

《心理学报》审稿意见与作者回应

题目：编码强度对字体大小效应的影响

作者：赵文博 姜英杰 王志伟 胡竞元

第一轮

审稿人 1 意见：

研究选题有一定的理论意义和应用价值。但论文中还有这样一些问题。

意见 1：实验 1 的意义似乎不大，建议删除。特别是第 2.2.2 部分，以字体大小为自变量，连续识别反应时为中介变量，学习判断为因变量，进行的知觉流畅性在字体大小与学习判断间中介效应模型分析，有数字游戏之感。

回应：感谢审稿老师的意见。因为审稿老师的意见，我们重新思考并梳理了引出实验 1 的逻辑(详见前言第 3 段和第 4 段，P10-11，以蓝色标记)，对实验 1 的必要性进行了分析，请审稿老师审阅。考虑到以下两点理由，实验 1 暂时未删除。

(1) 实验 1 的结论对本研究整体逻辑具有重要作用。

结合本研究的实验目的，实验 1 是弥补前人研究的局限性。Yang 等人(2018)研究中要求被试在成功识别出目标词后立即进行学习判断，但是由于大字体和小字体单词之间的识别反应时存在显著差异，这种即刻学习判断使知觉层面的流畅性线索更加凸显且更容易被个体注意，个体更倾向将流畅性线索作为学习判断的线索(之一)，放大了知觉流畅性在字体大小效应中的作用。因此，知觉层面的流畅性线索是否是字体大小对学习判断产生影响的机制，这一问题尚未被完全解答。实验 1 旨在验证字体大小效应的出现是否是知觉层面的线索在起作用？实验 1 通过贝叶斯多层中介分析发现，知觉流畅性是字体大小效应产生的机制。随后，实验 2 和实验 3 通过激活更深层的线索(如语义线索)，削弱知觉线索对学习判断的影响，进而矫正字体大小效应。

(2) 实验 1 采用贝叶斯多层中介分析，为基于知觉流畅性的字体大小效应提供了更为严谨的统计依据。

该方法共 4 条马尔科夫蒙特卡罗链(MCMC)，每链共产生 10,000 次迭代来估计中介效应，中介分析结果更加科学严谨(Vuorre & Bolger, 2018)。贝叶斯多层中介分析方法与 Monte Carlo 法和参数 Bootstrap 相比，点估计和区间估计都最准确(方杰，温忠麟，2018)。因此，该方法可以准确地揭示出大字体通过减少识别反应时(即增加知觉流畅性)，进而增强学习者

的学习判断值，说明这种影响的确是基于知觉层面线索的作用，为实验 2 和实验 3 的展开提供了前提和基础。

参考文献：

- 方杰, 温忠麟. (2018). 三类多层中介效应分析方法比较. *心理科学*, 41(4), 962–967.
- Vuorre, M., & Bolger, N. (2018). Within-subject mediation analysis for experimental data in cognitive psychology and neuroscience. *Behavior research methods*, 50(5), 2125–2143.

意见 2：一般说来，“结果”部分是要求作者描述所发现的事实；“讨论”部分是要求作者对事实进行解释。而文章有个“结果分析”部分，既有对所发现事实的描述，也有对事实的解释，不妥。建议把这两部分内容区分出来，分别描述。

回应：感谢审稿老师的意见。

根据原文的结果部分与讨论部分出现重叠，修改稿已将“结果与分析”部分重新梳理，并全文更改，详见 P14、P18 和 P20。

.....

审稿人2意见：

本研究探讨了编码强度对字体大小效应的影响，一方面通过建立中介模型，考察了知觉流畅性在字体大小对学习判断的影响中可能发挥的作用，另一方面试图通过改变编码强度来降低被试对无效线索（字体大小）的依赖，消除元认知错觉。该研究具有一定的理论意义和实际应用价值，但也存在一些问题需要解释和修改：

意见 1：前言中，作者介绍了 Yang 等人使用的知觉范式连续识别任务（CID），提到“与真假词判断任务相比，Yang 认为 CID 任务对测量项目的知觉流畅性有更高的敏感度。”因此在文中通过该任务来评估大小字体的知觉流畅性。但并没有分析 CID 任务对流畅性敏感度更高的具体原因，请分析该范式的优势和必要性，以及是否可以用更简单的方法替代，比如看到字之后敲击键盘直接记录反应时？

