IPDT 与数学成绩存在显著相关（ $r = 0.55, p < 0.05$ ）（Bender & Milakofsky, 1982）；而且 IPDT 还可用于儿童数学课程的教学准备，以使数学课程内容更适合学生的认知水平（Patterson & Milakofsky, 1980）。研究二的推理认知干预材料是参照瑞文图形推理和其他图形推理任务改编和整理，任务形式与图形推理任务有着较高的一致性。因此，本研究二以推理认知和数学成就来检验 IPDT 动态测验的有效性（相关内容已补充进前言 1.4 部分）。

方富熹，盖笑松，张丽锦. (2005). 皮亚杰认知发展量表(IPDT)中国城市常模的制订. *中国心理卫生杂志*, 19(12), 789–792.

张厚粲，王晓平. (1985). *瑞文标准推理测验*. 北京：北京师范大学出版社.

Bender, D., & Milakofsky, L. (1982). College chemistry and Piaget: The relationship of aptitude and achievement measures. *Journal of Research in Science Teaching*, 19, 205–216.

Furth, H. (1970). *An inventory of Piaget's developmental tasks*. Washington, DC: Catholic University, Department of Psychology, Center for Research in Thinking and Language.

Patterson, H., & Milakofsky, L. (1980). A paper-and-pencil inventory for the assessment of Piaget's tasks. *Applied Psychological Measurement*, 4, 341–353.

意见 2：前言部分关于低社会经济地位对儿童发展影响及原因的介绍有些过于细致了。由于本研究的目的主要是在潜能划分和针对性干预，并不在详细的社会经济地位影响的机制探

讨，因此过于细致的介绍会显得与研究本身距离较远。建议作者对此内容进行总结，概括性介绍。这一问题比较突出地表现在 1.1 部分第一段后半段，1.2 部分第二段第三段。

回应：感谢审稿专家的意见和建议，我们完全接受，已对前言部分（具体为 1.1 部分第一段，1.2 部分第二段和第三段）进行了概括与删减。

意见 3：前言部分可能需要对 IPDT 的结构等做更深入的介绍。从现在的阐述来看，并不知道 IPDT 共涉及多少个领域，因此也就不太能理解作者选择表征、分类、规律领域的原因和想法。类似，文章中前后几处提到这几个领域适合小学高年级儿童，究竟是这些能力发展在这个年龄阶段敏感，还是这个测验中相关领域的内容适合高年级，还是本研究专门选择了小学高年级这个年龄段而因此选择了适合的题目？

回应：（1）感谢并接受审稿专家的建议与意见，我们对 IPDT 各维度的内容，以及选择表征、分类和规律领域进行动态化改编的原因进行了补充（见前言 1.4 部分）。经开发者（H. Furth）授权，国内研究者已将 IPDT 进行了适合中国儿童的改编（方富熹，盖笑松，龚少英，刘国雄，2004），IPDT 包含守恒、关系、表征、分类和规律五个领域。根据研究所报告的测量学指标和国内常模分数（方富熹 等，2005），守恒和关系领域的任务难度适合小学低-中年级儿童，且有研究者已将其改编为动态测验并应用于小学三年级儿童的认知潜能评估（张丽锦，陈亮，方富熹，2011）。（2）本研究根据 IPDT 全国常模中报告的表征、分类和规律领域的题目对小学高年级儿童的难度适当、区分度良好的结果（方富熹 等，2005），将其应用于五年级儿童，并对这三个领域进行了动态化改编。相关理由解释已补充到正文中（见研究一 2.1 部分）。

方富熹，盖笑松，龚少英，刘国雄. (2004). 对儿童认知发展水平诊断工具 IPDT 的信度效度检验. *心理学报*, 36, 96-102

张丽锦，陈亮，方富熹. (2011). “儿童认知发展水平诊断工具”动态测验的初步编制与应用. *心理学报*, 43(9), 1075-1086.

意见 4：对于研究二选择推理能力和数学成就作为促进的目标领域的原因同样需要解释。为何选择了这两个维度，它们与表征、分类、规律领域的关系是怎样的？明确这些内容有助于读者更清晰地理解作者的研究思路和研究目的。

回应：该问题与第一个问题的第 2 点有共同之处，并到“第一个问题的（2）”中答复了。同时在正文中也增加了相应必要解释（见前言 1.4 部分）。请审稿专家审查并指正。

意见 5: 3.3.1 部分, 要考察静态测验和动态测验的预测力, 为何选取后测成绩? 可以理解作者想要表明动态测验预测的是潜能, 但本研究中的后测成绩一部分是干预过的、一部分是没有干预过的, 这两部分后测分数似乎不能加在一起来对待?

回应: 感谢审稿专家的细致评审, 这个问题确实是我们在数据分析中疏忽了, 完全接受审稿专家的意见。我们重新分析了数据并梳理结果, 在新的相关和回归分析中, 低 SES 干预组 40 名被试的数据不在分析之列, 这样就分离出了推理认知干预对后测成绩的影响, 以便对动态测验多项指标的有效性进行验证(修改后的结果见研究二 3.3.1 中的表 4 和表 5, $N = 209$)。

意见 6: 此外, 文中的表达中还存在一些小问题, 建议作者明确。(1) IPDT 动态测验的干预 0 级是什么都不做吗?(2) 为什么表征、分类、规律领域属于“复杂领域”或“更高级领域”? (3) 推理认知的干预就是做练习册吗? 有讲解吗? 此过程可能需要进一步明确说明。(4) 研究二的前后测都是用的瑞文测验, 但文中多处将其直接替换为“认知水平”, 瑞文测验测量的逻辑推理能力似乎不能完全代表总体认知水平?(5) 关于测验鉴别出的“高潜能”, 还是需要有一个更明确的定义。是指现有水平和可达到水平之间的差距较大? 还是未来可达到水平较高? 目前从测验划分标准看, 似乎是前者, 可是从研究二的检验逻辑看, 又似乎是后者。(6) 后测的百分位排名是否非升即降, 因此 50% 的提升只能算是随机, 不能算做提升?(7) 还有个别句子缺少出处。

回应: (1) 第 0 级干预是什么都不做, 审稿专家的提问有道理, 严格地讲, 这不是干预, 确实有些儿童不经提示就能正确作答。之所以叫“0 级干预”, 是为了反映逐级提示范式中整个干预过程中的一个环节, 也是为了表 3 结果中计算和体现各干预等级的通过率和累积通过率。(2) 在全国常模制订中, 10 岁组儿童在守恒、关系、表征、分类、规律 5 个领域的难度系数范围分别是 0.19~0.71、0.59~0.79、0.37~0.81、0.10~0.36、0.15~0.38(方富熹 等, 2005), 相较于守恒和关系领域, 表征、分类与规律领域各子测验的难度系数低(即通过率低)的测题更多(见对第 2 位评审专家第 5 (2) 问题的回答)。从 IPDT 题目内容上看, 表征领域里的运动表征、投影表征和观点表征 3 个子测验, 分类领域的类比推理、类相交、类包含所有子测验, 规律领域的概率、旋转、角度所有子测验也确实更复杂, 因此说是复杂领域或更高级领域。相关内容在前言 1.4 部分的第三段内容中做了概要说明。(3) 有讲解, 上一稿忽略了此问题, 感谢审稿专家的建议, 我们对研究二 3.2.3 程序部分“(4) 推理认知干预”的干预细节进行了补充, 请审阅。(4) 《瑞文推理测验》的编制依据斯皮尔曼的智力二因素论(即

