

《心理学报》审稿意见与作者回应

题目：多媒体学习中线索效应的元分析

作者：谢和平, 王福兴, 周宗奎, 吴鹏

第一轮

审稿人 1 意见：

意见 1：前言部分没有针对“注意引导”的研究结果不同的原因做出分析。

回应：感谢审稿专家的建议。与线索在学习效果上的作用类似，线索是否具有注意引导作用也没有得出特别一致的结论。关于“注意引导”的研究结果不同的原因，目前并没有很明确的解释。但是有研究者猜测，这可能与线索的相对凸显性有关。例如：当学习材料中的元素具有运动属性时，线索的凸显性容易被材料元素的运动性所抵消，因为材料中元素本身的运动也如线索一样具有很强的注意引导作用，这可能导致线索和运动的元素之间产生注意资源的竞争，线索的凸显性便相对降低。本文引言部分确实没有针对“注意引导”的研究结果不同的原因做出分析，因此我们在 1.2 部分进行了补充解释。摘录如下：

即线索没有引导注意。对于线索在注意引导上的不同结果，有研究者猜测这可能与线索的相对凸显性有关(Lowe & Boucheix, 2011)。例如：当学习材料中的元素具有运动属性时，线索的凸显性容易被材料元素的运动性所抵消，因为材料中元素本身的运动也如线索一样具有很强的注意引导作用，这可能导致线索和运动的元素之间产生注意资源的竞争，线索的凸显性则相对降低，不利于发挥线索在注意引导上的有效作用。

结合第二位审稿人的意见 1，且为尽量保证行文通顺性及段落的协调性，我们在 1.2 部分第一自然段的“不过，仍有少量研究未发现线索能够有效提高学习者的注视加工”这一句子前进行了分段处理(即另起一段)。具体见修改稿 1.2 部分。

意见 2：不如将“1.3.1 特征线索”改成“1.3.1 线索类型”。

回应：根据建议，已将标题“1.3.1 线索特征”更改为“1.3.1 线索类型”。相应地，为保障行文流畅性，对 1.3.1 部分的首句做了相应删减和更正，具体见正文红色字体。

意见 3：没有指出选用检索关键词的出发点，也没有给出选择“时间跨度为 1995 年 1 月至 2015 年 5 月”的原因。这会给读者带来很多疑问。例如，既然“cueing effect”是一个很重要的关键词，为何没把该词作为检索词？选择“animation”的目的是什么？该词在前言中是未涉及的。

回应：文献检索的最终目的在于尽可能全面地搜索到有关元分析研究主题的文献，这与检索关键词、检索时间、数据库等的选定密切相关。本研究中，文献检索关键词的选取主要基于以往关于多媒体学习线索效应研究中的标题、摘要、关键词等来确定，查阅发现“cue”、“cueing”、“signaling”、“attention guidance”、“multimedia learning”、“animation”、“eye tracking”、“eye movement”等是相关文献中经常出现的关键词汇(中文文献与英文文献类似)。因此本研究以上述词语为检索关键词。但是如专家所言，本研究正文中没有指出选用这些检索关键词的出发点，故作者在 2.1 部分增加了检索关键词的选用出发点。摘录如下：

检索关键词的选取主要基于以往关于多媒体学习线索效应研究中的标题、摘要、关键词等来确定……关键词的选取标准及文献检索时间跨度同上……

对于检索的时间跨度，截止时间即为作者检索文献的时间 2015 年 5 月，而起始时间之所以选择 1995 年 1 月主要出于三个方面的考虑：(1)在仔细阅读国内外有关线索效应的综述(De Koning et al.,

2009; 王福兴等, 2013)及其引用的参考文献后, 我们发现最早就多媒体学习环境中线索的作用进行探讨的实证研究可能为 Kalyuga, Chandler 和 Sweller (1999)的研究; (2)为避免遗漏, 作者将检索的起始时间适当地向前推移几年, 且为了大致搜索近 20 年的研究文献, 故将起始时间初步定为 1995 年 1 月; (3)之后在文献回溯的过程中, 我们也对所搜集到的合适文献中的参考文献进行了检索, 暂未发现关于多媒体学习的线索效应的更早实证研究, 因此最终将起始时间定为 1995 年 1 月。本文在 2.1 部分就搜索时间跨度的选择进行了补充说明。摘录如下:

纵观以往研究发现, Kalyuga 等人于 1999 年较早探讨了多媒体环境中加入线索的作用, 为尽量避免遗漏, 本研究文献检索起始时间稍作提前, 时间跨度为 1995 年 1 月至 2015 年 5 月近 20 年。