回应：感谢审稿老师的意见。修改稿中详细分析了CID任务对流畅性敏感度更高的具体原因，并分析了该范式的优势和必要性，已经将上述内容整理到修改稿中P10、P11，以蓝色标记，请审稿人把关。再次感谢您的建议！这里对该问题简要回答如下：

相较于真假词判断任务，CID任务对测量知觉流畅性敏感度更高的具体原因如下：

（1）**实验程序方面：**在真假词判断任务中，被试首先需要识别学习词，然后尽快而准确地辨别学习词为真词(e.g., orange)或假词(e.g., atgeor)，最后选择真假词相应的按键并输入自己的答案(z = 真词, m = 假词)。此时按键结束后，电脑记录的反应时作为衡量项目知觉流畅性

的客观指标(Yap, Sibley, Balota, Ratcliff, & Rueckl, 2015); 在连续识别(CID)任务中, 学习词和掩蔽刺激交替呈现, 被试需要看清目标词后立即按空格键并口头报告出该词(此时仅有一个按键, 不存在额外区分相应按键反应时的干扰)。每个目标词有14次识别循环, 第一次目标词呈现17ms, 掩蔽刺激呈现233ms; 第二次目标词呈现34ms, 掩蔽刺激呈现216ms; 以此类推, 目标词呈现时间以17ms递增, 而掩蔽刺激以17ms递减。如此循环, 直到被试做出按键反应或14次识别循环结束。如果被试识别出目标词则需要立即按下空格键, 此时目标词和掩蔽刺激消失, 被试口头报告(或输入)该词, 电脑记录识别反应时(Yang, Huang, & Shanks, 2018)。

(2) **两种任务的差异之处:** 与CID任务相比, 除个体在知觉层面所消耗的反应时外, 真假词任务所获得反应时还可能涉及额外的按键反应时或学习词的语义提取反应时等(Chumbley & Balota, 1984)。可见, 真假词判断任务获得反应时并不单纯考察个体对知觉层面线索的反应, 往往涉及深层加工反应时或额外的反应时间混淆。与之相对的CID任务, 仅获取被试识别刺激的反应时, 排除语义加工过程的潜在影响。掩蔽刺激的作用为了消除当刺激消失后残留下来的知觉后像, 避免不同字体大小的学习词所残留的后像对识别时间的影响。综上所述, 与真假词判断任务相比, CID任务对知觉流畅性(知觉层面线索)的测量更加敏感。

参考文献:

- Chumbley, J. I., & Balota, D. A. (1984). A word's meaning affects the decision in lexical decision. *Memory & Cognition*, 12(6), 590–606.
- Yang, C., Huang, S. T., & Shanks, D. R. (2018). Perceptual fluency affects judgments of learning: the font size effect. *Journal of Memory and Language*, 99, 99–110.
- Yap, M. J., Sibley, D. E., Balota, D. A., Ratcliff, R., & Rueckl, J. (2015). Responding to nonwords in the lexical decision task: insights from the English lexicon project. *Journal of Experimental Psychology Learning Memory & Cognition*, 41(3), 597–613.

意见 2: 在前言中, 作者强调了真实课堂情景下由于字体大小不同可能引发的元认知错觉, 目的是突出该研究的实际应用价值。但是文中相关的阐述有些生硬, 逻辑关系不够清晰。比如在实际教学中, 不同字体大小的安排部分是为了帮助区分概念的层级关系, 有利于学习内容的组织, 而且学习的时候大字或者加粗的字往往具有强调的含义, 会吸引学生更多的注意力去学习, 这与实验室实验中设定固定时间进行特定项目的测量并不同质。建议作者对这个部分进行修改, 更好地挖掘出研究的实际应用价值。

回应: 感谢审稿老师的意见。在修改稿中(P23, 以蓝色标记), 已将实际应用价值按照您的建议进行重新的组织梳理。

矫正基于知觉线索引起的元认知错觉, 对实际的教学具有重要的意义, 并且仍存在其他途径对这种错觉进行消除。以往研究关于字体大小效应的内在产生机制仍存在争议, 其中主

