

《心理学报》审稿意见与作者回应

题目：小学低年级儿童的等值分数概念发展及干预

作者：辛自强，韩玉蕾

第一轮

审稿人 1 意见：

意见 1：该项研究采用实验的方法考察小学低年级儿童等值分数概念的发展并进行了合理的促进研究，收到预期的效果，研究选题既有学术价值又有教学指导意义。问题提出有理有据，实验设计基本合理。但是，有几个问题值得商榷：

1.表 3 中各种三种实验条件下的被试人数各是多少没有标注，建议补充。

回应：我们已修改。在表 4 中补充了三种实验条件下的被试人数。

意见 2：既然对表 3、表 7 和表 14 中的数据所做的统计分析是重复测量的方差分析，那么，重复测量的变量是什么？为什么要做重复测量的方差分析？重复测量的方差分析应该怎样做？

回应：重复测量的变量是被试在部分绝对量相等和整体绝对量相等这两种干扰条件下的得分。由于干扰条件是被试内变量，被试在两种条件下的得分存在关联，因此，需要进行重复测量的方差分析。

意见 3：各种实验条件下的被试数量都是小样本，得出文中的发展规律的可靠性有多大？如何判断被试是否理解或掌握了等值分数的概念？

回应：小样本确实可能不利于实验结果的稳健。但我们（1）选择被试时尽量选择有代表性学校的被试，做了性别平衡、保证样本的同质性，这有助于保证样本的代表性；（2）更重要的是，心理学除了强调取样的随机性，更强调进入不同实验条件时的“随机分配”，我们严格遵循了随机分配原则，有助于控制样本个体差异的干扰（所以，大量已发表的心理学实验的单元格样本量也都不足 30 人）。（3）我们做的三个实验（具体是 4 项实验）能做到相互印证和支持。所有这些都助于保证所获规律的可靠性。（4）此外，虽然实验 1 的单元格样本量大多在 15-20 人之间，但总样本量已经超过 150 人，对于需要个别施测和文本编码的实验研究而言，已经是一个巨大的工作量，实在无力做更大规模的被试。

被试要真正理解等值分数概念，需要发展起等值分数的运算思维，包括守恒观念和乘法思维。获得守恒观念的前提是对相对量的认识，即认识到整体的大小是由两个部分量共同决定的。然而，正如本研究所欲阐明的，对于发展研究而言，我们不仅要揭示一个概念理解的最高阶段，更要揭示此前要经过哪些阶段，即说明概念如何从不成熟而逐渐趋向成熟。无论是皮亚杰还是新皮亚杰学派的发展研究都采用了类似做法。具体到每个发展阶段，表 2 提供了相应的操作定义和编码标准。

审稿人 2 意见：

意见 1：《儿童的等值分数概念发展及干预》通过分析各年级儿童的运算思维水平，概括出等值分数概念发展的三个阶段，接着基于“最近发展区”理论设计了两个干预实验，分别来提高一年级儿童的相对量概念和二年级儿童的乘法思维。研究揭示了低、高年级儿童的等值分数概念水平的差异，以及从低年级到高年级的概念发展过程。实验结构完整、设计严谨，对

以往研究有推进；数据统计科学、结果有效。不仅具备理论价值，同时对教学实践有较强的指导意义。文章结构清晰、层次分明、文笔流畅，引用文献较为全面，讨论较为有理有据。

但还存在一些值得商榷的地方：

1、文章题目叫《儿童的等值分数概念发展及干预》，但实际上只针对于小学低年级儿童，是否题目应该加以界定？虽然文章中说到为何选取这三个年级，但不能排除儿童的等值分数概念在三年级之后仍有发展。

回应：我们赞同这样做，已经修改题目为“小学低年级儿童的等值分数概念发展及干预”。并在正文，注意限定了结论的年龄适用范围。

意见 2：干扰条件设置不够科学，只有整体绝对量相等和部分绝对量相等，并没有完全不相等的作为对照。从直觉性整体量图式和学生采用的一些策略来说，整体绝对量相等可能在某种程度上不仅不造成干扰还减小了判断的难度。对于这样设置的理由在文章中论述还并不够充分，讨论也还较为单薄。