一般因素和特殊因素),是可用于团体施测的非文字智力测验,个体在该测验上的得分代表了其一般智力水平(Raven, 1941)。Raven把G因素划分为两种相互独立的因素,一种称为再生性能力,另一种称为推断性能力。前者与学校教学密切相关,后者则指一个人作出推理判断的能力,由此编制了《瑞文推理测验》。张厚粲等人(1989)在对《瑞文推理测验》进行常模制订时对该测验的效度进行了检验,瑞文推理测验的总成绩与韦氏智力测验总智商分数的相关达0.71($p < 0.01$);他们同样认可瑞文推理测验是对个体一般认知能力的测查。因此,我们在研究二中用瑞文推理成绩作为衡量儿童一般认知能力的指标。这样的思考是否合适,请审稿专家赐教。(5)感谢审稿专家的中肯建议,修改稿中对高潜能儿童的界定进行了说明。根据维果斯基的ZPD概念(即现有水平和潜在发展水平之间的差距)及动态测验的相关理论与研究,干预分、迁移分、后测分和增长分等都被认为是动态测验分数。本研究依据IPDT动态测验的这些分数指标所划分出的高潜能儿童,是指那些干预后进步更大、迁移更好的儿童。我们在文中(见前言1.3部分)对高潜能的含义进行说明,请审稿专家审阅。

(6)本研究中提到的百分位提升,是相对于数学前测,后测成绩的排名进步,不在提升之列的被试或者保持位次不变或者位次下降。表7中对各组的百分位数进行比较,是想说明低SES干预组,尤其是高潜能干预组(60%)比对照组儿童(39%、38%)经推理认知干预后在数学成绩上进步的人数比例更高,意在组间比较。坦白地讲,我们不能肯定后测比前测排名提升的人数比例哪个比值是随机概率,因为如果暂且不管提升还是下降,当人数较少时,所有儿童前、后测保持位次不变也是可能的。因此,是否有随机概率,哪个比值是随机概率可能还不好确定。不过,本研究关注的是不同组之间的比较,从百分位提升的比例来看,有明显的差距,故而从另一个侧面说明高潜能干预组的更大进步。不知我们的考虑是否合适,请赐教。(7)我们对写作中的疏忽表示歉意与惭愧,感谢审稿专家如此细致的评审。修改稿中对全文进行了检查与修改,对缺少出处的地方进行了文献补充。

张厚粲,王晓平.(1989).瑞文标准推理测验在我国的修订.《心理学报》,(2),113-121.

Raven J. C. (1941). Standardization of progressive matrices. *British Journal of Medical Psychology*, 19, 137-150

.....

审稿人2意见:

该研究基于“儿童认知发展水平诊断工具”,编制了测量儿童认知发展水平的动态测验,采用“前测-干预-迁移-后测”的程序,输出动态测验分数,能够对学业成绩和推理能力进行预测。总体来看,研究在儿童评估方面有一定的应用价值,希望在以下方面进行修改

或补充。

意见 1: 建议修改标题，到底是一个干预研究，还是介绍一种测评工具/方式，应该择其一。

回应: 经审稿专家提议，我们对拙作进行了反思，根据研究目的，我们将标题修改为“儿童认知发展水平诊断工具 IPDT 的动态化编制及其在低社会经济地位儿童中的应用”。研究一是对“皮亚杰认知发展量表 (IPDT)”表征、分类与规律领域进行动态测验的改编；研究二通过“前测-干预-后测”的研究设计，将基于 IPDT 动态测验潜能分类的优势应用于对低社会经济地位儿童认知与数学表现的干预之中，考查依据潜能分类后的不同干预效果。

意见 2: 前言部分主要聚焦于社会经济地位与儿童认知发展的关系，有以下两个问题：(1) 1.2 部分，文中提到“但是这些项目的设立和运行需要长期大量的财力、人力和时间投入，涉及部门多、实施难度大。因此，构建一套可操作的、更经济的方案，通过有效区分低 SES 中不同发展潜能的儿童，进而针对不同潜能水平儿童制定领域针对性的干预措施，力图保障有限教育资源的高效利用，是促进低 SES 儿童认知发展和学业改善的可行策略之一。”该部分逻辑不够连贯：为何区分“低 SES 中不同发展潜能的儿童”这一点是文章的核心，因此有待补充论述。(2) 虽然能够理解低 SES 儿童更有可能在认知发展上落后，但 SES 并非是影响认知发展的决定性（唯一）因素，高 SES 儿童同样可能发展落后。既然动态测试能够区分儿童的潜能，为何文章选择只在低 SES 儿童群体中验证？另外，由后文可以看到，研究对高低 SES 儿童的划分基于样本群体的标准分，因此该工具在今后的使用中该标准可能发生变化。

回应: 感谢专家就前言中的逻辑问题提出的细致意见。**关于问题 (1)**，我们对前言 1.2 部分的相关内容重新进行了梳理与阐述（见前言 1.2 最后一段），以回答为何在本研究中对 IPDT 进行动态化编制后，通过 IPDT 动态测验分数所进行的高、低潜能类别划分有助于检验后续干预对不同群体的不同效果。**关于问题 (2)**，根据动态测验的理论与实验，动态测验（维果斯基也称之为“最近发展区的评估”）是针对传统静态测验（指在没有干预和直接帮助下，对儿童现有认知水平的评估，主要指智力测验）对弱势儿童（包括学习无能、心理落后、不良条件和文化背景下的儿童以及新移民儿童等）不公平而提出并开发的测验范式。维果斯基及动态测验的研究者认为传统静态测验存在一定的局限性：如，只关注儿童现有认知水平而未能提供关于学习潜能和未来发展的重要信息；对学业失败最好的预测效果体现在 IQ 处于 50~80 的儿童上；为追求测验公平所采取的统一不提示、不反馈的中立态度不易消除弱势儿

童理解和适应测验的困难等 (Brown & French, 1979)。而动态测验以与智力测验相同或相似的任务为评价工具, 摈弃传统测验中不指导、不干预的静态评估范式考查儿童认知发展潜能, 以便更为公正、客观地评价儿童, 尤其是“弱势儿童”的能力发展 (Tzuriel, 2000)。因为动态测验的兴起是为了客观评估弱势儿童, 并探查其认知发展的最高点; 也因为与高 SES 相伴的较为丰富的教育资源相比, 低 SES 环境中的诸多因素都对儿童认知发展和学业成就产生不利影响 (见前言 1.1 和 1.2 部分)。所以, 本研究意在考查 IPDT 的动态测验对低 SES 不同潜能发展儿童的评估, 一方面关注低 SES 弱势儿童经潜能水平划分后表现出的不同程度的干预进步 (见表 7), 试图证实低 SES 高潜能儿童比低潜能儿童干预后的进步更大、细致分类对针对性干预及其效果的独特意义; 另一方面将中 SES (结合审稿专家的第 7 个关于低、中-高 SES 组划分问题, 我们对被试的 SES 水平参照父/母最高学历进行了进一步核实, 认为称为“中 SES”更为合适, 详见对第 7 个问题的答复) 高潜能儿童作为对照组以考查低 SES 儿童经干预提升后能否赶上同群体发展最好的中 SES 高潜能组儿童。维果斯基及动态测验的倡导者主张, 将干预后的表现或进步及迁移等作为衡量潜能发展 (或最近发展区宽度、认知改变) 的指标, 因此本研究的高、低潜能组划分采用了干预分、迁移分、增长分和后测分 4 个动态测验分数, 并且为了使高、低潜能组差距更大, 尽可能地降低误报, 采取了 4 个分数都居前/后 50% 位次这一较为严格的标准对高、低潜能组进行划分。诚然, 正如审稿专家所言, 高低潜能的划分是相对标准, 中 SES 组儿童也是做如此划分。动态测验对儿童进行高/低潜能的划分不采取绝对标准, 因为根据动态测验的主张, 潜能水平/最近发展区宽度在儿童某个/某些认知领域发展的不同阶段、不同干预阶段和干预方式中是动态变化的, 并且不同儿童在各个认知领域的潜能水平/最近发展区宽度也是不同的, 在学习的不同阶段也会有所不同。由于本研究关注的是低 SES 儿童, 所以被试的选取来自城乡结合部的小学, 就读学生的家庭社会经济地位普遍低于地处城市中心的小学。所以, 分类后的不同 SES 同一类型潜能组之间在 IPDT 动态测验分数上没有显著差异, 但不同潜能组之间有一定差异。