要的理论为加工流畅性假说和信念假说。陈颖等人(2019)发现,改变学习者对加工流畅性的信念或是告知被试流畅性与记忆无关时,将消除字体大小效应。这种矫正方式保持了学习材料中字体大小的突出、强调或区分概念等作用,同时又通过改变信念,消除其造成的元认知错觉对学习过程的副作用。本研究基于线索利用模型,旨在激活更多深层线索以削弱知觉流畅性(知觉层面线索)对学习判断的作用,从而有效直接地矫正知觉线索所引发的元认知错觉。

意见 3: 前言的最后一段,作者指出“被试对不同字体大小项目的学习判断值均远远高于实际回忆成绩”,但给出的相关研究主要强调了个体的回忆正确率低,并没有将学习判断值与实际回忆成绩做比较。同时,作者在此处也没有分析导致编码强度低的可能原因,比如为什么呈现 5 秒仍然回忆率很低(在后面才提到可能是注意力不集中)?但在实验 2 中还是试图通过改变学习时间来操控编码强度。请对这一段的逻辑进行整理。

回应: 感谢审稿老师的意见。

(1) 现已在修改稿中重新整理了前言最后两段,并补充了以往研究导致编码强度低的可能原因(P11-P12,以蓝色标记)。具体修改内容如下:“分析字体大小效应的相关研究,发现这种元认知错觉的产生可能与被试学习时编码强度较低、激活线索不充分有关。在以往字体大小效应的相关研究中,被试的自由回忆正确率均较低(0.12~0.25)。例如Rhodes和Castel(2008)要求被试对每个词语学习5s,在该研究最后的回忆测试中,个体的自由回忆正确率仅为0.16。相似的结果在Mueller等人(2014)和Yang等人(2018)的研究中重复出现。虽然学习时间已经长达5s(理论上可以充分学习),但是往往个体实际的学习时间并不总等于限定的学习时间,他们可能在限定时间之前结束学习,或是在长时间的学习条件下出现注意力不集中、分心等现象(Cooper & Pantle, 1967; Mueller, Dunlosky, Tauber, & Rhodes, 2014),即学习者的编码强度(encoding strength)偏低,导致学习时间并未有效激活更多的线索用于学习判断。并且,根据线索利用模型的观点,学习判断出现偏差的主要原因是个体进行学习判断时使用了与真实记忆强度无关的线索(Benjamin, Bjork, & Schwartz, 1998; Koriat, 1997)。也就是说,当编码强度较弱时,学习材料的知觉特征线索容易对学习判断产生误导性作用,出现字体大小效应。本研究推测当编码强度增强时,将激活更多的线索,这些线索与个体真实的记忆强度相关,此时知觉特征线索的作用就会被弱化,进而降低或者消除知觉特征引起的学习判断错觉”。请审稿人把关。

(2) 本研究实验2中操控学习时间的方式,与以往研究有所不同。为了增加被试的有效学习时间,提高学习时编码强度,避免被试在限定的时间之前结束学习,被试需要以每秒一次的速度(词语的下方会出现同等大小的闪烁星号**)对屏幕上呈现的词语进行**复述**,实验主试对其进行记录(Lupker, Harbluk, & Patrick, 1991)。

(3) 为了进一步说明以往研究存在注意不集中、分心,即编码强度较低的现象。本研究将实验2中4s学习时间条件(复述)和实验3中浅层加工条件(学习方式不做要求)的自由回忆成绩

进行比较。两种实验条件的学习时间均为4s，唯一区别在于学习时间组的被试需要按照每秒一次的速度对学习词进行复述，而浅层加工组不做任何学习要求。最后结果发现，浅层加工组显著低于4s复述组的回忆成绩。因此，说明以往研究虽然采用5s的学习时间，但被试实际的编码强度偏低，并未激活更多记忆强度的线索，出现字体大小效应。

参考文献：

- Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: when retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, 127(1), 55–68.
- Cooper, E. C., & Pantle, A. J. (1967). The total time hypothesis in verbal learning. *Psychological Bulletin*, 68(4), 221–234.
- Koriat, A. (1997). Monitoring one's own knowledge during study: a cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, 126(4), 349–370.
- Lupker, S. J., Harbluk, J. L., & Patrick, A. S. (1991). Memory for things forgotten. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(5), 897–907.
- Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font size effect on judgments of learning: does it exemplify fluency effects or reflect people's beliefs about memory?. *Journal of Memory and Language*, 70, 1–12.
- Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, 137(4), 615–625.