回应：本研究是在前人研究基础上进行设计的，着眼点是考察小学低年级儿童的等值分数概念的发展特点。根据文献综述，小学低年级儿童的困难主要在于受整数等值的干扰而出现按绝对量判断的错误，因而本研究设计了部分和整体绝对量相等两种干扰条件。对于未接受数学教学的学前儿童，可能通过直觉性整体量图式来解决问题；对于接受正式教学的儿童，由于首先接受整数教学，发展起整数等值图式，这可能会对后面学习分数等值产生干扰。本研究的研究对象是接受过正式数学教学的儿童，三项实验的结果均表明，在两种干扰条件下，都出现了按整数等值判断的错误。这说明这些被试尚没有获得数量化的相对量概念。只有对于那些获得成熟的相对量概念的个体来说，整体绝对量相等的条件才更容易判断。

本研究中采用的干扰项和原始项的值均位于 $1/2$ 两侧（即，如果原始项的分数值小于 $1/2$ ，那么干扰项大于 $1/2$ ，反之亦然）。Spinillo 和 Bryant（1991）等先前研究发现 6 岁以上儿童能够很好地区分 $1/2$ 两侧的分量。因此，“完全不相等”条件对于本研究的被试而言可能过于容易，有可能出现“天花板效应”，因而本研究中没有设置这一条件。表 4 的数据显示小学一至三年级的被试得分分别在 10、12、14 分之上（而满分 16 分），这说明题目不应再降低难度了。

意见 3：实验是在 Boyer 实验基础上进行改良，但并没有就改良的理由阐述清楚。加入新的“离散量”把原实验的离散量当做“过渡量”，作者给出的理由是被试可能会忽视原实验中离散量的刻度将其当做“连续量”，但从实验材料上来看，刻度非常明显，被试不可能注意不到（这一点从结果和作者自己的讨论中也可以看出来）。作者的改进确实可以在测量被试的概念发展上起到一定作用，但是文中这种思想表达不够明显。

回应：对此，我们在问题提出部分的倒数第二段作了补充阐述。Boyer 实验的离散量任务虽然有刻度标志，被试应该可以注意到，但这无法排除被试将之还原成连续量来处理，因为刻度只是一个冗余信息，不考虑它也可以解决问题。所以 Boyer 实验的离散量任务，同时提供了连续量和离散量信息，这是一种“混合”了两种信息的条件。我们在修改时采纳了审稿专家建议，将这种情况称为“混合量”条件。

意见 4：作者将自己新增材料作为“离散量”，原实验的“离散量”作为“过渡量”其实是没有完全理解 Boyer 的实验。在 Boyer 的实验中“离散量”本来就是一种“干扰”。作者改进后的实验中“过渡量”的难度比“离散量”更高（干扰程度大），从渐进角度上来讲是不科学的。其实“过渡量”的实质并不是过渡，而是一种“混合量”，否则命名易导致歧义，产生过渡量难度低于

离散量的印象。

回应：已修改。确实如您所说，应将文中“过渡量”修改为“混合量”，这样更容易解释结果。

意见 5：实验一中 F 值的自由度显得很突兀，从作者给出的被试数量来看找不到依据，这一点需要补充交代。

回应：已修改。表 4 中增加了各条件下的被试数量。

意见 6：严格说实验一没有操作自变量，因此准确地讲应该是测量而不是实验。

回应：感谢审稿专家的建议，这实际上是一个非常复杂的问题。我们这样理解：对于发展性结论（年龄差异）而言，确实只是测量而非实验，因为年龄不是可以操纵的变量，只能通过观察和测量描述发展差异；但是实验一，在数量性质和干扰条件两个变量上都是人为操纵的，能够推论因果关系，是一个实验设计。因此，实验一同时有测量和实验的含义，但为了简化提法，我们还是笼统称为“实验”一。谢谢专家提醒。

意见 7：在实验一中，对干扰条件进行了介绍，但对匹配的条件缺乏介绍，使读者阅读实验材料时不够清晰。

回应：我们补充介绍了有关内容。详见实验一 2.1.2 的“刺激材料和仪器”第 2 段。

意见 8：图 1 展示的实验材料中最后一部分没有达到整体绝对量绝对相等。

回应：图 1 已修改、完善。

意见 9：实验二 a 与 b 中对照组与实验组被试做的题量不一样，两者练习量不等，其结果的差异可能是练习量的不等所致。

回应：实验二 a 中，对照组没有练习任务，干预组则做了 8 道连续量的练习任务，结果表明干预组在后测的离散量任务上的表现更好。为了排除单纯练习效应的影响，在实验二 b 中，一组被试先做连续量题目，再做离散量题目，另一组被试先做离散量题目，再做连续量题目，这两组被试的做题量是相同的。结果发现，在离散量任务上，第一组被试表现更好；而在连续量任务上，两组被试的表现无显著差异。这排除了练习效应的影响，也说明只有连续量任务对离散量任务具有迁移作用，而离散量任务对连续量任务没有迁移作用。

