

# Investigating the Utility of Eye-Tracking Information on Affect and Reasoning for User Modeling

K. Muldner<sup>1</sup>, R. Christopherson<sup>2</sup>, R. Atkinson<sup>2</sup>, and W. Burleson<sup>1</sup>

Arizona State University, School of Computing and Informatics/Arts, Media and Engineering<sup>1</sup>, Psychology of Education<sup>2</sup>, Tempe, Arizona  
{Katarzyna.Muldner,rmchris3,Robert.Atkinson,Winslow.Burleson}@asu.edu

## 调研眼动信息在用户情感和推理建模中的效用

**【摘要】** 我们调研了使用眼动仪为用户的情感和推理提供信息的效用。为了完成调研，我们进行了用户研究，结果显示用户瞳孔对于积极和消极的情感状态反应差别很大。就推理而言，我们的分析发现大的瞳孔尺寸往往伴随着更具建设性的推理事件，这也表明，如果想要消除不同推理事件之间的模糊歧义，可能需要额外的信息。我们的结果显示，对于提高用户模型带宽，参考瞳孔反应是非常有前途的非侵入式的途径。

## 1 介绍

在不干扰用户和自适应系统的交互的情况下，提高模型 *带宽*（即建模时可用的信息数量和质量），是用户建模的关键挑战<sup>1</sup>。可以证明的是，想要获得的信息层次越高，用户模型的建立越复杂，因为这样可能需要重要的人工智能技术和创新的传感器。因此，证明（1）获取必要的用户的状态是可行的（可行性要求）和（2）越来越复杂的模型提高了系统的可用性（可用性要求）越来越重要。

在这篇论文中，我们关注可行性需求，调研了眼动仪提供的瞳孔数据对于研究高层次的情感和推理两种用户状态的用户建模的效用。关于用户的感受和/或者推理的信息具有很高的价值，它使得自适应系统对于用户需求和偏好做出合适的响应。比如，用户经历一段投入的计算响应之后，再使用计算机处理一个令人沮丧的任务的话，所消耗的时间更长（参见 2）；如果计算例子提供了量身定做的提示来提升用户的元认知技能（比如领域无关的推理能力，参见 3），用户学习的产出将会增加。然而，高层次的状态不易观测，很难轻易获取。创新的传感设备是非常有前途的，它可以很自然地将用户与自适应系统交互过程中的生理状态作为副产品捕获。比如，D'Mello 和 Graesser（参见 4）靠机器学习表明，对话和姿态特征可以区分无聊、混乱、流和沮丧等情感状态。Burleson et al.（参见 5）指出，一个整合了压力鼠标、姿势椅子、摄像头和皮肤电导手镯的学习同伴会影响学生的动机和对于同伴的态度。

也有人研究眼动仪的凝视模式信息是如何影响用户模型的，比如决定（1）注意力转移和/或者专注；（参见 6,7）（2）通过自我解释的高层次推理（参见 8），即解释和消化教学资源的过程（参见 9）。另一个分支的眼动追踪研究专注于瞳孔扩张状态。在精心控制的实验条件下，脑力负荷和瞳孔膨胀，情感和瞳孔膨胀之间有清晰的联系，情感反应和脑力劳动都会提高瞳孔尺寸。（参见 10,11,12）然而，这样的评价结果依赖于这样一种不现实的实验协议——和自适应系统之间自然交互的上下文信息大部分都被移除了。举例来说，被试在有限的时间内给充满感情的词分类或者听一段诱发感情的音频。当转移到更真实的应用场景中的时候，对于瞳孔信息的可靠性有各种各样的结果。一些人尝试去找到阅读困难程度和脑力负

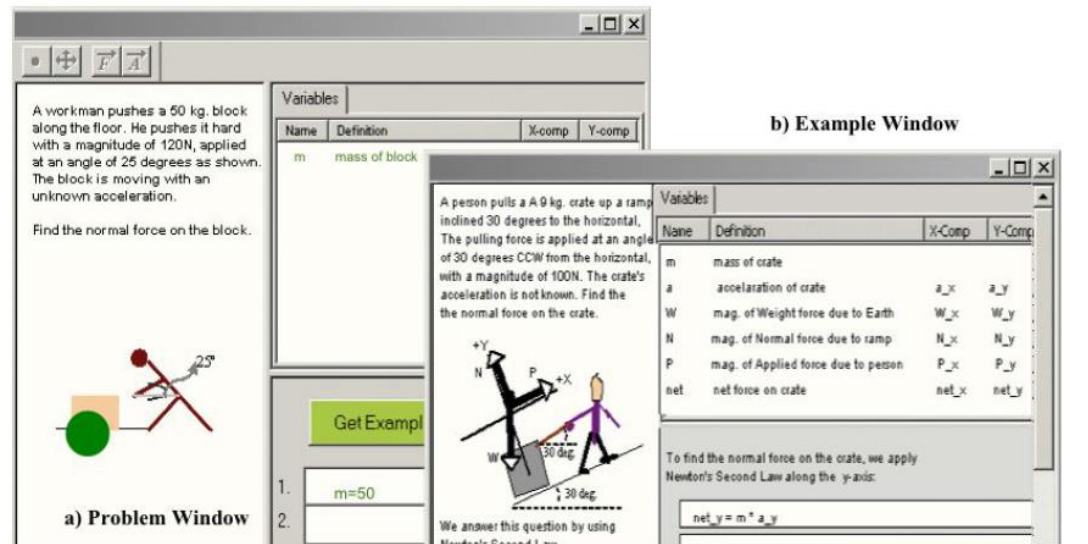
荷之间的关系，但是失败了（参见 14,15），尽管 Igbal et al.<sup>14</sup> 确实发现处理复杂的文件的时候瞳孔尺寸增加。Conati et al.<sup>8</sup> 未能发现瞳孔反应和自我解释（可假定和脑力负荷有关）之间的关系，瞳孔尺寸也是一样。现在比较清晰的是，我们需要投入更多的精力才能解开脑力负荷、情感和瞳孔反应，以及它们对于用户建模效用之间的关系。我们的研究正是这个方向的一小步。