意见 4：2.1.4 中没有说明反应错误如何处理，另外是否存在 14 次识别循环结仍未能给出正确答案的试次？如果有，怎么处理的？

回应：感谢审稿老师的意见。

在修改稿中，现已在实验1的“结果与分析”中对其进行补充，补充内容如下(详见P15，以蓝色标记)：“在实验1中，连续识别任务分别计算被试的连续识别正确率和连续识别反应时两个指标。连续识别正确率具体计算公式如下：

$$\text{连续识别正确率} = \left(1 - \frac{\text{错误识别个数} + \text{未成功识别个数}}{\text{识别总数}} \right) \times 100\%$$

由于连续识别任务难度较低，被试的连续识别正确率为97.84%，出现天花板效应。因此，实验1选用连续识别反应时作为衡量知觉流畅性的指标。如果被试口头报告错误，或14次识别循环结束，被试仍未进行识别反应(未按键或口头报告学习词)，则将剔除报告错误和未成功识别的试次，计算余下正确识别试次的反应时”。

意见 5：3.2.2 中在 2 秒、4 秒和 8 秒条件下都出现了字体大小效应，但 8s 学习条件下字体大小效应有所减弱。虽然从图 4 中能够发现这一趋势，但是缺少相应的统计检验，建议对三种时间条件下大小字体学习判断的差异量进行计算和比较，作为 8s 学习条件下字体大小效应减弱的证据。

回应：感谢审稿老师的意见。

现已在修改稿中补充了三种学习时间条件下(实验2)，大字体和小字体词语的学习判断值差异量的计算和比较。修改如下(详见3.2.2 学习判断值，P18-P19，以蓝色标记)：“将三种学习时间条件下，大字体和小字体词语的学习判断值差异量作为衡量字体大小效应变化的指标。结果显示2s和4s学习时间条件下的字体大小效应显著强于8s条件下的字体大小效应($p_1 = 0.001$; $p_2 = 0.02$)。2s和4s两种条件间的字体大小效应无显著差异($p = 0.97$)。值得强调的是，虽然随学习时间的增加字体大小效应逐渐减小，但是8s学习条件下大字体词语的学习判断值仍显著高于小字体词语， $t(24) = 2.87, p = 0.01, \text{Cohen's } d = 0.20$ ”。

意见 6：在 3.3 讨论中，作者提到：“随着编码时间的增加，编码强度逐渐成为学习判断的有效线索，弱化了学习判断对知觉特征线索的依赖，导致字体大小效应出现减小的趋势。”该表述存在一定问题。在该实验中，学习时间是编码强度的评估指标。从结果看，学习时间越长，字体大小效应越小，可以说编码强度越深，学习判断对知觉线索的依赖越小。但是学习时间主效应并不显著，也就是说学习判断并没有以编码强度作为评估线索，如何得出“编码强度逐渐成为学习判断的有效线索”这一结论？请作者进行解释或修改。

回应：感谢审稿老师的意见。在修改稿P19，以蓝色标记。重新梳理了本研究的实验逻辑，现已对3.3讨论进行修改，此处语言修改如下：“但是，随着有效编码时间的增加，可能激活更多关于记忆强度的线索用于学习判断，知觉特征线索对学习判断的作用得到弱化，导致字体大小效应出现减小的趋势”。

意见 7：4.2.1 中提到事后多重比较，但是自变量只有两个水平：深层加工和浅层加工，请修改相关表述。

回应：感谢审稿老师的意见。原文表述不准确，已将其删除。

意见 8：在 5.2 中作者提到“知觉流畅性的增加会间接地提高个体的学习判断值”，而实验 1 结果发现，知觉流畅性是作为中介因素影响学习判断的，这个作用是直接的。几处表述是否存在矛盾？

