

《心理学报》审稿意见与作者回应

题目：动态测验对“数学学习困难”儿童的进一步甄别

作者：张丽锦 张臻峰

第一轮

审稿人 1 意见：论文在学困生精细筛查和准确区分方面有独到的见解，对动态测验应用思路成熟，对于国内动态测验思想的推广运用有启示价值。修改意见如下：

意见 1：研究中提到小学四年级是语文和数学成绩开始产生高低差别的关键时期，因而以往研究大多选取四年级儿童作为研究对象，但如果工作记忆动态测验有筛查功能，是否可以在更低年级进行早期筛查与预测，请在讨论中给出建议。

回应：选取小学四年级是基于以往许多以数学困难儿童为被试的研究及其筛查标准（见引言部分第 5 段开头），这时儿童在语文和数学成绩上有了明显差异，可以采用能力 - 成绩差异模型。正如文中所述，基于操作便利性的考虑，许多研究采用的都是年级 - 水平离差法。本研究的目的是用动态测验进一步区分年级 - 水平离差法筛查出来的数学困难儿童，所以选取的是四年级儿童。动态测验是基于潜能评估指标进一步区分传统能力测验所筛查出来的特殊群体，是传统测验的有益补充；但它不能脱离和独立于传统测验。如果传统鉴别方法（如能力 - 成绩差异模型）可以对更低年级儿童进行数学困难的筛查，动态测验则可以应用于更低年级的儿童。目前，对更低年级儿童（甚至学前儿童）数学能力的评估开始关注对其数感发展水平的测量，并且发现儿童早期的数感不良是其日后数学学习困难的强预测变量（Chard, et al., 2005; Jordan, et al., 2010; Locuniak & Jordan, 2008; Seethaler & Fuchs, 2010）。对数感的准确评估尤其需要在干预中评估这样的动态测验范式。特别需要说明的是，尽管以往研究较为一致地发现了工作记忆与数学困难的密切关系（见引言部分第 2 段），但数学困难的核心缺陷至今仍不明确（Mazzocco & Myers, 2003），可以确认数学困难儿童的工作记忆有明显缺陷，但反过来却不一定成立，因此还不能将工作记忆与数学能力进行对应。本研究所采用的工作记忆动态测验考查的是个体视-空工作记忆和言语工作记忆认知加工的潜能水平，还不属于数学认知能力的测验，因此，尚不能单纯使用 S-CPT 针对更低年级进行筛查。我们在讨论中补充了这一观点。

意见 2：研究发现，与能力 - 成绩差异模型鉴别方法相比，工作记忆动态测验可以对筛查出的儿童进行进一步的分类，这说明传统鉴别方法存在一定的误报。那么，对于同样基于成绩、能力测验选取的对照组，其中是否存在漏报的“发展落后儿童”？当前研究中经过干预后达到“正常组”水平的发展落后儿童是否真的达到了正常水平，有待进一步分析说明。

回应：本研究对对照组的要求是智力与实验组匹配，且不符合实验组初筛的条件（“每个班数学成绩处于后 25%、语文成绩处于中等或中等以上的儿童(42 人)”），就是说，对照组被试不是数学成绩落后儿童，这就不太可能是漏报的“（数学）发展落后”儿童。本文在被试筛选的介绍方面有不严谨之处，漏掉了“在 42 人之外”的特别说明，修改稿中已补充。感谢审稿专家的细致审阅！

至于“发展落后”儿童是否真的达到了正常水平，本研究关心的是通过动态测验的检验，发现这些“发展落后”数学困难儿童经短期干预后能够达到正常水平，且成绩明显好于对于干预不敏感的“发展缺陷”儿童，可以确定这些“发展落后”儿童不是真正的数学困难儿童，更

准确地说，他们和数学困难儿童有着分属于不同水平的差异。而以往大多数研究者倾向于采用的“年级-水平离差法”则不能对所筛查的群体作进一步区分，这不利于对这类儿童进行针对性干预，也不利于对干预的方法、策略和方案的效果进行评估、总结与推广。