意见 10：实验三通过量的成倍增长来发展儿童的相对概念，有些牵强。因为乘法思维的本质是倍数关系，是加法的简便运算，培养相对量似乎不是最合适的材料。

回应：实验三的干预对象是二年级儿童，他们已经发展起相对量概念，但还不能进行精确的乘法运算。因此本实验的主要目的是提高二年级儿童的乘法思维，干预原理是通过让儿童对两种事物按一定的对应关系进行组合，促进其对一种情境中两个维度之间对应数量关系的理解。结果表明，干预组的得分和使用正确策略的频次均显著高于对照组，且干预组的一些被试开始使用比例计算策略，而干预组比对照组在加减运算的使用上明显降低，这说明干预条件能有效改变儿童错误的运算思维。

不过，干预组出现了部分未对应的新错误，这可能是由于干预组过度关注数量关系，而忽视了各部分之间的对应。这是实验设计的一个瑕疵。

意见 11：在实验三中，实验组被试做 16 道离散量题目，对照组做 16 道过渡题目。为什么对照组与实验组做不同类型的题目？两种类型题目的难度不同，两者统计的差异很可能是难

度不同导致的，怎么解释？

回应：实验组和对照组在正式测验阶段做的都是“混合量”（过渡量）题目，原稿中是笔误，已修改。

意见 12：实验三的材料中只有双人桌材料的正式测验材料与示例材料不完全匹配（即不是整数倍匹配），而别的都是整数倍的匹配。对这一点没有交代原因和实验程序的不同。

回应：双人桌也是整数倍匹配，原稿中是笔误，已修改。

意见 13：还有一些小问题及笔误，见全文中的批注。

回应：我们认真核对了文稿的错误，所有修改之处都以蓝色标记。参考文献和英文摘要也进行了细致的修改。

第二轮

审稿人 1 意见：

意见 1：作者既然是做重复测量的方差分析，那么，您对重复测量的变量重复测量了几次？重复测量的结果之间的差异是否显著？这是应该报告的。如果您是利用 SPSS 作的重复测量方差分析，那应该报告数据是否符合作重复测量方差分析的前提条件，这也是需要报告的。希望您能够回答上述问题。

回应：本研究之所以采用重复测量的方差分析，是因为本研究的三个实验均采用了混合设计，其中被试内变量是指两种干扰条件。为考察儿童对部分和整体的加工特点，实验设置了两种干扰条件的题目，一种条件是干扰项和原始项的橙汁量相等（即部分绝对量相等），另一种条件是干扰项与原始项的橙汁和水总量相等（即整体绝对量相等）。每个被试要做这两种条件下的测试题目（详见实验一 2.1.2 刺激材料和仪器）。相应地，实验结果的报告包括每个被试在“干扰一”条件题目上的得分和“干扰二”条件题目上的得分。因此，对实验结果的分析采用了重复测量的方差分析，三个实验的结果分析均表明干扰条件的主效应显著，效果量达到 0.16-0.2，说明儿童在加工过程中倾向于关注部分，而非整体。

对于真正的重复测量（如前后 2 次的重复测量）以及被试内设计（1 名被试完成 2 个任务）都需要采用“重复测量的方差分析”，本研究属于后一种情况，我们已经在文中作了说明，说明了实验设计类型和方差分析的自变量及相应水平有哪些，这样读者就可以理解为什么采用重复测量的方差分析了。

审稿人 2 意见：

意见 1：所有的干扰项的分值都比匹配项的分值大？在讨论分析应说明。

回应：我们在设计实验时尽量保证干扰项与原始项的“差值绝对量”大体一致，并对两种条件进行了平衡，其中 8 道题的干扰项大于原始项，另外 8 道题的干扰项小于原始项（如表 1），原表中的差值原是指差值绝对量，为防止误解，现已修改为“干扰项-原始项的差值”。研究 2 中同一类型表也修改。

意见 2：自由度数值对吗，请核对，上下同。

回应：原始计算有错，已修正了。多谢。

第三轮

审稿人 1 意见:

意见 1: 您没有对“怎样使用重复测量方差分析”的问题作出正面回答。请您仔细看看, 您回答的是我问的问题吗? 请您请教统计学专家, 弄清楚究竟什么是重复测量的方差分析? 怎样运用重复测量的方差分析?