至于关键词“cueing effect”的问题, 作者承认该词比较重要, 但并未将其作为文献检索的关键词, 原因在于: (1)确定的检索关键词中已有“cueing”一词, 从文献检索的角度来讲, “cueing”要比“cueing effect”一词更为宽泛, 即: 用“cueing”检索到的文献理应包含用“cueing effect”检索到的文献, 因此为避免重复检索而增大不必要的工作量, 作者未将“cueing effect”一词作为检索词; (2)以往关于线索效应的研究论文中使用“cueing effect”作为文章关键词的频率并不是特别高, 更多采用“cue”、“cueing”、“signaling”等词, 因此为慎重起见, 我们使用了范围更广的“cueing”一词作为检索词。

之所以选择“animation”一词作为关键词之一, 主要是由于近年来随着技术的发展, 动画等动态材料在多媒体学习中的作用愈发凸显(Höffler & Leutner, 2007; Tversky, Morrison, & Betrancourt, 2002), 越来越多研究者采用动画形式的多媒体学习材料作为其研究的正式实验材料, “animation”一词也经常以关键词的形式出现在研究中(Amadiou, Mariné & Laimay, 2011; Kriz & Hegarty, 2007; Lin, 2011; Skuballa, Schwonke, & Renkl, 2012; Spanjers, Van Gog, Wouters, & Van Merriënboer, 2012)。因此本研究也选择了“animation”一词作为检索关键词。不过, 该词在引言中确实疏于提及, 可能不助于读者理解, 因此作者在 1.3.2 部分做了增补。摘录如下:

材料动态性可分静态(static)和动态(dynamic)两种(Tversky et al., 2002)。静态材料在教学中较为普遍, 其元素往往以静止的形式呈现, 例如: 静态的 PPT 材料; 而随着技术发展, 动态材料的运用也愈发广泛, 其元素在空间位置上往往具有运动性, 例如: 动画(animation)。虽然线索……

意见 4: 排除标准中第二条的目的不清晰, 即为何要排除“仅含画面或仅含语词的研究”?

回应: 根据 Mayer (2009)的定义, 多媒体学习是指从画面(picture)和语词(word)材料中进行的学习。由此, 多媒体学习材料应同时包含两个基本元素——画面和语词。缺乏其中任何一个基本元素, 则不属于多媒体学习材料。由于本研究关注多媒体学习环境中的线索效应, 所以排除“仅含画面或仅含语词的研究”的目的在于保证所纳入研究的学习环境均为多媒体学习。

2.2 部分的第二个标准已对排除“仅含画面或仅含语词的研究”的目的略有介绍, 但可能并不直截了当。因此如专家所言, 有必要在该处进一步介绍清楚。摘录如下:

(2)研究中使用的学习材料应属于多媒体学习材料, 即包含画面(如: 图表、照片、视频、录像、动画)和语词(如: 视觉文本、声音解说)两个基本元素(Mayer, 2009), 二者缺一则不属于多媒体学习范畴, 故仅含画面或仅含语词的研究予以排除。

意见 5: “95% CI”应该与段落中的 95%置信区间表达一致, 否则应对缩写有所标示。即便标示也应该在第一次出现时说明, 而不是后面的表 5 才说明。

回应: 非常感谢专家的细致审阅。文中“95% CI”的标示以及段落中 95%置信区间的表达确实不合规范。

从全文来看, “95% CI”主要在表 2、表 4 和表 5 中出现。虽然所有的 CI 均代表置信区间, 但它们每次出现时所指代的对象均稍有变化。因此, 我们在表 2、表 4 和表 5 下边重新作了补充注明或修改。为统一表述, 我们将段落中的“置信区间”一词统一改为“CI”。各表下的注明摘录如下:

[表 2] 此处 95% CI 指各结果变量对应的效应量 g 的 95%置信区间(含下限和上限)。

[表 4] 此处 95% CI 指各调节变量对应的效应量 g 的 95%置信区间(含下限和上限)。

[表 5] 此处 95% CI 指 Egger's intercept 的 95% 置信区间(含下限和上限)。

意见 6: 有些结果的报告太突然, 如“3.3 异质性检验”部分的“I2 值”是干什么的? 是如何得来的? 此后的 Q_B 、失安全系数(Rosenthal's N_f s)等同样存在类似的问题。

回应: 原文对这类统计指标没有做明确介绍, 读起来的确比较突兀。为此, 我们同意分别在文中合适的地方对这些统计指标或方法(Q 检验、 I^2 检验、失安全系数、Egger 线性回归检验、剪补法)进行简单介绍。为方便审稿专家理解, 现对此进行详细说明。