回应：感谢审稿老师的意见。在修改稿中，已对此处进行修改并对“总讨论”重新调整，详见 P22，“5.1 加工流畅性与字体大小效应”。

意见 9：在 5.2 中，作者还提到“向人们呈现与编码强度相关的有效性线索时，人们更倾向于选择有效的记忆线索作为学习判断的基础，字体大小对个体学习判断的影响则被弱化，从而

减小由于知觉特征引起的元认知错觉。”但是根据实验 2 和实验 3 的结果，无论是学习时间还是加工深度对学习判断都没有影响（主效应均不显著），如何能够推出“人们更倾向于选择有效的记忆线索作为学习判断的基础”？

回应：感谢审稿老师的意见。本研究考察编码强度操控有效性的因变量为自由回忆成绩，从实验2和实验3的结果来看，学习时间和加工深度对回忆成绩均具有显著影响(主效应显著)，说明不同条件下的编码强度确实存在显著的差异，也表明本研究的编码强度的操作是有效的。根据您的修改意见，在修改稿中已补充了编码强度对学习判断没有影响的解释(详见P24，第5至16行，以蓝色标记)，具体解释如下：“但是，与前人研究结果不一致的是，本研究实验2和实验3中均未发现编码强度对学习判断的影响，即学习时间和加工深度的主效应均不显著。造成这种现象的原因，可能是本研究将编码强度进行被试间操控。根据Koriat (1997)提出的线索利用模型，学习判断对不同线索的权重会受到学习词表形式的影响。具体来说，与被试间操控不同，当线索采用被试内操控时，外部线索(例如，学习时间)不仅影响学习者的学习程度，而且影响学习者的元认知判断。被试内操控会将这些因素的潜在作用放大，并促使个体利用这些外部线索。本研究为防止学习策略间的干扰(实验2：复述和实验3：造句)，将编码强度进行被试间操控。当编码强度作为被试间操控条件时，学习判断对被试间操控的因素并不敏感，这已经被大量研究所证实(Susser, Mulligan, & Besken, 2013)，所以编码强度对学习判断的主效应不显著，实验2和实验3仅发现编码强度对学习程度的影响。虽然，编码强度对学习判断的主效应不显著，但是增加编码强度有可能激活更多的深层线索(如语义线索等)用于学习判断，进而削弱知觉特征线索对学习判断的影响(即字体大小效应)”。

参考文献：

- Koriat, A. (1997). Monitoring one's own knowledge during study: a cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, 126(4), 349–370.
- Susser, J. A., Mulligan, N. W., & Besken, M. (2013). The effects of list composition and perceptual fluency on judgments of learning (jols). *Memory & Cognition*, 41(7), 1000–1011.

意见 10：该研究包括两个部分，一是考察知觉流畅性在字体大小影响学习判断中发挥的作用，二是试图通过改变编码强度来降低被试对无效线索（字体大小）的依赖，消除元认知错觉。但这两各部分之间的联系不够紧密。对作用机制的探讨聚焦在知觉流畅性的中介作用上，但干预却不是从知觉流畅性的改变入手，而是增加编码的强度。二者的联系需要作者进一步澄清。

回应：感谢审稿老师的意见。本研究在修改稿中重新组织了三个实验之间的联系(详见P12，前言最后一段，以蓝色标记)。具体说明如下：

“综上，本研究目的：

- (1) 验证知觉流畅性线索(知觉层面的线索)在字体大小效应中的作用;
- (2) 考察增加编码强度所激活的深层线索对字体大小效应的影响。

为达到上述目的,本研究开展三个实验,实验1通过将CID任务和JOL任务分别实施,在减小知觉线索的凸显性后,考察其对学习判断的影响。实验2通过增加有效学习时间,激活更多线索,考察编码强度对字体大小效应的影响。实验3,将进一步明确激活线索类型(语义线索),以提高加工深度的方式增加编码强度考察深层线索对字体大小效应的影响(Lupker, Harbluk, & Patrick, 1991)。对以上问题的探讨,不仅有助于全面揭示字体大小效应的内在机制,而且为矫正基于知觉特征的元认知错觉提供了重要启示”。简而言之,本研究旨在验证知觉层面的线索(即知觉流畅性)是否是字体大小效应的内在机制(实验1),进而随编码强度的增加,激活更多的线索,削弱知觉特征线索对学习判断的作用(实验2),并进一步明确深层线索能完全矫正基于知觉特征引起的元认知错觉(实验3)。