Brown A L, French L A. (1979). *The zone of potential development: Implications for intelligence testing in the year 2000*. In: R J Sternberg, D K Detterman ed. *Human intelligence: Perspectives on its theory and measurement*. Norwood, NJ: Ablex, 217–235.

Tzuriel, D. (2000). Dynamic assessment of young children: Educational and intervention perspectives. *Educational Psychology Review*, 12(4), 385–435.

意见 3: 由于该研究开发的动态测验基于“儿童认知发展水平诊断工具” (Inventory of

Piaget's Developmental Task, IPDT), 前言部分应加入对该原始工具的更详细的介绍, 包括对其中分测验和子测验的介绍。

回应: 抱歉我们对这个问题的考虑过于主观了, 以为有以往相关研究便在拙作中简略甚至忽略了, 信息不完整无疑会使审稿专家及读者产生疑惑。我们感谢并接受审稿专家的建议, 并在前言中增加了对 IPDT 更为详细的介绍(见前言 1.4 部分)。并请参见第 3 位审稿专家的第 1 条意见及回复。

意见 4: 每个研究应该在一开始就明确提出实验设计和明确的假设, 请作者修改补充。

回应: 感谢并接受审稿专家的意见, 我们对研究假设和实验设计做了明确的表述, 分别补充在研究一 2.1 和研究二 3.1 部分。

意见 5: 研究一, 2.3.1 部分, 有以下三个问题: (1) “根据前人对 IPDT 各测题通过率的结果(方富熹 等, 2004)以及本研究前期实验对表征、分类、规律领域的各测题通过率的验证, 确定各阶段测题(见表 1)”——前期实验如何进行, 结果如何? 这部分应该进行报告。(2) 文中提到, “前测题为该子测验内难度中等的题目”(2 个), “干预题是与前测题中至少一道难度相匹配的题目”(1 个), “迁移题是同一子测验中难度较高的题目”(1 个), 既然全部 4 个题目均已用上, 那么为什么只有中等或高难度题目, 低难度题目在哪里呢? 这四个题目各自难度又如何?(按难度值来看的)。(3) 迁移阶段, “干预结束后要求被试独立作答迁移题, 以检验干预后的学习能力是否能解决同领域更高难度的题项。”——比前测题目难度高多少为宜?

回应: 抱歉我们在撰文中只考虑了字数限制而忽略了这些内容。审稿专家的 3 个问题都是针对动态测验编制中的题目设置而提问的, 在此综合回复。对于专家的问题(1), 本研究前期实验以城市普通小学 50 名五年级儿童(女生 25 人, 10.48 ± 0.27 岁)为被试, 对 IPDT 进行施测并计算了表征、分类与规律领域子测验的通过率。结合 IPDT 常模制订中各子测验的通过率结果(方富熹, 盖笑松, 张丽锦, 2005), 通过二比率差异显著性检验发现, 各子测验的难度系数与方富熹等人(2005)的研究无显著差异($ps > 0.05$)。由于表格太大, 考虑到所占版面因素, 我们将这部分内容放到补充材料中, 请评审专家查阅并指正。关于问题(2)(3), 我们是根据 IPDT 各子测验中 4 个测题内部的相对难度, 将各子测验中的测题分配至动态测验的不同阶段。十分抱歉我们在文稿中的表述不全面也不准确, 测题分配的原则应该是这样的: 各子测验的 4 道题中难度最高的定为迁移题; 考虑到干预后的进步需留出提升空

间，前后测题目要尽可能保证有 1 道中-高难度的题目（当然，IPDT 作为评估 7-15 岁儿童的认知发展测验，不同领域的子测验难度对不同年龄/年级儿童差异会比较大。对于本研究的五年级儿童来讲，同一认知领域中有些子测验的难度可能会偏高或偏低，而偏低的子测验在做干预题设置时可能不太理想，即符号表征子测验和水平面表征子测验。但为了保证 IPDT 测验动态化编制的完整性，加之这类子测验数量很少，我们还是保留了这些难度指数对五年级儿童不理想的子测验）；干预题与前后测题目中的 1 道难度相近。例如，观点表征子测验中 4 个测题的难度系数分别为 0.31（25 题）、0.36（26 题）、0.52（27 题）、0.49（28 题），我们就将其中难度最高的 25 题分配至迁移测试阶段，将 26 和 28 题分配至前后测阶段，将 27 题分配至干预阶段。上一稿中确实没有表述清楚，现已对文中相应部分的表述做了修改（见结果部分 2.3.1），请审稿专家指正。

方富熹, 盖笑松, 张丽锦. (2005). 皮亚杰认知发展量表 (IPDT) 中国城市常模的制订. *中国心理卫生杂志*, 19(12), 789-792.

意见 6: 研究二包含了对推理能力干预的考察，请作者对图形推理练习进行更多具体操作说明，并提供这种干预方法能够有效提升认知能力的研究证据。

回应: (1) 根据审稿专家的建议，我们在研究二中对推理认知干预的过程进行了具体说明（见 3.2.3 程序中“（4）推理认知的干预”部分），并在讨论中论述了支持证据。(2) 本研究采用的瑞文图形推理测验、皮亚杰认知发展水平诊断工具 IPDT 与自编的图形推理认知干预都有一些共同特点：都是图形形式的测验、都关乎推理认知能力的考查。瑞文推理测验（张厚粲, 王晓平, 1989）、皮亚杰测验任务的纸笔形式（即 IPDT）都可以作为考查个体一般认知能力的测验，本研究参照多种图形推理测验（见研究二 3.2.2 (5)）所设计的 24 册图形推理练习册也与儿童的认知能力密切相关。而且，推理认知干预对个体认知能力的促进作用也被其他研究所证实。例如，鉴于存在聋生的认知推理能力不及健听生的现象，研究者采用单一被试多次干预的方法考查了序列推理（如，图形推理、空间次序、动作序列、符号推理、数字推理等）干预对认知能力的促进作用，（与本研究相关的）结果发现，图形推理干预对四年级聋儿（邱天龙 等, 2017）和对学前幼儿（陈彦, 杜晓新, 黄昭鸣, 2009）的认知能力均有显著的促进效果。张登科等人（2012）对 60 名脑外伤患者进行为期 12 周（每周一次）的心理推理能力的干预，发现干预组在心理推理任务和心理推理他评量表上的后测得分均显著高于对照组（ $F(4, 87) = 25.36, p < 0.001$ ）。此外，本研究采用的非标准化干预方法对儿童认知能力（如，执行功能）的促进效果也得到了证实。例如，Röhliberger 等人（2011）采用前

测-干预-后测设计, 通过有难度梯度的系列游戏(以小组为单位由幼儿教师在日常教学中进行干预)对幼儿执行功能进行干预, 在控制前测分数后发现干预组幼儿的多项认知后测分数均显著高于控制组 ($F(1, 67) = 3.0\sim 12.8, ps < 0.05$), 证实了非标准化干预方法的有效性。

邱天龙, 杜晓新, 郭雯, 刘琳, 刘伟, 杨集梅, 蔡勇刚. (2017). 四年级聋生序列推理能力特征及干预研究. *中国特殊教育*, (01), 53-57.