回应: 您好, 非常感谢您的建议, 我们认真学习了重复测量方差分析的问题, 也请教了两位统计专家, 并按照您的建议对论文的这部分统计做了仔细修改。具体说明如下:

首先, 从研究设计本身来说, 实验一的设计是 3 (年级: 一、二、三年级) \times 3 (数量性质: 连续量、混合量、离散量) \times 2 (干扰条件: 部分绝对量相等干扰、整体绝对量相等干扰) 的三因素混合设计。其中年级、数量性质属于被试间变量, 干扰条件属于重复测量变量, 因变量为正确答题数。相当于对 9 组被试, 逐一施测两种干扰条件下的题目, 因而需要采用重复测量方差分析。实验二和实验三的设计均为 2 (干预条件: 干预组、对照组) \times 2 (干扰条件: 部分绝对量相等干扰、整体绝对量相等干扰) 的两因素混合设计, 其中干预条件属于被试间变量, 干扰条件属于重复测量变量, 因变量为正确答题数。相当于对两组被试, 分别施测两种干扰条件下的题目, 因而也需要采用重复测量方差分析。

其次, 我们修改了原来并不完善的统计程序。在具体的数据运算中, 由于重复测量变量只有两个水平, 球形检验的 $df = 0$, 因此需校正单变量检验的自由度, 取 Greenhouse-Geisser Epsilon (G-G) 校正系数, 在结果报告时我首先报告了重复测量变量的主效应, 以及重复测量变量与被试间变量的交互作用, 继而报告了被试间变量的主效应。具体的结果报告请见文中 2.2, 3.1.5, 4.2。

不知上述修改是否符合要求, 无论有任何问题, 我们都愿虚心学习, 改进。

第四轮

审稿人 1 意见:

意见 1: 既然“干扰条件属于重复测量变量”, 那么请问: 重复测量了几次?

回应: 我们没有重复几次, 只是同时测量的被试内变量。重复测量的方差分析这种统计方法里的“重复”二字, 并不能只做字面理解, 它包括两种情况: (1) 真正的重复测量 (如一个工具的前后测, 重复两次及以上); (2) 同时测得的被试内变量 (我们的研究属于这类)。更准确地说, 只要研究设计里包含被试内因素, 就要首先考虑使用这种重复测量的方差分析, 因为被试内因素的不同水平的得分可能有相关。如果是混合设计, 即同时包含被试内和被试间的因素, 也需要用这种重复测量的方差分析。

在审稿专家的提醒下, 我们认真阅读了有关方法类书籍, 也请教了统计专家。举例来说:

- (1) 在舒华, 张亚旭 (2008, pp.187-195) 两位教授的方法书里有同样的三因素混合设计的例子, 她们认为要用重复测量的方差分析。这个例子的实验设计和我们的研究结构相同。原文如下:

一、重复测量一个因素的三因素混合设计

这种设计的特点是，研究中包含三个因素，其中一个因素为被试内变量，另两个因素为被试间变量。假设一项阅读研究包含如下三个因素：(1) 年级 (A)，分初中二年级 (A1)、初中一年级 (A2) 和小学五年级 (A3) 三个水平；(2) 推理能力 (B)，分高 (B1) 和低 (B2) 两个水平；(3) 名词的词频 (C)，分高频 (C1) 和低频 (C2) 两个水平。其中，年级和推理能力为被试间变量，词频为被试内变量。因此，这是一个 $3 \times 2 \times 2$ 混合设计，包含 12 种条件。研究的目的是考察同高频的名词相比，

p.188

被试对低频的名词的阅读时间是否更长，以及二者之间的差异，即词频效应如何受被试年级和推理能力的影响。

p.190 [接上页]

第二步，重复测量方差分析。在重复测量两个因素的三因素混合设计中，为了确定每个因素是否真的起作用，以及所起的作用是否受其他因素影响，例如，规则性所起的作用是否受字频、阅读能力的影响，研究者通常需要进行重复测量方差分析，即 F 检验。具体步骤如下。

(1) 激活 Analyze 菜单，选 General Linear Model 中的 Repeated Measures... 命令项，弹出 Repeated Measures Define Factor(s) 对话框。在对话框的 Within-Subject Factor Name 和 Number of Levels 的后面，填

(2) 张文彤 (2002, pp.36-43)关于统计分析的书里也有类似例子。只要有被试内因素就首先要考虑重复测量的方差分析。

我们也仔细阅读了其它一些版本的统计教科书或实验设计的教材，到目前为止我们对这个问题的理解大致如此。我们的统计学知识有限，冒昧之处，请见谅。欢迎审稿专家的任何建议，请明确告诉我们应该如何做，如何处理这种包含被试内因素的混合实验设计的数据。

参考文献：

舒华，张亚旭 (2008). 心理学研究方法：实验设计和数据分析. 北京：人民教育出版社.