首先, 关于异质性检验。一般情况下, 纳入到同一个元分析中的所有研究都会在不同的方面(例如: 被试、材料等)存在差异, 因此在统计上通常将元分析中不同研究间的各种变异统称为异质性(heterogeneity)。按照元分析的统计原理, 只有具有较好同质性的资料才能进行合并, 因此在元分析时需要多个研究的结果进行异质性检验(heterogeneity test)。异质性检验的目的主要有二: (1)判断模型的选定是否合理, 如果效应量异质, 表明适合选择随机效应模型进行元分析; (2)判断是否需要进一步进行亚组分析(即调节效应检验), 如果效应量异质, 表明各研究效应量间的变异来源除了抽样误差外, 还可能来自其他来源(例如: 被试性别), 这些来源可能对效应量起到调节的作用(例如: 男性和女性可能存在显著不同的效应量), 此时需要将所有纳入元分析的数据分成更小的单元(例如: 男性、女性)以尽量保证各单元的同质性, 从而进一步进行比较。而 Q 检验和 I^2 检验则是识别异质性的两种具体的、不同的方法。 Q 检验具有统计学的显著性, 以 p 值形式呈现, 通常情况下如果 $p < 0.05$, 则表明具有异质性。至于 Q_B , 它是指在亚组分析时对组间的异质性进行的 Q 检验, 用于检验亚组之间的效应量是否存在差异(例如: 男性的效应量是否显著大于女性的效应量)。而 I^2 检验反映的是由效应量的真实差异造成的变异在总的变异中所占的比重, 以 I^2 值的形式呈现, 一般将 25%、50%、75% 的 I^2 值作为异质性低、中、高的界限(Higgins, Thompson, Deeks, & Altman, 2003)。如果 I^2 值较低, 则可大致认为各研究间的效应量同质, 否则异质。

其次, 关于发表偏差检验。失安全系数(Rosenthal's fail-safe number, N_{fs})、Egger 线性回归检验(Egger linear regression test)以及剪补法(trim and fill method)是检验发表偏差的常用指标或方法。失安全系数是由 Rosenthal 在 1979 年提出的一种指标, N_{fs} 值反映的是: 当元分析结果在统计上显著时, 还需要多少项没有统计学意义的研究才能让该元分析结果变得不显著。该值越大表明存在发表偏差的可能性越小, 元分析的结果越稳定, 结论被推翻的可能性也越小。有研究者以“ $5k + 10$ ”(k 指纳入元分析的独立效应量个数)作为是否存在发表偏差的界限(Rothstein, Sutton, & Borenstein, 2006), N_{fs} 值如果大于 $5k + 10$, 则可认为不存在发表偏差。Egger 线性回归检验采用线性回归法检验发表偏差, 该方法以精确性为自变量, 以标准正态离差为因变量建立回归方程。理论上, 如果纳入元分析的研究不存在发表偏差, 则回归方程的截距(Egger's intercept)应为 0。但实际操作中, 该截距基本不为 0, 此时一般会求出线性回归方程的截距及其 95% CI, 再对该截距是否为 0 进行假设检验, 如果不显著, 则表明不存在发表偏差。剪补法的基本思想是首先剪掉漏斗图中不对称的部分(漏斗图是评价发表偏差的直观方法, 但因其不够客观, 故本研究未选用该方法检验发表偏差), 用剩余的对称部分估计漏斗图的中心值, 然后沿着中心两侧粘补上被剪切部分以及相应的遗漏部分(如图 1 中的实心圆点), 最后基于粘补后的漏斗图重新估计合并效应量的真实值。在元分析中运用该方法检验发表偏差时, 研究者主要判断剪补后效应量的显著性与剪补前是否一致即可, 一般认为如果剪补后的效应量并没有发生显著变化, 则可认为不存在发表偏差。

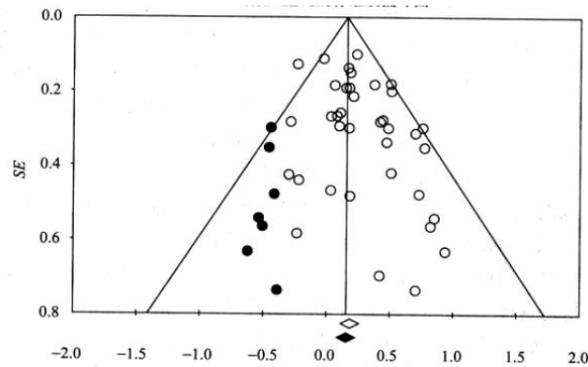


图1 剪补法在漏斗图中的示例(选自: Borenstein et al., 2009, p213, 仅供理解)

异质性检验和发表偏差检验的统计数值均会在 CMA 软件中给出。

为尽量避免结果的报告过于突然, 修改稿中, 我们在方法的 2.5 部分对发表偏差检验指标或方法进行了补充说明, 并对结果的 3.5 部分进行了相应的增减处理。另外, 也在 2.6 部分的第一自然段对异质性检验指标或方法进行扩充介绍, 并对 3.3 部分也进行了相应增减(Q_B 于 3.3 部分进行了补充说明)。摘录如下:

2.5 发表偏差

元分析过程中, 如果实际纳入的研究与所有理应被纳入的研究之间存在系统性误差, 则认为出现了发表偏差(Borenstein et al., 2009)。一般来说, 有统计学意义的研究更容易发表, 而发表的研究更容易被纳入元分析。这可能会导致对综合效应的高估, 从而影响单个元分析结果的真实性(Franco, Malhotra, & Simonovits, 2014; Kuppens, Laurent, Heyvaert, & Onghena, 2013)。本研究采用常用的失安全系数(Rosenthal's fail-safe number, N_{fs})、Egger 线性回归检验(Egger linear regression test)以及剪补法(trim and fill method)评价发表偏差。失安全系数 N_{fs} 值反映的是: 当元分析结果在统计上显著时, 还需要多少项没有统计学意义的研究才能让该元分析结果变得不显著。该值越大表明存在发表偏差的可能性越小, 元分析的结果越稳定, 结论被推翻的可能性也越小。有研究者以“ $5k + 10$ ”(k 指纳入元分析的独立效应量个数)作为是否存在发表偏差的界限(Rothstein, Sutton, & Borenstein, 2006), N_{fs} 值如果大于 $5k + 10$, 则可认为不存在发表偏差。Egger 线性回归检验采用线性回归法检验发表偏差, 一般先求出线性回归方程的截距(Egger's intercept)及其 95% CI, 再对该截距是否为 0 进行假设检验, 如果不显著, 则表明不存在发表偏差(Egger, Smith, Schneider, & Minder, 1997)。剪补法主要通过先剪后补的方式使得各研究在平均效应量的左右两边尽量对称分布, 并重新估计合并效应量的真实值(Duval & Tweedie, 2000), 若剪补后的效应量未发生显著变化, 则可认为不存在发表偏差。

.....

2.6 模型选定及软件

.....此外, 研究通过事后异质性检验(heterogeneity test)进一步验证模型选择的合理性。选择 Q 检验和 I^2 检验识别异质性。 Q 检验具有统计学的显著性, 以 p 值形式呈现, 若 $p < 0.05$, 则表明显著异质。而 I^2 检验反映的是由效应量的真实差异造成的变异在总的变异中所占的比重, 以 I^2 值的形式呈现, 该值越大表明异质性程度越高。根据 Higgins 等人(2003)的观点, 可将 25%、50%、75% 大致可看作异质性低、中、高的界限。若效应量显著异质且程度并不低, 则表明选择随机效应模型合理。

.....

3.3 异质性检验

分别对 4 种结果变量进行异质性检验(见表 3)。结果发现 Q 检验均显著, 这表明效应量显著异质。对于保持测验和迁移测验, I^2 值分别为 79.39%、67.99%, 说明保持测验和迁移测验中由效应量的真实差异造成的变异各占总变异的 79.39%、67.99%; 对于注视时间和注视次数, I^2 值分别为 43.57%、62.66%, 表明在对线索区的注视时间和注视次数上由效应量的真实差异造成的变异各占总变异的 43.57%、62.66%。从异质性低、中、高的界限(Higgins et al., 2003)来看, 本研究 4 类结果变量中效应量真实差异所占比重属于接近中等及中等以上的异质性, 进一步表明异质性程度较高。故本研究采用

随机效应模型合理。此外，效应量异质还意味着线索对学习的影响可能存在潜在的调节变量(Cooper, 1989)。因此，需进一步进行亚组分析(调节效应检验)，即对亚组间的异质性进行 Q 检验(Q_B)，用于考察亚组之间的效应量是否存在差异。

.....

3.5 发表偏差检验

采用以下方法评价发表偏差(见表 5)。从失安全系数(Rosenthal's N_{fs})来看，保持测验和迁移测验分别为 1737 和 849，注视时间和注视次数分别为 141 和 185，均远远大于“ $5k + 10$ ”，据此，该指标表明本研究并不存在显著的发表偏差。从 Egger 线性回归检验来看，在保持测验上，该指标的值为 2.03，不接近 0，且 95% CI 不包含 0， p 值也小于 0.05，暗示保持测验可能存在发表偏差；但在迁移测验、注视时间、注视次数上，该指标的值分别为 0.76、-0.66、1.02，接近 0，且 95% CI 均包含 0， p 值大于 0.05，暗示迁移测验、注视时间、注视次数存在发表偏差的可能性较小。针对以上检验结果(特别是保持测验)的差异，进一步采用剪补法对合成效应量左右两边的文献进行剪补后，发现效应量仍显著。综上，可以认为本研究并不存在较大的发表偏差。