意见 3: 文章在动态测验的文献综述部分最着重介绍的关于健康老人、轻微认知损伤者、早老性痴呆患者的研究，虽然与当前研究使用动态测验范式对某一群体进一步细分的研究设想相一致，但其被试群体和所测内容与当前研究相差较大，似不能作为该研究选用动态测验方法的核心依据。

回应: 我们认为对健康老人、轻微认知损伤者、早老性痴呆患者的研究（Ballesteros, Maria, Lluís, Rosa, & Julio, 2003）首先证明了针对潜能评估的动态测验范式能够对某一群体进行更为细致的类别区分。其次，动态测验是基于维果斯基的“最近发展区”思想而发展起来的通过积极干预对个体的感知、思维、记忆与学习乃至问题解决进行评估的测验范式（Tzuriel, 2000），是指向于潜在发展水平的测量。所谓“潜能”，也即“潜在能力水平”，指的是“潜在认知发展水平”，也称为“认知改变力”（cognitive modifiability）（Feuerstein, et al., 1979）。Ballesteros 等人（2003）的研究和本研究运用动态测验对某一群体做细致区分所考量的指标都是“认知潜能”，都是以认知能力（视觉-空间记忆、言语/词汇记忆等）的变化作为潜能的指标，所依据的动态测验的理念和范式也都是一样的。另外，引言中提到的 Swanson 和 Howard（2005）的以动态测验对阅读困难儿童进行进一步甄别的研究，也是重要的依据之一。因此，我们认为这些研究可以作为选用动态测验进行本研究的核心依据。

意见 4: 该研究旨在探查动态测验指标对传统筛查的补充，应该注意到，研究中静态测验与动态测验筛查效果的比较是以工作记忆为测验内容的，而目前国内普遍采用的数学困难鉴别的年级水平离差法是基于数学成绩和一般认知能力测验的，因此客观上存在着工作记忆测验与能力、成绩测验在数学困难鉴别中效果的比较，研究中并未提及。

回应: 本研究的目的是对数学困难进行进一步分类，首先就要筛选出数学困难儿童，而目前对数学困难儿童的诊断和甄别大多是基于能力-成绩差异模型的理论主张和操作定义。本研究所主张的并非工作记忆对数学困难儿童甄别的效用，而是动态测验在其中的作用。只是 S-CPT 是针对与数学困难密切相关的工作记忆而编制的动态测验，指向于其他认知能力的动态测验（参见，张丽锦，盖笑松，方富熹，& 方格, 2003）只要适合于小学四年级儿童的认知水平应该都能够探查到被试的认知潜能，并依据潜能作进一步分类。从上世纪 60 年代对学习困难进行识别、诊断和训练开始到现在，数学困难鉴别方法经历了从能力-成绩差异模型到干预反应范式（RTI）的重要改进。动态测验和 RTI 的鉴别并不是完全摒弃传统的诊断标准，而是在传统诊断标准的基础上进行动态评估和干预，这样得以更为准确、细致地鉴别和区分学习困难儿童。从本研究一系列动态测验分数上也可以看出，经过动态干预，“发展落后”组的分数有了明显优于“发展缺陷”组的改进，证明了动态测验的区分和鉴别效用。另外，工作记忆与一般认知能力并不矛盾，或者说，是认知能力的一部分，二者并不是对立的关系。相应的观点我们将在讨论部分做补充。对问题 4 的说明还可参见作者对问题 1 的答复。

意见 5: 结果部分 3.2.2 中，三类儿童在初始分上差异显著，S-CPT 是对工作记忆的考量，那么，从工作记忆角度区分或甄别数学学习困难儿童是否也是一种诊断方法？而不必使用“前测-干预-后测”的动态测验范式？

回应: 实际上，作者在实验设计时曾预测或假设：在初始分上，对照组成绩好于“发展落后”和“发展缺陷”组，而后两组之间没有差异。但统计结果却发现三组之间都有显著的差异。