意见 11: 在前言的最后一段和结尾处,作者均提到,无效线索(如知觉流畅性)与有效线索(如编码强度)共同出现时,个体学习判断会更多地依赖编码强度等有效性线索。但是研究结果却并没有发现编码强度对学习判断的直接影响,编码强度更像是在发挥调节作用。所以,文中从线索整合角度对结果进行的分析逻辑不够清晰,建议调整,并修改第2个结论。

回应: 感谢审稿老师的意见。在修改稿中,已补充了编码强度对学习判断没有影响的解释(详见 P24, 第 5 至 16 行,以蓝色标记)。虽然,本研究中编码强度(学习时间和加工深度)对学习判断的主效应不显著,但增加学习时间和提高加工深度可能激活更多知觉特征以外的线索(如语义线索),且这些线索将同时整合于个体的学习判断中。在修改稿中,重新调整了对 Undorf 等人(2018)研究的表述,修改如下(P25, 第 2 段以蓝色标记):“Undorf 等人(2018)发现人们进行学习判断时并不是根据单一线索,向被试同时呈现多种线索(例如字体大小、词语的情感属性和词性等),这些线索会在个体的学习判断中进行线索整合(cue integration)和相互竞争(competition),削弱彼此对学习判断的影响。具体来说,该研究发现,在具体词条件下,情绪词的学习判断值显著高于中性词,出现情绪性效应(the emotionality effect);在抽象词条件下,虽然这种情绪性效应仍然存在,但显著减小。也就是说,学习判断的词性线索削弱了词语情感属性线索的作用”。同时,需要说明的是,本研究采用学习时间和加工深度作为操控编码强度的方式,并通过自由回忆成绩考察操控编码强度的有效性。

意见 12: 还有一些细节问题需要修改。如英文摘要质量可以进一步提高,如有些句子过长表意不清,存在拼写错误等。还有病句,表达不够流畅的情况。格式也存在一些问题,如(Vuorre, 2017)在参考文献中没有引用等,希望作者认真检查修改。

回应: 感谢审稿老师的意见。修改之处以蓝色标记,详见P31-P32。

已通读全文,仔细按照APA格式对全文进行修改。为了更好的完善英文摘要,邀请伦敦

大学学院(University College London)心理学与语言学部的一名副研究员(心理学博士,研究方向元记忆领域)对英文摘要内容以及语言进行了修改。为了进一步确保英文拼写正确,使用专业的英文核对软件(Grammarly)进行了核查。

第二轮

审稿人 2 意见:

经过修改,文章质量有了较大的提升,但仍存在一些问题需要作者考虑:

意见 1: 修改稿中,作者通过与真假词判断任务进行比较,对知觉范式连续识别任务(CID)的优势进行了较为清楚的阐述。但读者想了解的是相对于其他任务,CID 任务对知觉流畅性的测量有何优势,这个比较的对象并不特指某一个任务(如真假词判断任务)。作者不必对真假词任务展开分析(研究中并未涉及),可以考虑提炼一下 CID 的优势。

回应: 感谢审稿老师的意见。在修改稿中 P14,以红色标记,我们重新组织了 CID 任务对项目间知觉流畅性的测量优势。具体修改内容如下:

“为进一步考察知觉流畅性在字体大小效应中的作用, Yang 等人(2018)直接比较了 CID 任务与真假词判断任务测定项目知觉流畅性的有效性。CID 任务是一种视知识识别任务,被试需要在识别出呈现的单词后立即按键反应(Ward, Berry, & Shanks, 2013)。真假词判断任务要求被试尽可能快而准确地辨别屏幕上呈现的字符串为真词或假词,并作出相应的按键反应。结果发现,在 CID 任务中,所有被试识别大字体单词的速度均显著快于小字体单词。然而,在真假词判断任务中,大字体单词和小字体单词的判断反应时却没有差异。Yang 等人(2018)认为造成这种测量结果差异的主要原因是, CID 任务仅涉及个体对项目的知觉加工过程,所获取的连续识别反应时并未受到额外因素(例如,语义加工或按键选择)的潜在干扰(Chumbley & Balota, 1984; Yap, Sibley, Balota, Ratcliff, & Rueckl, 2015)。以上结果表明, CID 任务对知觉流畅性的测量具有更高的灵敏度。”

此外,需要说明的是,知觉流畅性的测定方式包括两个方面,即主观的自我报告和客观的判断反应时和判断准确率。目前,直接与字体大小效应相关的研究中,常采用真假词判断任务和 CID 任务测量知觉流畅性的客观指标(e.g., Mueller, Dunlosky, Tauber, & Rhodes, 2014; Undorf & Zimdahl, 2019; Yang, Huang, & Shanks, 2018),并且这两种范式测量的反应时指标分别作为支持各自理论的重要证据(加工流畅性理论和信念理论)。我们希望通过两种范式的差

异比较，使读者不仅理解 CID 任务测量知觉流畅性的有效性，而且帮助理解信念理论(否定流畅性线索对学习判断的影响)和加工流畅性理论(并不否定在字体大小和学习判断之间可能也存在其他线索作用，例如信念)之间的关系。

综上所述，我们重新组织了对两种任务的描述，将叙述重点转移为 CID 任务获取的反应时仅为项目的知觉加工过程，排除额外反应时的潜在干扰。请审稿老师把关，再次感谢您的建议。

参考文献：

- Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font size effect on judgments of learning: does it exemplify fluency effects or reflect people's beliefs about memory?. *Journal of Memory and Language, 70*, 1–12.
- Undorf, M., & Zimdahl, M. F. (2019). Metamemory and memory for a wide range of font sizes: What is the contribution of perceptual fluency?. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(1), 97–109.
- Yang, C., Huang, S. T., & Shanks, D. R. (2018). Perceptual fluency affects judgments of learning: the font size effect. *Journal of Memory and Language, 99*, 99–110.

意见 2：对于研究的实际应用价值的探讨，此次修改作者进行了进一步的梳理和澄清。作者在相关问题的回复中提到了陈颖等人(2019)的研究，并进行了对比分析。正文中，前言对实际应用价值做了弱化处理，在讨论中指出“实验 2 和实验 3 发现激活深层线索是矫正基于知觉特征线索(实验 1)引起的元认知错觉的有效途径，该结果对提高个体学习判断准确性和帮助学习者制定合理的学习计划具有重要的应用价值。”但这个部分的论述感觉没说透，往后翻也没找到对实际应用价值的进一步阐述，仅在最后一段看到对不足和展望的分析。

回应：感谢审稿老师的意见。在修改稿中对本研究的实际应用价值进行了修改和补充(详见 P26，以红色标记)，具体修改内容如下：

“相比于其他类型的知觉线索(例如，文本清晰度和分辨率, Strukelj, Scheiter, Nyström, & Holmqvist, 2016; 字形和呈现方式, Rummer, Schweppe, & Schwede, 2016), 字体大小线索在实际教学情境中出现的频率更高(Magreehan, Serra, Schwartz, & Narciss, 2016)。所以，字体大小所引发的元认知监测错觉对自我调控学习的潜在影响更应引起关注。例如，在教学材料排版上，教育工作者不仅要考虑使用不同字体大小强调知识要点，还应当考虑如何避免由于字体大小而引起的学习判断错觉。本研究结果表明，促进学习者进行更深层次的语义加工是一种消除字体大小效应错觉的有效措施。因此，教育工作者可以考虑在不同字体大小的知识内

容后面添加测试题(如, 造句或样例提名等), 以促进学习者对知识要点进行更深层的语义加工并激活更多的语义线索, 从而削弱或消除由于知觉特征而引起的学习判断误差。”

意见 3: 修改稿中对于编码强度发挥作用的相关讨论是合理的。作者在前言中提到了线索利用模型 (“根据线索利用模型的观点, 学习判断出现偏差的主要原因是个体进行学习判断时使用了与真实记忆强度无关的线索”), 在讨论部分也阐述了线索整合理论(Undorf 等人(2018)的研究)。但相关理论在问题提出的时候是不是就应该有所铺垫?