陈彦, 杜晓新, 黄昭鸣, (2009). 听障儿童五项认知能力评估与训练的个案研究. *听力学及言语疾病杂志*, 17(2), 183-184.

张登科, 徐水琴, 苏巧荣, 潘一楠, 沈雪艳, 陈三妹, 葛建荣, 单兴尧. (2012). 脑外伤患者心理推理能力缺陷的认知性和技巧性干预: 一项 12 周随机单盲对照试验. *中国心理卫生杂志*, 26(12), 906-912.

张厚粲, 王晓平. (1989). 瑞文标准推理测验在我国的修订. *心理学报*, (2), 113-121.

Röhlisberger, M., Neuenschwander, R., Cimeli, P., Michel, E., & Roebbers, C. M. (2012). Improving executive functions in 5- and 6-year-olds: Evaluation of a small group intervention in prekindergarten and kindergarten children. *Infant & Child Development*, 21(4), 411-429.

意见 7: 研究二, 关于 SES 的计算, 有以下两个问题: (1) SES 的计算中, “运用项目反应理论对家庭资产和教育资源进行参数估计, 作为家庭收入的计数指标。”——如何进行参数估计, 此处应进行更详细的介绍。另外, SES 计算公式的来源应进行引用标注。(2) 研究采用了 PISA 中计算 SES 的方法, 并将 SES 分数高于平均数 0.5 个标准差的被试作为中 SES 组, 低于 0.5 个标准差的作为低 SES 组。对此的疑问是, PISA 以国家为单位, 计算 SES 的时候各变量取值范围广, 变异大, 也更能反映出 SES 绝对水平的高低差异。但在本研究中, 仅采用了银川城乡结合部 2 所小学的五年级学生家庭, 是否会由于取样缺乏代表性, 导致所有被试的 SES 绝对水平都偏低, 在这种情况下区分出的相对 SES 水平, 是否能推广到一般儿童?

回应: (1) 我们参照 PISA (OECD, 2003) 以及方晓义教授团队(方晓义, 范兴华, 刘扬, 2008; 范兴华, 方晓义, 刘扬, 蔺秀云, 袁晓娇, 2012) 的研究中对 SES 的计算方法(文中加注了参考文献), 采用 IRTPRO (Scientific Software International Inc., USA) 软件对每一个被试的家庭资源进行参数估计。具体参数如下: Comments: Two-parameter logistic for each item Estimation: Method = BAEM; E-Step = 500, 1e-005; SE = S-EM; M-Step = 50, 1e-006; Quadrature = 49, 6; SEM = 0.001; SS = 1e-005; Scoring: Pattern = EAP; Score Persons; Mean = 0; SD = 1。这些参数估计的具体参数和内容, 以往研究在正文中都不报告, 本研究沿用了以往

的方式。(2)项目反应理论根据试题的难度和学生的反应用数学概率模型来预测某生答对题的概率,从而估计学生个体的能力或心理特质。经济地位与学生能力有类似特征,都是潜在特质,通过学生对某个选项的选择概率的计算,可以获得经济地位的估计值。项目反应理论的优点在于能够将不同样本群体在一个相同的尺度上标定位置(即本研究中对教育水平、职业、家庭资源数据的标准化处理),基本不受样本分布的影响,因此能够对不同样本群体经济地位做等值运算,还可以开展不同研究之间的比较(见任春荣,2010)。本研究中分数标准化并合成后的SES分数与以往研究一样,近似标准正态分布(平均数为-0.003,标准差为0.996)。调研以往相关研究,对高、低SES儿童分组(特别是低SES)的研究(行为或脑电)大致有以下做法:(1)直接从位于不同地区的学校选取,即当地的高SES学校和低SES学校,并报告了两者的SES分数(e.g., Kishiyama, Boyce, Jimenez, Perry, Knight, 2008; D'Angiulli, Herdman, Stapells, Hertzman, 2008; Jordan, Huttenlocher, Levine, 1992);(2)根据母亲教育水平(大学以上或以下)把儿童直接分为两组,即用母亲受教育水平作为SES的指标(e.g., Stevens, Lauinger, Neville, 2009);(3)直接选取开端计划或当地扶贫计划的儿童(Siegler, Ramani, 2008; Wilson, Dehaene, Dubois, Fayol, 2009);(4)编制若干考查家庭经济状况的题目,低于某分数的为低SES组,其他为中或高SES组。例如,冯晓霞教授团队(肖树娟,冯晓霞,成丽媛,苍翠,2009;陈敏倩,冯晓霞,肖树娟,苍翠,2009)采用SES分数分组的研究在同一地区(北京)6所学校选取被试(143人),然后按照SES的相对高低,将被试进行高、中、低组划分。本研究关注的是低社会经济地位儿童的认知改变,为了保证这部分被试筛选的正确击中率,选取了宁夏回族自治区银川市城乡结合部的小学儿童。因为被试来自城乡结合部的小学,为了使中、低SES组划分更符合所属组的经济水平,采用了上下0.5个标准差的筛选标准,将这个标准之内的被试剔除,这是为了尽可能地保证两组,尤其是低SES组的正确击中。作为来自西北地区城市城乡结合部学校的被试,即便SES处于上0.5个标准差的儿童,也有可能如审稿专家所说,其家庭经济水平相对于其他地域的儿童来说可能仍然较低,所以我们在修改稿中将这一组改为中SES组。即便有极个别儿童家境富裕,但就整个群体来说,更符合中等经济条件特征。有研究者认为,相对于家庭收入的隐蔽性与不稳定性,父母受教育水平与儿童的认知能力呈稳定相关(e.g., Noble, McCandliss, & Farah, 2007; Stevens, Lauinger, & Neville, 2009)。为了检验原有分组的有效性,我们比较了本研究中不同SES儿童父母的受教育年限。发现低SES组父母平均接受了初中教育($M = 8.37$),中SES组父母平均为高中文化水平($M = 12.83$);同时,中SES组儿童的父母中有35.3%接受过大专或本科教育,而这一比例在低SES儿童中仅为1.5%(具体比例见表a)。我们上述考虑是

否合适，请审稿专家指正。

表 a 不同 SES 组别儿童父母受教育年限（父或母较高一方）的人数比例（%）

	<i>n</i>	0 年	5 年	8 年	11 年	14 年	15 年
低 SES	128	2.3	13.1	54.6	28.5	1.5	0
中 SES	121	0	0	8.4	56.3	28.6	6.7

范兴华, 方晓义, 刘杨, 蔺秀云, 袁晓娇. (2012). 流动儿童歧视知觉与社会文化适应: 社会支持和社会认同的作用. *心理学报*, 44, 647-663.