仔细分析，我们认为这个结果是可解释的。S-CPT 毕竟是基于工作记忆而编制的动态测验，而工作记忆不足与数学困难有着显著的正相关关系，数学学习困难被证明与数学能力（计算加工技巧、数字事实提取、位值概念、数感）或/和一般加工能力（工作记忆、加工速度）缺陷有关（Chan & Ho, 2010），那么，不同程度的数学困难儿童的工作记忆就有可能不同。但数学困难的核心缺陷目前仍不清楚，虽然同为学习困难的分支，数学困难与它的“孪生兄弟”阅读困难却大不相同。数学困难表现出具有很大内部差异的特殊症候群，而且还存在与阅读困难共生的现象（参见：孙金玲，张承芬，2007），由此体现的多种亚型表现使数学困难更难找到统一的定义。即便数学困难儿童存在某一个或多个方面工作记忆缺陷，但工作记忆缺陷却不单独指向于具有多种亚型的数学学习困难。因此，用工作记忆单纯诊断和进一步分类数学困难儿童并不十分可靠。对该问题的回答还可参见对审稿专家问题 1 的答复。

意见 6：图 1 和图 2 几个动态测验分数有相似的特征和性质，应该综合到一个图里。

回应：接受审稿专家的建议对图的结果进行整合。因初始分、获得分和保持分是 S-CPT 的基本分数；而加工差异分、加工稳定分、策略效能分和指导效率分是延伸出来的动态测验分数，因此将图 2 和图 3 的结果合并，把延伸的动态测验分数整合到一个图（图 2）里。

审稿人 2 意见：

动态评估作为一种探查儿童潜在发展水平的测验方法，已在国外得到研究者和教育实践者的大量关注，并在提升儿童认知能力的发展等方面做出了很大的贡献。《动态测验对“数学学习困难”儿童的进一步甄别》一文，是利用“Swanson 认知加工测验”（S-CPT）对“数学学习困难”儿童的能力进行动态评估的尝试，具有很好的理论意义和教育实践价值。整体看，论文思路清晰，数据统计合理，结果可靠。但以下几个问题需要进一步澄清：

意见 1：为什么要选取四年级的儿童做被试？应当在引言中提到这一点，而不是在讨论中。

回应：接受审稿专家的建议，在引言中介绍被试选取的理由。

意见 2：引言部分对于获得分和保持分没有明确定义，读者很费解。

回应：因动态测验分数类型繁多且有自己独特的解释含义和操作定义，本文对动态测验各种分数的介绍放到了“2 研究方法·2.3 施测程序”部分了，并且相关动态测验分数用黑体显示，这样结合测验工具集中介绍可能会更清晰些。引言部分主要集中阐述为什么进行该研究的文献回顾和问题阐述了。根据审稿专家的建议，我们对“2.3.2 S-CPT 的施测（程序）”中“(4) 延迟后测”里关于保持分的计分移到了“2.3.3 S-CPT 的分数类型、计分及分数转化”，这样使内容更集中些。

意见 3：P5 中“发展变化的角度跨时诊断”的内涵应阐述清楚。

答：这句话是结合文中前面提到的“年级-水平离差法”来讲的，该测量模型不考虑文化教育背景而采用同一个范式和标准进行测量和筛查，而后续发展的测验范式——动态评估和由此拓展的干预反应范式（RTI）则考虑了环境、教育带来的变化而进行多时段的诊断。感谢审稿专家的建议，已在文中做了进一步解释。

意见 4：应对研究实施的流程进行更清晰的描述。如四层提示的实质是什么？干预是如何进行的？干预过程的时间有多长？目前的阐述太概括。同样，对干预后不同类型分数（得分）的介绍也要在程序中提前讲清楚。

回应：S-CPT 是指向于工作记忆的标准化的干预范式，Swanson 在编制测验时对施测程序和步骤已作了具体规定。采用的是即时干预的做法，干预程序就是文中“2.3.2 S-CPT 的施测”中的“(2) 干预”中所介绍的。因每个被试经提示后在不同难度的项目上反应不一样，有的只需一级提示(1 个词汇)就能正确回忆，有的则可能需要三级或四级提示(3 或 4 个词汇)。我们考虑，介绍测验的程序一般是先介绍测验，后介绍计分；也由于动态测验的计分种类比较繁多复杂，所以本文在方法部分单独列出一节 2.3.3 来介绍 S-CPT 的分数。动态测验的程序和分数与传统静态测验不同，我们在文中做了一些补充性的解释(以蓝色字体标注)，以增加可读性。