张文彤 (2002). SPSS11 统计分析教程. 北京：北京希望电子出版社.

第五轮

审稿人 1 意见：

意见 1：作者对统计方法的运用做了有根据的说明。建议作者：(1) 根据 3 个实验的结果逐条概括研究结论。

回应：首先感谢审稿专家耐心细致的审稿，更感谢专家能够认可我们对统计方法的运用及相应的解释。

对于3项实验的具体结果我们已经在每个实验的讨论部分做了概括。最后第6部分的结论是对上述结果的综合概括，“结论”是基于“结果”的，但不等同于结果，所以结论是最后的总体性结论，不再重复罗列结果了。

意见2：讨论部分可以合并在一起写，并根据按实验结果作出讨论；

回应：文章目前的结构是：每个实验后有相应的讨论，所有3个实验后有综合讨论，这样更为清晰。似乎不需要把所有讨论合在一起写。这一写法，审稿人在前四次的审稿中都没有表示异议，应该不存在明显错误。

意见3：中文摘要可以再精简、准确一些。

回应：已经作了精简，摘要的字数从312个字压缩至268个字，精简掉44个字。

意见4：删掉修改批注，按照学报的写作格式认真修改全文。修后再审。

回应：已经按照要求删掉批注和修改文字的颜色，按照学报格式要求做了加工。

第六轮

审稿人3意见：

意见1：审稿专家：先有发展的基础研究，后有干预研究，设计上比较严谨、可靠。建议发表。

回应：感谢审稿人对我们实验设计逻辑的认可，并作出“建议发表”的决定。

意见2：图一的标注不是很清楚。需要说明每一条件中两个选项的含义。

回应：我们在图1上面那段文字中，增加了对每个选项（干扰项、匹配项）设置原则和意义的说明。由于说明文字较多，没有放在图1的标注里。

意见3：表1中不需要标明人数。

回应：表1是关于实验材料的介绍，本来没有人数信息。表1最左边一列是材料中题目的序号，不是人数。

意见4：实验一中每一条件下两个选项的顺序是否有控制，作者没有说明。如果没有有效平衡，需要做讨论。

回应：实验一已经对所有题目的两个选项（干扰项和匹配项）的相对位置（或顺序）进行了平衡，并对两种干扰条件下的题目的呈现顺序进行了平衡。我们在图1之下那段的最后一句做了说明。

意见5：文章很长，是否可以把实验一中的有关策略的说明、结果、分析与讨论等内容删掉？

回应：实验一是为了说明儿童等值分数概念的发展水平或阶段，不仅要根据“定量”的结果来确定，更要根据儿童所使用策略的性质，对儿童概念水平进行“质”的判断，故而对儿童所用策略进行分析、说明和讨论。认知发展的研究通常需要结合使用这种定量和定性的分析，以便确定认知发展水平或阶段。当然文章确实有些长，为此，我们咨询过学报的政策，并不严格限定文章的长度。文章投稿后已经经过六轮审稿和修改，每次我们都尽量精减了文字。

意见6：实验二（a）的结果很好，建议去掉实验二（b），后者的逻辑不成立。

回应：实验二（a）的结果确实能大致说明研究的问题，但为了控制实验中的练习效应，我们专门设计了实验二（b）以排除任务熟悉性和练习效应造成的可能的干扰风险。对实验二的逻辑我们在这部分内容开始的一段做了交代。

意见 7：建议去掉实验 3，与前面关系不大。

回应：我们在实验一中证明了儿童等值分数概念发展分成三个阶段，实验二和实验三设计了专门的干预方法促进相邻两个发展阶段之间的转换，探索提高儿童认知水平的方法。其中，实验二，通过干预实验促进“一年级”儿童的“相对量概念”水平的提高；实验三，设计干预实验以促进“二年级”儿童的“乘法思维”的发展。实验二和三是两个并列的研究，针对不同年级被试，促进不同的概念水平的发展。实验二和三，通过干预实验进一步确证了实验一关于等值分数概念发展的三阶段模型；而干预是建立在对发展的现实水平的确定和最近发展区的估计上。因此，我们的三个实验的逻辑是清楚的，实验三与前面的实验是相辅相成的，不可或缺。

感谢各位审稿专家耐心细致的工作，我们的文章投稿后已经经过 6 轮审稿和修改，经过 10 余人次的专家审阅，我们受益良多，也钦佩《学报》审稿专家和编委精益求精的精神。