回应: 感谢审稿老师的意见。在修改稿中, 我们分别在前言和讨论中更改了对线索整合与竞争观点的描述(详见修改稿的 P14 和 P27, 以红色标记), 其主要观点就是, 学习判断中存在对多种线索的利用, 并且个体的学习判断会将各种线索进行策略性整合(Peynircioğlu & Tatz, 2019; Undorf & Bröder, 2019; Undorf, Söllner, & Bröder, 2018)。目前, 该观点仍处于发展阶段, 本研究的结果为该观点提出了支持性的证据。同时, 在前言中, 我们补充了本研究通过深层线索的激活矫正基于知觉特征引发的元认知错觉的理由, 分别从线索利用模型和学习判断中存在多种线索整合与竞争的观点切入。具体内容如下:

“近期研究发现, 人们可以整合多种线索用于学习判断(Undorf & Bröder, 2019; Undorf, Söllner, & Bröder, 2018)。字体大小效应的产生是否与除知觉流畅性以外其他深层线索的激活不足有关? 对以往字体大小效应的研究进行分析, 研究者发现这种元认知错觉的产生还可能与学习时编码强度(encoding strength)不足、诊断性线索(diagnostic cues)的激活不充分有关。在以往研究中, 被试的自由回忆正确率均较低(0.12 ~ 0.25)。例如, Rhodes 和 Castel (2008) 要求被试学习每个单词 5s 并对所学单词进行学习判断, 最后参加自由回忆测试。在测试中, 个体的自由回忆正确率仅为 0.16, 相似的结果在 Mueller 等人(2014)和 Yang 等人(2018)的研究中重复出现。需要指出的是, 这种编码强度问题并不是由于学习时间不足所导致。Rhodes 和 Castel (2008)的实验中, 每个单词的学习时间已长达 5s, 但是个体可能并没有充分利用所有的学习时间对单词进行持续性编码。换句话说, 被试的编码时间并不总等于所学项目的呈现时间。被试可能在所学项目消失之前就提前结束了编码加工, 或是在长时间的学习过程中出现注意力不集中或心智游移(mind wandering)等现象(Cooper & Pantle, 1967; Mueller, Dunlosky, Tauber, & Rhodes, 2014; Seli, Risko, Smilek, & Schacter, 2016)。

根据线索利用模型的观点, 学习判断出现偏差的主要原因是个体在做出学习判断时利用了与真实记忆强度无关的线索(Benjamin, Bjork, & Schwartz, 1998; Koriat, 1997)。当编码强度较弱时, 学习材料的知觉特征线索容易对学习判断产生误导作用, 进而出现字体大小效应错

觉。增强信息编码强度能够激活更多和更有效的线索(如语义线索)，这些线索与个体真实的记忆强度相关且诊断性更高，此时则弱化了知觉特征线索对学习判断的影响。为验证该推断，本研究将分别考察增加有效学习时间(并要求被试在学习过程中进行复述，以确保被试充分利用学习时间对信息进行持续性编码加工)和提高语义加工深度(即要求被试对学习词进行造句)能否有效减少或消除字体大小效应。”

主要参考文献：

Peynircioğlu, Z. F., & Tatz, J. R. (2019). Intensifying the intensity illusion in judgments of learning: Modality and cue combinations. *Memory & cognition*, 47(3), 412–419.

Seli, P., Risko, E. F., Smilek, D., & Schacter, D. L. (2016). Mind-wandering with and without intention. *Trends in cognitive sciences*, 20(8), 605–617.

Undorf, M., & Bröder, A. (2019). Cue integration in metamemory judgements is strategic. *Quarterly Journal of Experimental Psychology*, 73(4), 629–642.

Undorf, M., Söllner, A., & Bröder, A. (2018). Simultaneous utilization of multiple cues in judgments of learning. *Memory & Cognition*, 46(4), 507–519.