意见 5：“Swanson 认知加工测验”(S-CPT)中的“押韵词测验”是否经过作者的修订？修订后的信度与效度如何？在材料的音节数量上是否考虑到汉语的特点？是如何根据测验提供的百分等级计算得分的？作者应该在研究方法部分充分报告工具的可靠性等问题。

回应：S-CPT 中的子测验一押韵词测验是用来测试被试对韵脚相似词汇的听觉记忆。施测者以 2 秒 1 个词的速度读一组押韵词，询问被试这一组中有没有某个指定的词。例如，施测者阅读“run-fun-gun”，然后询问被试“sun 或 fun”是否在这一组词中。接下来，施测者再要求被试按顺序回忆刚才呈现的一组词(即 run-fun-gun)，如果被试遗漏、错误回忆，则按照测验规定的顺序呈现不同等级的提示。本研究在修订押韵词测验时，没有按照英文词汇进行翻译，而是根据韵脚相似的标准选择中文词汇，因为英文里押韵的词汇翻译成中文并不押韵。对于测验项目内的元素(本测验中即为押韵词)而言，词汇均为双字词的押韵词，例如，押韵词测验的项目 3 修改后是“信封-/蜜蜂-刮风-/山峰”。关于词汇的熟悉性，S-CPT 是适用于从 6.5 到 70 岁人群的工作记忆的动态测验，对于听觉加工测验来讲，不必保证所有词汇对于所有被试都是熟悉的。我们熟悉的传统测量理论中更关注测验性质的考查；而动态测验理论更关注“前测-后测”间干预后的差量变化及认知改变。因此，我们认为修订后的押韵词测验与 Swanson 的原测验的目的和检验效果是一样的。

本研究的分数转换方式根据 S-CPT 制定的方案标准，首先查阅《S-CPT 手册》中附表 A-2~A-7，将各子测验原始分数转化为量表分数，再参照附表 A-1 将量表分数转化为 100 为平均数、15 为标准差的标准分数。A-8 和 A-9 是对指导策略和策略效能的赋值表。S-CPT 没有采用百分等级的方式。这些信息都是手册中的细节，所以没有在文中体现。我们在文中的表述是“根据《S-CPT 手册》中的 9 个参考表，……”(见 2.3.3)。

意见 6：作者在实验材料中提到选取瑞文标准推理测验中的 20 题，满分 20 分。在选择被试时又将测验低于 90 分被试排除。请作者阐述清楚在研究中是如何使用瑞文推理测验工具的。

回应：已在文中补充瑞文测验的介绍，见 2.2 实验材料(1)。将原始分转换成了 100 为平均数、15 为标准差的标准分数，以 90 分作为筛选被试的分界点。文中原来含糊的表述是我们的疏忽，感谢审稿专家的细致评审。

意见 7：文中的图中要标识出标准误。三类儿童应作为横轴的变量，可以更清晰的看到不同儿童的能力变化。

回应：接受审稿专家的建议，已将标准误在图中标识。

关于图的横纵轴转换问题，动态测验分数比较的是每种分数(如初始分、获得分、保持分等)里各类被试组的差异，而不是每类被试中初始分、获得分和保持分的差异，所以分数类别应该是横轴的变量。

意见 8：作者提到“快速”区分，请明确阐述“快速”的意思。

回应：关于快速的含义，在文中 4.2 讨论的最后一段进行了阐述。经审稿专家提示，我们对一些表述又做了补充和修改，使对“快速”的描述更为清晰。

意见 9：作者在讨论中提到小学四年级是语文和数学成绩开始产生高低差别的关键时期，但又强调动态测验对数学困难儿童的早期诊断和干预有帮助。作者需要在讨论中阐述是如何考虑这一问题的。