方晓义, 范兴华, 刘杨. (2008). 应对方式在流动儿童歧视知觉与孤独情绪关系上的调节作用. *心理发展与教育*, 24, 93-99.

任春荣. (2010). 学生家庭社会经济地位 (SES) 的测量技术. *教育学报*, 6(5), 77-82.

肖树娟, 冯晓霞, 成丽媛, 苍翠. (2009). 不同社会经济地位家庭儿童的入学数学准备状况比较. *学前教育研究*, (3), 3-13.

陈敏倩, 冯晓霞, 肖树娟, 苍翠. (2009). 不同社会经济地位家庭儿童的入学语言准备状况比较. *学前教育研究*, (4), 3-18.

D'Angiulli, A., Herdman, A., Stapells, D., & Hertzman, C. (2008). Children's event-related potentials of auditory selective attention vary with their socioeconomic status. *Neuropsychology*, 22, 293-300.

Jordan, N. C., Huttenlocher, J., & Levine, S. C. (1992). Differential calculation abilities in young children from middle- and low-income families. *Developmental Psychology*, 28, 644-653.

Kishiyama, M. M., Boyce, W. T., Jimenez, A. M., Perry, L. M., & Knight, R. T. (2008). Socioeconomic disparities affect prefrontal function in children. *Neuropsychology*, 21, 1106-1115.

Noble, K., McCandliss, B., & Farah, M. (2007). Socioeconomic gradients predict individual differences in neurocognitive abilities. *Developmental Science*, 10, 464-480.

OECD. (2003). *The PISA 2003 technical report: Scaling procedures and construct validation of context questionnaire data*. Paris, France: OECD Publishing.

Siegler, R. S., & Ramani, G. B. (2008). Playing linear numerical board games promotes low-income children's numerical development. *Developmental Science*, 11, 655-661.

Stevens, C., Lauinger, B., Neville, H. (2009). Differences in the neural mechanisms of selective attention in children from different socioeconomic backgrounds: An event-related brain potential study. *Developmental Science*, 12(4), 634-646.

Wilson, A. J., Dehaene, S., Dubois, O., & Fayol, M. (2009). Effects of an adaptive game intervention on accessing number sense in low-socioeconomic-status kindergarten children. *Mind Brain and Education*, 3(4), 224–234.

意见 8: 研究二，3.2.2 部分，使用的各种测验应报告基于本研究数据的信度。此外，在两个研究中，请报告 IPDT 前测、后测分数、干预分、迁移分、增长分的信度。

回应: (1) 本研究 IPDT 前测分数的内部一致性信度为 0.845，方富熹等人（方富熹，盖笑松，龚少英，刘国雄，2004）对 IPDT 信效度检验发现，5 年级儿童的内部一致性为 0.77 ($n = 130$)（在研究二的 3.2.2 工具部分补充了该信度系数）；(2) IPDT 动态测验中涉及对被试的干预，干预效果在不同（组）儿童身上是不同的（这也恰恰是本研究所探查的内容），这样所得到的动态测验分数（即干预分、迁移分、后测分和增长分）上测题两两之间的高相关（即 Cronbach's α 系数）不是动态测验研究所追求的，对内部一致性程度的考查并不适用于干预研究。

意见 9: 研究二，3.2.3 部分，有以下三个问题：(1) 干预组和对照组的具体人数、性别等人口学信息以及各组在各分数指标上的平均分和标准差均应进行报告。(2) 推理认知干预部分，是否为双盲测试，如是，需说明。(3) “对于中-高 SES 儿童，以上述同样的方法筛选出高潜能儿童 41 人”，高 SES 组为何没有纳入低潜能组？与之相关，3.3.2 部分，文中提到，“因此假设，相对于低潜能组儿童，由 IPDT 动态测验筛选出的高潜能儿童经相关认知训练后将表现出更为明显的认知和学业提升，可以在一定程度上缩小与中-高 SES 高潜能儿童的发展差距。”——为什么不一起比较高 SES 低潜能的？作者应对这一点进行解释。

回应: (1) 上一稿我们将各组人数信息放置在表 7 中确实不容易让读者快速加工，而且性别样本量信息也不好呈现，感谢并接受审稿专家的建议，修改稿中将各干预组和对照组的人数、性别信息补充至研究二 3.2.3 程序 (2) 中。各组被试瑞文和数学前后测的平均分与标准差在表 7 中进行了报告（见研究二结果部分的 3.3.2）。请审稿专家审阅。(2) 动态测验强调在干预中施测、在过程中评估，动态干预的目标是使学习者或被测者更好地掌握，力图实现最好的学习效果。比如著名以色列教育心理学家 Reuven Feuerstein 以“中介学习理论”解释动态测验的干预促进过程：教育者（父母、教师、年长儿童）介于学习者和知识任务之间，通过设计、选择、集中、反馈、创造合适的学习环境等方式对学习者的训练，最终实现学习者的结构化认知改变和不同程度的迁移。这一过程中，教育者就是“中介者”（mediator），教授过程就是“中介着”（mediate）的过程（R. Feuerstein, & S. Feuerstein, 1979）。所以，

干预过程是以学会和掌握为目的，对单盲和双盲不做更多要求。而本研究的前后测是瑞文推理纸笔测验和数学纸笔测验，均为团体施测。尽管所测内容相似，但动态测验的干预施测与一般意义的认知实验和传统的静态测验有着不同的理论基础、目的过程、测验关系和结果解释，这些反映了不同的测量理念和测量哲学（Grigorenko & Sternberg, 1998）。（3）本研究中 SES 高潜能组是作为对照组存在的，目的是考查低 SES 不同潜能儿童经过干预后与中 SES 高潜能儿童相比，是否能赶上甚至超过那些处在家庭教育资源较好环境，且自身认知潜能较高、发展较好的儿童。这个问题与审稿专家第 2 个问题的（2）前半部分的问题观点相似，请参见上面对“2（2）问题”的回复。

Feuerstein, R., & Feuerstein, S. (1991). *Mediated learning experience: A theoretical review*. In R. Feuerstein, P. S. Klein & A. Tannenbaum (Eds.), *Mediated learning experience (MLE)* (pp. 3-52). London: Freund.

Grigorenko, E., & Sternberg, R. (1998). Dynamic testing. *Psychological Bulletin*, 124, 75–111.