意见 7：表 4 不是三线表。

回应：感谢专家的细致审稿。表 4 的画法确实并非严格的三线表，例如：在该表中间部位显示了细横线。本文中之所以未严格画成三线表，主要出于以下考虑：该表中包含 4 类结果变量(保持测验、迁移测验、注视时间、注视次数)，而每类结果变量下又含有多个基本一致的调节变量，如果不在每个结果变量之间加入横线以示类别，可能阅读起来显得杂乱无章，因此为了方便阅览该表，我们以结果变量为单位，在其中加入细横线以将各类结果变量区分开来。

另外，原稿中表 4 最后几行的边框看起来虽不符合三线表格式，但经排查，这并非人为造成，而是由于表 4 过长，转页时自动生成所致。合理地调整该表在文中的位置，应该不会出现该问题。

意见 8：应该添加适当的理论描述。此外，应该对今后研究的方向和实践启示给予一定的说明。

回应：同意审稿人的建设性意见。本研究在理论上的贡献主要表现为从元分析的角度直接支持了线索的注意引导假说、图文整合假说，并间接支持了认知负荷理论。虽然作者在原文 4.1 部分的第二自然段对相关的理论有所描述，但可能内容不太清晰突出，表述不够深入完整，为此，我们对该段进行了重新组织和表述，以期尽量提高本元分析的理论意义，为读者提供理论参考。摘录如下：

从理论上讲，线索在注意引导和学习效果上的积极作用可结合注意引导假说(attention-guiding hypothesis)、图文整合假说(text-diagram integration hypothesis)和认知负荷理论(Cognitive Load Theory, CLT)来进行解释。注意引导假说认为，线索能够在早期的知觉加工(perceptual processing)阶段有效发挥对信息的选择功能(De Koning et al., 2009)，即：线索可将学习者的注意力引导到特定的关键区域，增加对这些区域的注视时间和注视次数，忽视任务无关区域，从而为后期的深入理解提供保障(Jamet, 2014; Jian et al., 2014; Ozcelik et al., 2009)。该假说得到了本元分析中眼动指标(注视时间和注视次数)的直接支持。图文整合假说认为，仅仅注意到被线索化的信息并不能保证一定会提高最终的学习成绩，实现有效的学习还需要学习者进行更深入的建构性的认知活动。而线索还具有组织和整合的功能(De Koning et al., 2009)，它的加入能够在后期的认知加工(cognitive processing)阶段进一步强调教学材料中图文元素内部及图文元素之间的组织或逻辑结构(Mautone & Mayer, 2001, 2007)，进而将其准确地整合成连贯而一致的心理表征(Scheiter & Eitel, 2015)，这种整合作用对于学习者的识记和理解都具有重要意义。该假说也得到了本元分析中学习效果指标(保持测验和迁移测验)的直接支持。另外，根据认知负荷理论，个体在信息加工过程中的认知资源是有限的，如果超出了认知资源总量，个体的学习效率就会受到影响(Sweller, Van Merriënboer, & Paas, 1998)。由于多媒体学习材料的相对复杂性，学习过程中涉及较多的认知资源消耗，因此存在超出认知资源总量的较高可能或风险。而在材料中合理地加入线索后，线索可能通过上述选择、组织和整合信息的方式降低外在认知负荷(extraneous cognitive load)，从而提高学习效果(De Koning et al., 2009; 王福兴等, 2013)。不过，由于本研究并未直接考察线索对认

知负荷的影响,因此仅从间接角度支持了认知负荷理论。总之,本研究或直接或间接地支持了注意引导假说、图文整合假说和认知负荷理论,线索可能通过信息选择、组织和整合的方式来降低学习者的外在认知负荷,从而保证对学习材料良好的识记和理解效果。

对于实践启示及今后研究的方向,根据专家建议,我们在4.3部分的第二自然段给予了补充说明。摘录如下:

线索研究在多媒体学习中不仅具有重要的理论意义,还具有较强的教学实践价值。例如:在多媒体教学材料设计过程中,建议合理地加入线索以促进对信息的选择、组织和整合,降低认知负荷,特别是当材料属于静态、陈述性或理工科时,更应加入线索以使学习效果最优化;从重视最终学习成绩(特别是迁移成绩)的角度来看,教学设计时鼓励加入时空线索(而非物理线索)以更好地提高学习者对材料的理解深度。未来研究可考虑从以下角度深入开展:(1)继续采用眼动技术考察线索在不同条件下的注视加工作用,同时注重探讨眼动注视和学习效果之间的相互关系(Boucheix & Lowe, 2010; Scheiter & Eitel, 2015),这样既可深入了解线索对学习过程中注意分配模式的影响,也可为后续元分析提供充足的数据来源;(2)重视探讨不同的学习环境、学习材料、学习者经验水平等其他潜在调节变量对线索作用的影响,从而考察线索效应的其他边界条件;(3)尽管本元分析在某些潜在的调节变量(如:呈现通道、呈现步调)上并未得出显著的调节效应,但这不应成为后续研究继续探讨这些变量的绊脚石,未来研究亦可从其他改进的角度(例如:采用更为严格的实验设计)对这些变量的作用进行深入挖掘。