回应：对数学困难的鉴别大多使用的是能力 - 成绩差异模型，该模型的计算方法经历了年级 - 水平离差法、期望公式法、标准分数法、回归分析法的改进历程，这些计算方法都是基于学习成绩的鉴别方法。如本文所述，“小学四年级是语文和数学成绩开始产生高低差别的关键时期”，这时依据学习成绩的鉴别才有可能，也才有意义。近些年，对数学能力的鉴别又有了突破“能力-成绩差异模型”框架的新进展，即动态测验和基于动态测验理念发展起来的干预反应范式（RTI）的方法。对于更低年级儿童数学能力的考查，并不适合使用“数学学习困难”的概念，因为学习困难毕竟是学习能力方面的欠缺，涉及到与学校课程相关的学习能力和学习成绩，而更低年级学生成绩分化还不甚明显。后来，有研究者使用了“数感”的概念和测评方式，并且证实数感是数学困难的强预测变量，而数感是个体在学前期就有表现和发展的数学能力。以动态测验范式对儿童的数感进行评估，可以及早发现不同程度的数感不良儿童，有利于及早诊断和及早施以针对性干预。我们将在讨论中对这些内容进行介绍；同时也见对审稿专家一问题 1 的答复。

意见 10：文章行文中断句有不少问题。请仔细阅读，在需要使用句号的地方使用句号。

回应：我们已对文稿中的句子进行了修改，并请同行进行了语句语法的修改。

意见 11：英文摘要的个别语句需要修改。建议直接从动态评价入手（本研究关注的问题）入手来写。

回应：从本文的题目可以看出研究的主要内容是对数学困难儿童进行进一步区分，所以摘要先提出了问题，即以往研究诊断标准的缺陷，由此引入了“动态测验”的概念和研究目的。我们认为这样的摘要框架更易于读者了解本文的意图和研究目的。

另，英文摘要已请在国外大学工作的发展心理学研究者把关、修改。

最后，因目前还不涉及排版事宜，跨页表的问题暂未处理。

第二轮

审稿人 1（这个审稿专家是第一轮的第 2 位）意见：

作者较好地回答了审稿人提出的问题。但仍有两处还需要根据评审意见进一步修改。

意见 1：图的横纵轴问题。如果作者坚持以分数类型作为横轴，那么请在文章结果的表述（如标题）中不要从三类儿童的角度表达，而要从得分类型上描述三类儿童的特点；

回应：经审稿专家再次评议，我们又仔细思考了图表的表达方式，把动态测验各分数类型放到横轴上而不区分确实不太恰当。由于动态测验成绩比较的是每种分数（如初始分、获得分、保持分等）里各类被试组之间的差异，所以，我们把各类分数之间加上了明显的区分线，以示各类分数之间是非连续的，详见图 1、图 2。在结果 3.2.2 中对标题和统计表述也做了修改（见深蓝色字体）。非常感谢审稿专家的提醒！

另外，这次修改检查时发现上一稿在图中标识的是标准差，这一稿做了修改，图中标识的是标准误。为我们上一次的疏忽表示歉意！

意见 2：英文摘要。作者没有理解审稿人的意见。审稿人建议以本研究的研究问题为出发点写英文摘要，这样更符合西方读者的阅读习惯，也使得研究问题更加突出。作者目前的写法是提出原有以往研究中存在的问题，然后介绍研究目的。审稿人对于摘要中的语言表达没有意见。

回应：谢谢审稿专家的建议，很抱歉我们没有把修改思路交待清楚。《心理学报》要求的英文摘要是 500 字的大摘要，其中对 background 的写作要求建议为“为主要的研究问题和假设提供简要的理论背景介绍。理论背景和文献综述要求精炼简洁，但也不能仅仅列出研究问题或研究假设。在描述事实信息时请使用现在时态，报告你的研究问题和假设时使用过去时态。背景部分不得超过 200 字”。作者考虑，如果以“研究问题”开头来撰写，对理论背景的介绍就得放置其后，就得以倒叙手法撰写。不知这样是否合适？如果需要这样构架，作者将进行修改。目前，我们采取了折中的处理办法，即，把原来的 background 的两段并为一段，并进行了删减凝练，使这部分更紧凑，便于读者阅读时尽快捕捉到研究问题（详见英文摘要部分）。不知这样可否？