意见 10: 研究二，3.3.1 部分，“由于动态测验在评估儿童认知发展现有水平的同时，还可以考查其潜在水平，因此可以推测动态测验分数对儿童的认知水平和数学成就的预测力更强。”——请作者说明为什么可以考察潜在水平，因此对认知水平和数学成就的预测力更强，其中的逻辑是什么？

回应: 对儿童认知潜能评估的动态测验最早来自维果斯基“最近发展区”的理论与实验，针对传统能力测验视能力为相对固定的、注重儿童现有认知水平的观点，动态测验采用与智力测验相同或相似的项目作为评价工具，摒弃传统能力测验中不指导、不干预的静态方法来考查儿童认知发展潜能，通过干预探查儿童认知发展的最高点，以便更公正、客观地评价儿童，尤其是“弱势儿童”的能力发展。维果斯基及其早期动态测验的研究者（e.g., Vygotsky, 1978; Brown, 1979, 1985; Guthek, 1992）以学习速度、迁移能力、体现干预进步的认知改变力为学习潜能或 ZPD 宽度的指标。随着动态测验领域研究的不断推进，更客观、更可靠的评估方法不断产生，综合起来有：（1）测验者指导下认知成绩的最大改善；（2）前后测之间因教育干预带来的增长；（3）由逐步提示法所测的达到任务要求的提示量；（4）控制认知技能的元认知策略的伸展等。无论是先前的动态测验还是其后发展出的变式“干预反应范式（Response to Intervention, RtI）”和“心灵工具（Tools of the Mind, TotM）”课程，其测量范式都是“前测-干预-(后测-干预-……) 后测”。可见，动态测验分数不仅涉及静态测验的前测分，更重要的是包含了考量干预后进步的增长分、体现对干预敏感性的干预分、考虑初始水平（即前测）和增长进步的后测分、考量应用性的迁移分，这些动态测验分数包含了

更多的潜能发展信息，关注儿童的结构化认知改变。因此，动态测验的理论及多年的实验研究一直反复证明了动态测验分数比传统静态测验分数对儿童能力和学业成就有着更好的预测（Grigorenko & Sternberg, 1998）。关于动态测验的相关理论与研究成果以往已有研究进行了介绍（e.g., Allal & Ducrey, 2000; Grigorenko & Sternberg, 1998; 张丽锦, 盖笑松, 方富熹, 方格, 2003），因此本文只做了简要概述。确如审稿专家所指，一篇研究报告文献综述的完整、完备与逻辑性不能忽略，尤其是核心观点，因此，修改稿中对这部分内容予以适当丰富和完善，请审稿专家指正。

张丽锦, 盖笑松, 方富熹, 方格. (2003). 儿童认知发展动态测验. *心理科学进展*, 11(6), 104-160.

Allal, L., & Ducrey, G. P. (2000). Assessment of--or--in the zone of proximal development. *Learning and Instruction*, 10, 137-152.

Brown, A. L., & French, L. A. (1979). The zone of potential development: Implications for intelligence testing in the year 2000. In R. J. Sternberg, D. K. Detterman (Ed.), *Human intelligence: Perspectives on its theory and measurement*. Norwood, NJ: Ablex, 217-235.

Brown, A. L., & Ferrara, R. (1985). Diagnosing zone of proximal development. In J. V. Wertsch (Ed.), *Culture, communication, and cognition: Vygotskian perspectives*. New York: Cambridge University Press, 273-305.

Grigorenko, E., & Sternberg, R. (1998). Dynamic testing. *Psychological Bulletin*, 124, 75-111.

Guthke, J., & Wingefeld, S. (1992). The learning Tests concept: Origins, state of art, and trends. In H. C. Haywood, D. Tzuriel (Ed.), *Interactive testing*. New York: Springer-Verlag, 64-93.

Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

意见 11：研究二，3.3.1 部分的结果报告中：（1）表 4 的相关矩阵应该呈现所有变量，例如前测瑞文、前测数学。（2）回归采用的是何种具体的回归方式？应对包含所有自变量的完整回归结果进行报告。如果要证明动态测验各分数的价值，应该采用多重回归，并分别报告 R^2 变化值和显著性，证明其增益效度。另外，在预测源变量中应加入前测的瑞文成绩，否则无法排除变化是来自于推理能力的差异。

回应：（1）接受审稿专家的建议，修改稿表 4（研究二 3.3.1 部分）的相关分析中纳入了瑞文和数学前测的数据。（2）本文对瑞文与数学后测分数的预测分别采用阶层回归，将瑞文与数学前测分作为 Step 1 中的预测变量，Step 2 中纳入 IPDT 的前测分，Step 3 中加入 IPDT 动态测验指标（干预分、迁移分和后测分）作为预测变量，重新进行了统计分析（见研究二 3.3.1 的表 5）。

意见 12: 研究二, 3.3.2 部分, “对不同组别儿童的瑞文和数学前测成绩分别进行比较”——不应采用单因素方差分析+5 个 t 检验, 这样会增大一类错误, 而是应该采用 2*5 重复测量方差分析, 并检验简单效应。

回应: 本研究采用的是协方差分析的方法来探查各组间的差异。根据朱滢(2000)在《实验心理学》中对随机实验组控制组前测后测设计的显著性检验的说明(见书中第 20~21 页), 对这种实验设计的数据统计方法有两种: 一种是对增值分数(后测-前测)进行显著性检验; 另一种是以前测分数为协变量, 对不同组别被试的后测分数进行协方差分析, 这样可以使组间的后测成绩比较不受前测成绩的影响。我们对五组的后测成绩进行差异分析时是把前测分作为了协变量, 十分抱歉我们前一稿中的表述错误, 尽管做的是后测组间比较, 但不能叫“单因素方差分析”或“单因素协方差分析”, 而应该是“协方差分析”。修改稿中已更正。

朱滢. (2000). *实验心理学*. 北京: 北京大学出版社.

.....

审稿人 3 意见:

作者的目的是似乎包括两个方面, 一是发展小学高年级的“儿童认知发展水平诊断工具”, 二是利用这一工具评估一方案(推理认知干预)干预学生认知与数学学业的发展。

意见 1: 在第一个关于工具的研究中, 没有对工具的研制过程、维度分类的方法、测验以后等因素加以说明。例如, 没有子测验任务的举例说明、维度来源及其依据、数据收集后的信度、效度等信息。

回应: 第 2 位审稿专家的第 3 条意见也提到了类似的问题, 感谢并接受审稿专家提出的意见。我们在前言中增加了对 IPDT 相关内容的介绍(见前言 1.4 部分), 测题示例见附录 1。在研究二中报告了 IPDT 前测分数上的内部一致性信度(见研究二 3.2.2 部分)。动态测验是基于维果斯基“最近发展区”理论而发展出的一种测验范式, 动态测验中测验者积极参与到测验中进行提示、指导和反馈, 用儿童对这些干预的反应来表征认知改变能力、学习潜能、学习熟练和迁移能力。这是一种将测验和干预相结合, 测量学习过程和结果的范式(Grigorenko & Sternberg, 1998)。研究一是对皮亚杰认知发展量表(IPDT)中表征、分类与规律领域的动态化改编, 前人研究中 IPDT 的内部一致性信度为 0.77 ($n = 130$), 本研究中为 0.845。前人对 IPDT 效度检验证实, IPDT 有良好的同时效度(IPDT 纸笔测验与皮亚杰实物操作任务的相关为 0.73, $p < 0.001$)和构念效度(1、3、5 年级儿童在 IPDT 分数上的差异显著, $F(2,$

387) = 208.94, $p < 0.001$) (方富熹, 盖笑松, 龚少英, 刘国雄, 2004)。相关内容我们已补充至文稿的相应部分, 请审稿专家审核并指正。

方富熹, 盖笑松, 龚少英, 刘国雄. (2004). 对儿童认知发展水平诊断工具 IPDT 的信度效度检验. *心理学报*, 36, 96-102.

Grigorenko, E., & Sternberg, R. (1998). Dynamic testing. *Psychological Bulletin*, 124, 75-111.