审稿人2意见:

本研究对多媒体学习的线索效应进行了元分析,进一步为线索效应提供了直接稳健的证据,具有重要的意义。本文写作流畅,思路清晰,文献完备可靠,方法适当。但是尚有一些小问题。

意见1: 1.1-1.3的总结,建议合并移到“方法”前。目前的组织显得碎片化。

回应: 本文当前的组织略显碎片化。根据审稿专家建议,作者同意将1.1-1.3部分总结性的段落(主要包括1.2部分的最后一个自然段,以及1.3部分第一自然段的后半部分)进行合并,统一置于新增的“1.4本研究假设”中,并整体放在“方法”前。为尽量行文流畅,对部分语句进行修改,且将1.3.1部分、1.3.2部分和1.3.3部分中的研究预期也统一移至1.4部分中。摘录如下:

1.4 本研究假设

综上,一方面,线索在学习效果和注意引导上的积极作用并不是绝对稳定的(中文综述见:王福兴等,2013)。针对线索作用的这种非稳健性,本研究拟采用元分析来探讨加入线索与否对多媒体学习效果和注意引导将产生怎样的影响。在多媒体学习领域,由于研究者多采用保持测验或迁移测验考察学习者的学习效果(Mayer, 2009),眼动研究多采用线索区的注视时间或注视次数考察线索的注意引导作用(Boucheix & Lowe, 2010; De Koning et al., 2010a),因此本研究分别以保持测验、迁移测验作为学习效果的结果变量,以线索区的注视次数、注视时间作为注意引导的结果变量。根据线索的信息选择、组织和整合功能(De Koning et al., 2009),研究推测:加入线索能够促进学习者的学习效果,表现为更高的保持测验和迁移测验成绩;同时线索也能够有效引导学习者的注意,表现为更长的注视时间和更多的注视次数。

另一方面,线索效应可能受到线索类型、材料特征以及学习者特征等潜在变量的调节,因此有必要对线索效应进行调节效应检验。由于元分析技术不仅可以合成总的效应量,同时也能够根据该效应量的影响因素对其进行亚组分析(subgroup analysis),即可将影响该效应量的因素看作调节变量进行分组对比(Borenstein, Hedges, Higgins, & Rothstein, 2009),所以能较好地实现该研究目的。研究推测:线索效应受线索类型、材料特征(包括材料动态性、呈现通道、呈现步调、知识类型、学科属性)和学习者特征(即教育水平)等因素的调节。

意见2: 引言第二段:“线索是…教学技术”,定义不太通顺。

回应: 感谢专家的细致审稿。作者认真查阅De Koning等人(2007, 2009)研究中对线索的定义及解释,

并经慎重考虑后认为：多媒体学习中的教学技术更多地是一种旨在提高学习效果的处理方式或手段，它应该体现出一种动态的过程，例如：加入线索。而“线索”本身只是一种非内容信息，它是静态的，并不能反映教学技术的动态性。因此，我们在引言的第二自然段对线索的定义进行了纠正，将落脚点放在了“非内容信息”上，而非“教学技术”。摘录如下：

线索是指在多媒体学习材料中加入的一种能够吸引学习者的注意力，并旨在促进学习者对材料的选择、组织和整合的非内容信息(non-content information, 如：箭头、颜色、高亮等)。

这样，线索的定义变为“线索是...一种...非内容信息”的格式。另外，我们对引言的第一自然段和第二自然段中的部分表述也进行了相应的更正，见文中的红色字体文字。

第二轮

审稿人 1 意见：

意见 1：前言中应该有较为深入的理论（如注意引导假说等观点），而不是仅在讨论部分涉及。

回应：感谢审稿专家的建议。引言中的理论依据确实存在一些不充分的地方。因此，在修改稿的 1.1 部分和 1.2 部分中，我们分别加入了与线索效应直接相关的两个理论假说(图文整合假说和注意引导假说)，以突出前言部分的理论观点，充实内容，使之更加符合逻辑。修改的部分内容摘录如下：

线索对学习效果的促进作用从理论上得到了图文整合假说(text-diagram integration hypothesis)的直接支持(Scheiter & Eitel, 2015)。该假说认为，线索具有信息组织和整合的功能(De Koning et al., 2009)。在多媒体教学材料中合理地加入线索，有利于对材料中的图文元素本身及图文元素之间的内在逻辑关系加以强调(Mautone & Mayer, 2001, 2007)，促进学习者将图文信息整合成为连贯、一致的心理表征(Scheiter & Eitel, 2015)，从而实现更为深入的、建构性的认知加工，最终提高学习成绩。

.....