意见 2: 在第二个研究中, 干预研究的价值很大, 但是样本太少, 一个条件只有 20 人左右。建议增大每一小组的被试量。

回应: 答: 感谢专家对干预研究意义的肯定。确如审稿专家所言, 本研究低 SES 各组样本量有些偏少, 这是本研究的一个局限。从文中所描述的程序可以看出, 动态干预程序繁琐且耗时长, 增加被试量确有难度。研究二以“前测-干预-迁移-后测”的程序对儿童进行一对一的施测, 认知潜能评估后再进行分组(为使高、低潜能组差异明显, 我们采用了 4 个动态测验分数都处于上或下 50%位置的被试, 这个过程被试量损耗比较大), 之后再对干预组被试进行为期 8 周的推理认知干预, 整个过程繁琐耗时, 无疑使增大被试量的工作变得比较困难。正是由于动态测验的繁琐程序, 以往一些干预研究的样本量也不够大。例如, Holmes 等人(2009)对儿童工作记忆的训练中干预组 22 人、对照组 20 人; Alloway 等人(2009)对学习障碍儿童的工作记忆进行训练, 干预组 8 人, 对照组 7 人; Shavelson 等人(2008)采用计算机认知训练促进儿童认知能力, 干预组 18 人, 对照组 19 人。另外, 目前来看, 研究二采用协方差分析考查不同潜能儿童干预后的能力提升, 根据相关要求, 运用 G*Power 3.1 软件, 保证 Effect size = 0.35 的前提下, 设定 $\alpha = 0.05$, power = 0.85 时, 计算出所需最小样本量为 115 人。本研究中用于协方差分析的样本量是 125, 符合要求, 样本量的有效性保证尚可接受。当然, 如审稿专家所指出的, 增加样本量确实可以得到更为可靠的结果, 未来研究确实需要在这方面予以改进。

Alloway, T. P., & Alloway, R. G. (2009). The efficacy of working memory training in improving crystallized intelligence. *Nature Proceedings*. Retrieved from <http://proceedings.nature.com/documents/3697/version/1/files/npre20093697-1.pdf>

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, 12, 9-15.

Shavelson, R. J., Yuan, K., Alonzo, A. C., Klingberg, T., & Andersson, M. (2008). *On the impact of computerized cognitive training on working memory and fluid intelligence*. In D. C. Berliner & H. Kuppermintz (Eds.),

意见 3: 需要说明干预的效应量（减去了控制组的）。从表 7 中效果不是很明显。“低 SES 的两个干预组儿童的数学成绩增长分（ $\Delta_{\text{高潜能}} = 7.15, \Delta_{\text{低潜能}} = 6.59$ ）高于两个对照组（ $\Delta_{\text{高潜能}} = 4.91, \Delta_{\text{低潜能}} = 4.61$ ）”，对于上面的结果，在表 7 中没有。这些结果应该都是关键的结果。

回应: 感谢审稿专家的意见。通过查阅资料（朱滢, 2000）发现，虽然我们以前测分数为协变量，进行了不同组别儿童后测分数的协方差分析，但成绩增量的差异性检验能够从另一个角度描述干预的效果。因此，根据评审专家的意见，我们在研究二的结果中增加了对不同组别儿童瑞文和数学成绩**增长分**（后测分数－前测分数）的进一步分析（以组别为自变量、以增长分为因变量的单因素 ANOVA）（见 3.3.2 部分的表 7）。结果表明，瑞文增长分的组别差异显著， $F(4, 120) = 8.01, p < 0.001, \eta_p^2 = 0.211$ ，多重比较发现，低 SES 高潜能干预组儿童的瑞文增长分显著高于其对照组（ $p < 0.001$ ），但与低 SES 低潜能干预组儿童的差异不显著（ $p > 0.05$ ）；而低 SES 低潜能干预组儿童的瑞文增长分与其对照组无显著差异（ $p > 0.05$ ）。数学增长分的组别差异不显著， $F(4, 120) = 2.13, p > 0.05, \eta_p^2 = 0.066$ ，故而又做了各组数学百分位数人数提升的比较。

朱滢. (2000). *实验心理学*. 北京: 北京大学出版社

第二轮

审稿人 1 意见:

意见 1: 在 4.3 部分第二段的讨论中（可能也包括前面的一些阐述），对于“干预效果”和“动态测验的预测效果”有部分混淆。具体表现为，在这一段的前半部分阐述高潜能儿童接受干预后有提升，低潜能儿童经过干预训练也得到了一定程度的提升，而未接受干预的对照组提升程度明显较少。这一结果所说明的，应该是干预训练的作用明显，但后面一句却解释为“这些结果说明由动态测验区分的不同认知潜能儿童，其发展的可能性和发展空间确有不同”，然后后面又回到了关于干预训练有效性的描述，然后再接着回到讨论划分潜能水平的必要性。这样的讨论显得思路十分跳跃。建议作者进一步明确，“干预组的提升高于对照组”这一结果只能说明研究所采用的认知干预的效果较好，其后的讨论应围绕干预有效性本身展开，不宜区分不同水平的儿童。而区分不同潜能儿童并给予其选择性干预的问题，是建立在

“高潜能儿童经干预后提升高于低潜能儿童经干预后提升程度”这一结果的。两个方面分开讨论会比较清晰。

回应：非常感谢审稿专家对讨论部分提出的指导性建议，根据该意见，已对 4.3 部分的行文逻辑进行了重新梳理与修改，讨论中增加了“4.4 低 SES 不同潜能儿童从推理认知干预中的获益不同”部分，相应的修改见文中棕色字体。

意见 2：关于上一次意见中提到的“后测中排名提升 50%”的说法，是针对 4.3 部分第二段中提到的“对于低 SES 低潜能儿童，经过推理认知训练也得到了一定程度的提升，其数学后测的百分位排名也有相当比例（50%）的提升”这一句所说的。由于排名非升即降，那么这些儿童中有 50%是提升，必然有 50%是下降，那么对于这个群体（低 SES 低潜能儿童）来说，认知训练真的有促进作用吗？既然一半孩子上升，一半孩子下降，是否可以认为干预作用不明显，甚至是随机的？

回应：我们仔细揣摩、认真分析了审稿专家的意见，尝试做如下回答：首先，我们整理分析了所有被试的排名，将不同组别儿童数学前后测成绩的百分位变化的具体情况做如下整理（见表 a）。可以看到，对每一组儿童，其数学前后测成绩排名除了发生提升与下降之外，还有持平的情况，即前、后测成绩排名顺序并没有发生变化。审稿专家的意见提醒我们进一步思考这个问题，但究竟是 50%算作随机概率，还是 33%算作随机概率（考虑到还有持平的情况），我们确实没有考虑成熟。结合审稿专家第一轮评审中提出的第（6）个问题及我们的回答，在本研究中我们关注的是不同组之间的比较，从表 a 中可以看到，高潜能干预组的百分位提升的比率是最高的，两个干预组的提升比率明显高于两个对照组，这从另一个角度证实了高潜能干预组的更大进步（见讨论 4.4 部分第 1 段）。不知我们这样的思考是否合适，敬请审稿专家指正。

表 a. 不同组别儿童数学前后测成绩百分位排名变化情况

	高潜能干预组 (n = 20)	高潜能对照组 (n = 23)	低潜能干预组 (n = 20)	低潜能对照组 (n = 21)	中 SES 控制组 (n = 41)
提升	12 (60%)	9 (39%)	10 (50%)	8 (38%)	22 (54%)
持平	3 (15%)	1 (43%)	2 (10%)	4 (19%)	4 (10%)
下降	5 (25%)	13 (57%)	8 (40%)	9 (43%)	15 (36%)

审稿人 2 意见：

意见 1：文章新标题的重点是“IPDT 的动态化编制”，然而前言开篇 1.1，1.2 大段描写低 SES 的相关问题，显得本末倒置。尤其 1.2 的倒数第 12 行出现了动态测验，此前却对此毫无介绍，可能会增大读者的阅读难度。建议从测验本身开始讲，之后再过渡到动态测验的价值——低 SES 群体的应用。

回应：非常感谢审稿专家对前言部分的逻辑框架提出的中肯建议。在一审的修改中我们在修改标题后确实忽略了标题与前言的逻辑对应关系。根据该意见，我们对前言部分的逻辑框架进行了重新梳理和修改。由于主要是对前言框架和相应内容的顺序进行调整，所以只将前言部分的二级标题标注为棕色字体。

意见 2：在 1.4 中，作者提到“传统智力测验仅在成人智力模型的基础上笼统评估儿童智力（Furth, 1970）”，事实上，目前的主流智力测验，例如 WISC，并不是在成人智力模型的基础上来发展儿童智力模型，相反，是先发展儿童智力模型，再应用到其它群体，故此描述略为不妥，建议删除。

回应：的确，我们在上一稿的表述不够严谨准确，接受审稿专家的建议对此予以删除。

意见 3：对于 2.3.3，根据作者的计分逻辑，通过率必然随干预等级的逐级提升而增高（也就是说即使干预方案有问题，例如前后顺序颠倒，也不可能出现逐级降低的情况），且当提供第 V 级提示时，已经给出了答案，必然是所有被试均通过。故此，从逻辑上看，此步骤恐无法成为检验干预步骤有效的证据，建议删除这一部分内容，或者从其他角度考虑检验有效性的方法。

回应：我们认真思考了审稿专家的建议，对该问题作如下回复：维果斯基及其同事最初针对智力落后儿童进行最近发展区评估时，使用的方法是将训练分成一系列小步骤，并设计好顺序进行干预，但他们并没有从干预量的角度对儿童的最近发展区（或潜能水平）进行量化评定。上世纪 60~70 年代以色列著名儿童心理学家 R. Feuerstein 开发了评估认知潜能的测验——“工具强化方案”（Instrumental Enrichment, IE）和“学习潜能评估工具”（Learning Potential Assessment Device, LPAD）并在以色列广泛使用。Feuerstein 的理论与测量研究激起了更为广泛的针对动态测验研究方法的探索；但对其的批评意见也很中肯——除了时间和技术上的高投入，主要批评临床法干预的标准化过低，“提示”描述不够详尽，只是“成人帮

助”的笼统提法（见 Grigorenko & Sternberg, 1998; 张丽锦 等, 2003）。后来，这个问题在 Campione 和 Brown（1987）开发的逐步提示技术中得到了一定程度的解决。他们采用**逐级提示直到问题解决的干预策略**，以“前测-提示-后测-提示性后测……”的程序，测量儿童潜能水平。与其他动态测验强调最大程度地改善作业成绩不同，该方法关注的是达到某一成绩的最小帮助量，将干预的步骤量化和标准化，用对应于干预步骤的提示量表示 ZPD 的宽度或学习潜能的水平，使 ZPD 的测量有了可操作性（Ferrara et al., 1986）。借鉴这一方法，本研究中的干预步骤本着由抽象到具体、层级式指导的逻辑，获得不同儿童达到掌握水平所需的提示量。结果 2.3.3 中的表 3 是为了说明我们所设计的这 5 级提示的干预步骤对被试具有**逐级提示**的作用。从各干预等级的通过人数可以看到，被试对从抽象到具体的干预步骤的敏感性在逐步提升，直至最终通过。一旦证明了所设计的干预步骤的逐级性，就可以对干预组被试所接受的提示量进行赋分（从第 0 级到第 V 级依次赋分 5 至 0 分，见研究二方法 3.2.2 中的“（2）改编的 IPDT 表征、分类、规律领域的动态测验”部分），最后得到“干预分”这一动态测验分数指标。为了更清晰地表述我们呈现表 3 的意图（即证明干预步骤具有逐级提示性），我们将研究一结果 2.3.3 的标题改为“逐级干预步骤的有效性检验”。此外，我们对这部分的文字表述进行了适当修改，见棕色字体部分。

Campione, J. C., & Brown, A. L. (1987). Linking dynamic testing with school achievement. In C. S. Lidz (Ed.),

Dynamic testing (pp. 82–115). New York: Guilford Press.

Ferrara, R. A., Brown, A. L., & Campione, J. C. (1986). Children's learning and transfer of inductive reasoning

rules: Studies of proximal development. *Child Development*, 57, 1087–1099.

Grigorenko, E., & Sternberg, R. (1998). Dynamic testing. *Psychological Bulletin*, 124, 75–111.

张丽锦, 盖笑松, 方富熹, 方格. (2003). 儿童认知发展动态测验. *心理科学进展*, 11(6), 104–160.

意见 4: 对于 3.2.1，为什么设定“Effect size = 0.35, power = 0.85”，请附上参考文献。

回应: 我们在对一审中审稿人 3 提出的样本量问题的回复与修改时，采用了 G*power 对样本量进行了估计。但根据本研究的实际情况，我们未在实验前进行样本量估计，因此用这个方法在事后去验证样本量是有失妥当的。本研究应进行事后的统计功效检验（而非进行先验分析），以验证样本量是否能够达到符合要求的统计功效，进而说明样本量的合理性（Faul, Erdfelder, Lang, & Buchner, 2007）。参考冉雅璇等人（2020）研究中的参数设置（效应量为 0.4、显著性水平为 0.05），我们采用 G*power 3.1 进行了 Post hoc 统计功效检验，相关内容在研究二 3.2.1 部分进行了更改（见棕色字体），请审稿专家审阅。

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191.

冉雅璇, 刘佳妮, 张逸石, 卫海英. (2020). “一”人代言的魅力: 品牌代言人数如何影响消费者的品牌态度. *心理学报*, 52(3), 371–385.

意见 5: 对于一审中的意见, “3.2.2 部分, 使用的各种测验应报告基于本研究数据的信度”, 作者并没有增添, 例如瑞文渐进矩阵推理测验。再如“数学成就测验”, 作者例举出了多种信度指标, 但难以判断是当前研究数据分析的结果, 还是原量表报告的数据。

回应: 首先我们为一审意见回复过程中的粗心大意进行自我批评。数学成就测验中的信度指标为前人研究中的数据, 本研究中各量表的信度指标在研究二 3.2.2 部分进行了补充报告(见棕色字体)。

第三轮

审稿人 1 意见: 修改稿较好回答了之前的问题, 我没有其它问题了。

回应: 感谢审稿专家的评审和认可。

审稿人 2 意见: 作者对问题进行了回应, 基本解决了审稿人的疑惑, 少量不同意见可视为正常的学术观点差异。另外, 表 2 中“讲解 1 道错误选项”建议改为“讲解 1 个错误选项”。

回应: 感谢审稿专家的评审和认可。接受审稿专家的意见并在表 2 中做了修改。

编委意见:

该研究证实其构建的 IPDT 动态测验, 相比传统静态测验, 多项指标可以更有效地预测儿童认知发展, 且认知干预对低 SES 不同潜能儿童的认知能力具有不同促进作用。研究有较重要的理论意义和较高的教育实践应用价值。

作者根据审稿人意见进行了多轮认真细致的答复和修改, 同意发表。

主编意见:

同意外审和编委意见, 建议录用