线索在眼动加工中的优势从理论上得到了注意引导假说(attention-guiding hypothesis)的直接支持(Jamet, 2014)。该假说认为，线索具有信息选择的功能(De Koning et al., 2009)。在多媒体学习材料中合理地加入线索可增强关键信息的知觉凸显性，这有助于迅速将学习者的注意力引导到这些被线索化的关键信息上(Ozcelik et al., 2010; Ozcelik et al., 2009)，从而提高学习者对这些信息的注意维持倾向，即对它们投入更多的(视觉)注意资源并促进其加工(Boucheix & Lowe, 2010; Jamet, 2014; Jian et al., 2014)。

意见 2：首次提出保持测验和迁移测验时，应该对他们做出对比说明。

回应：由于保持测验和迁移测验是本研究所选取的两个重要的结果变量，从逻辑上而言确实需要对其进行适当的解释，根据审稿人的建议，在二者首次提出时做出对比说明。但为了尽量不影响文章的通顺性，作者以脚注形式在 1.1 部分(正文第 1 页)进行说明。脚注内容摘录如下：

¹ 保持测验和迁移测验均为多媒体学习研究所关注的与学习效果有关的重要后测指标，二者在测验方式和目的上有所不同：保持测验通常要求学习者学习完后对材料内容尽可能多地回忆，测验答案多为学习材料本身，因该测验可能存在死记硬背现象，故主要用以考察学习者的浅层识记效果；而迁移测验通常是在学习者学习完之后要求其回答或解决与材料相关的另一些迁移性问题，测验答案需要学习者从材料中加以推理，该测验能较好地避免死记硬背现象，因此主要用来考察学习者的深层理解效果。

意见 3：眼动研究和非眼动研究之间的异同性最好再深入一下，因为行文在摘要等部分中似乎将这两个不同的角度区别来说明的，并没有将他们关联到一起。但结论部分似乎并没有从眼动研究和非眼动研究异同的角度表达。这样操作的意图是什么？

回应：线索效应的非眼动研究主要探讨线索与学习效果之间的关系，而线索效应的眼动研究一般来说既探讨了线索与学习效果之间的关系，同时也考察了线索对视觉注视加工的影响。因此相同之处在于：

二者都考察了线索与学习效果的关系；不同之处在于：眼动研究还采用眼动技术考察了线索对视觉注意的引导。

如果将眼动研究和非眼动研究关联到一起，则自然而然会引出一个比较重要的问题：注意引导和学习效果之间是什么关系？结合本研究来看，即：对线索区的注视时间、注视次数与保持测验、迁移测验有何关系？比较遗憾的是，该问题被大多数研究者所忽视，仅极少部分研究者采用相关等分析方法探讨了注意引导和学习效果的关系(Ozcelik, Arslan-Ari, & Cagiltay, 2010; Ozcelik, Karakus, Kursun, & Cagiltay, 2009)。因此，本文之所以没有探讨眼动与非眼动研究的关系，主要是由于现有关于注意引导和学习效果之间关系的研究并不多，因此可生成二者关系(例如：相关系数 r)的独立效应量数量极为有限，不足以说明上述研究问题。

不过，必须承认的是，眼动技术越来越成为多媒体学习研究者探讨线索效应的重要技术之一(Hyönä 2010; Mayer, 2010; Scheiter & Van Gog, 2009; Van Gog & Scheiter, 2010)，而其中探讨注意引导和学习效果之间的关系也越来越成为研究线索效应乃至多媒体学习领域其他问题的重要趋势(Ozcelik et al., 2010; Ozcelik et al., 2009; Scheiter & Eitel, 2015)。因此，在原文中 4.3 部分，我们对此问题进行了展望(为便于审稿，在文中标为蓝色)，以期引起多媒体学习领域研究者的高度关注。摘录如下：

未来研究可考虑从以下角度深入开展：(1)继续采用眼动技术考察线索在不同条件下的注视加工作用，同时注重探讨眼动注视和学习效果之间的相互关系(Boucheix & Lowe, 2010; Scheiter & Eitel, 2015)，这样既可深入了解线索对学习过程中注意分配模式的影响，也可为后续元分析提供充足的数据来源……

意见 4：建议将文中的表改为三线较为合适。如果觉得放在某一表中的内容太多，可以拆分成不同的表，而不应该破坏心理论文常用的三线表达的规则。

回应：根据建议将表 4 按照 4 类结果变量拆分为 4 个不同的三线表，原文表 5 则改为表 8，同时我们在表下备注及正文文字部分等处作了相应变更。详见 3.4 部分的修改。

意见 5：表 1 中，有些文献出现了两次，如“Tabbers, Martens, & Van Merriënboer, 2000”和“Tabbers, Martens, & Van Merriënboer, 2004”均出现了两次，这样操作的意图是什么？是各自有多个实验还是有其它的考虑？

回应：感谢专家的细心审稿。在表 1 中，之所以有些文献看起来出现了两次，主要是为了在调节效应检验中尽量生成充足的效应量数量而考虑。部分文献尽管只有一个实验(例如：Crooks, Cheon, Inan, Ari, & Flores, 2012; Jamet, Gavota, & Quaireau, 2008; Tabbers, Martens, & Van Merriënboer, 2000, 2004)，但其中的自变量除了线索(有、无)外，往往还有其他被关注的自变量，而这些变量可能是本元分析所关注的潜在调节变量。比如专家所提及的“Tabbers, Martens, & Van Merriënboer, 2000”和“Tabbers, Martens, & Van Merriënboer, 2004”两篇文献均为单个实验，但这些实验都是 2(线索：有、无)×2(呈现通道：视觉单通道、视听双通道)实验设计，“呈现通道”是其自变量之一，同时也是本研究关注的一项调节变量。为了考察在不同呈现通道条件下的线索效应是否存在差异，需在视听单通道和视觉双通道条件下各自生成一定数量的效应量以方便进行调节效应检验的分组对比，故当这些变量为本研究关注的调节变量时，并未被合并成单一的效应量，而是分开编码并生成各自对应的效应量(文中其他文献与此类似)。这一点在正文 2.4 部分和 3.1 部分中已做相应说明(为便于审稿，在文中标为蓝色)。另外，许多元分析研究中也存有类似的现象(参见：Adesope & Nesbit, 2012; Gegenfurtner, Lehtinen, & Säljö, 2011; Rowland, 2014; Wittwer & Renkl, 2010)。因此，表 1 中有些文献出现了两次。

意见 6：标点不统一，如“如：箭头”和“比如：Ozcelik 等人(2009)研究发现……”中的冒号不统一。

回应：感谢审稿专家极为细致的审阅。修改稿正文中采用中文输入法将冒号进行了统一，并对一部分其他标点符号进行了更正。

第三轮 主编终审意见

研究思路清晰，逻辑通顺，写作较为规范，数据分析得当，讨论细致。建议就以下几点进一步修改完善：

意见 1：元分析只是一种分析方法，作为视角是否恰当请进一步思考并斟酌题目。

回应：感谢主编建议。“视角”一词欠妥。经考虑，将题目改为《多媒体学习中线索效应的元分析》。

意见 2：摘要需注意相关表述，例如学习成绩对应的是保持测验成绩和迁移测验成绩，而非只是测验，可考虑修改为学习效果的保持和迁移，或同作者在脚注中提到的，浅层/深度学习效果；或者，将此句话修改为“分别以保持测验、迁移测验的成绩作为学习效果的结果变量，以线索区的注视次数、注视时间作为注意引导的结果变量”。事实上，研究探讨了线索对学习效果和注意引导两方面的作用，其中学习效果主要以保持测验和迁移测验中的成绩为指标，注意引导主要以注视时间和注视次数等眼动数据为指标。但是，在多处的表述存在不一致，建议通读全文，全文保持一致。

回应：感谢主编细致审阅和建议。如审稿意见所言，测验成绩为学习效果的结果变量，而非只是测验。接受主编建议，将摘要中相关表述更正为“分别以保持测验、迁移测验的成绩作为学习效果的结果变量，以线索区的注视时间、注视次数作为注意引导的结果变量”。同时通读全文，对表述不一致之处也进行了更正，强调成绩，而非测验(已在正文中标为绿色字体)。

意见 3：文中指出，线索效应可能受到多方面因素调节，如线索类型、材料特征和学习者特征等。此处，“材料特征”指代容易引起歧义，是指线索材料特征还是学习材料特征，建议修改为“学习材料特征”。相对应的，“呈现通道”、“呈现步调”等均应指明是学习材料的。

回应：为尽量避免可能存在的歧义，已将“材料特征”修改为“学习材料特征”。另外，也在正文中各处适当地指明“材料动态性”、“呈现通道”、“呈现步调”均属于学习材料的特征。

意见 4：在“1.4 本研究假设”部分，提到“研究推测：”，建议修改为“研究假设：”。

回应：根据建议，已将“1.4 本研究假设”部分的“研究推测：”修改为“研究假设：”。

意见 5：“2.1 文献检索”部分，可将“检索关键词的选取主要基于…时间跨度为”挪至本段第一句话之后，再分别指出中英文文献搜索时采用的关键词。

回应：已根据主编的合理建议进行相应更正。具体见 2.1 部分。

意见 6：结果部分，“3.1 文献纳入及编码结果”部分，应出现在方法部分，而非结果部分。

回应：感谢主编建议。为保证文章通顺，我们将文献纳入结果移至 2.2 部分中，将编码结果移至 2.3 部分中。具体见 2.2 部分和 2.3 部分绿色字体文字。此外，移动之后，相应部分的标题或标题号稍有变动。