作为我们的实验程序，我们依赖例子类比系统 EA-Coach<sup>16,17</sup>，它是我们开发的自适应学习环境，来支持基于例子的学习的元认知。尽管已经有关于 EA-Coach 的评估表明，一般而言它可以有效促进元认知水平，并且一些学生需要更多的帮助。因此，我们想要扩展这个有用情感和元认知脚手架的系统来帮助所有的学生有效地通过 APS 进行学习。因为这个脚手架将会基于 EA-Coach 用户模型，我们首先调研了提升模型带宽的方法来提供关于学生状态的精确信息。

我们以 EA-Coach 的介绍和它的用户模型开始。然后，我们将会描述我们实施的用户研究来评估情感和推理是否会影响瞳孔反应。我们在展示结果之后，进行了总结，并且对于未来的一些工作给出了建议。

## 2 EA-Coach 系统

EA-Coach 是一个自适应的例子类比学习环境，在分析解决牛顿力学领域的问题的过程中培养学生的元认知能力，也就是说它使用例子来帮助解决问题。有两种元认知能力和 APS 相关，因此它们也是 EA-Coach 系统的目标：



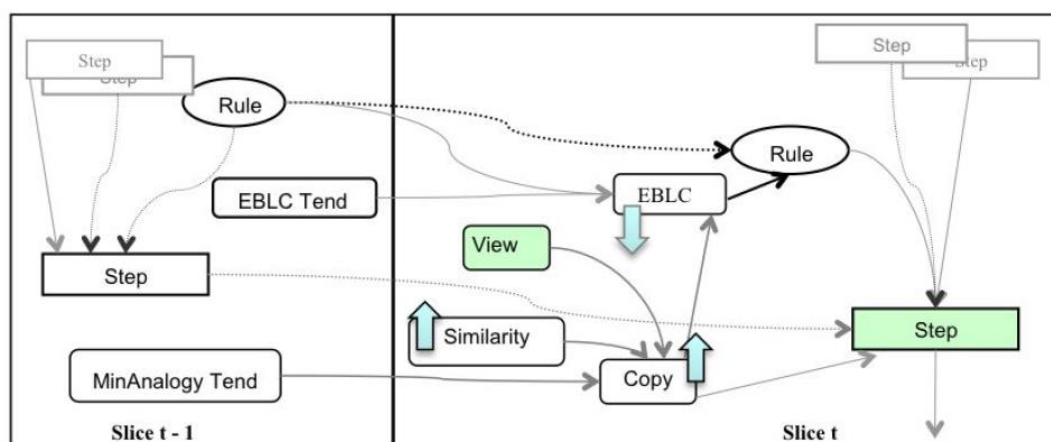
图表 1 EA-Coach 用户界面: (a)问题窗口和(b)例子窗口

- 最小类比：尽可能多地靠自己解决问题，而不是参考例子；（参见 18）
- 基于解释的学习的正确性（EBLC）：一种自我解释形式，使用学习者的现存常识和领域知识来推断新的规则，解释例子答案步骤产生的原因。（参见 19）

EA-Coach 包括学生用来解决题目和参考例子的界面（分别是图 1 中的 a 和 b）。学生在题目窗口画图、输入公式来解决问题。系统不严格限制问题答案的输入，学生可以按任何顺序输入答案甚至直接跳过一些步骤。系统通过将正确和错误的部分标红或者标绿来提供即时的正确性反馈，但是不会提供任何其他的反馈或者提示（比如，和物理有关的反馈）。

解决问题的时候，学生可以向例子求助（通过获取帮助按钮）。作为反馈，EA-Coach 系统从问题库中选出最有可能帮助学生解决问题并且能够让学生学到东西的例子显示在例子窗口里。例子的选取在我们另一篇文章（参见 17）中通过决策论过程完成。在选取的过程中，模型预估（1）学生特性和（2）问题和备选例子之间的相关性对于最小类比和 EBLC 的影响，以及对于后续学习和解决问题的产出的影响。例子展示给学生之后，模型依赖同样的信息源（问题和例子的相似度，学生特性）和学生在界面上所采取的动作来更新对于该学生的评估。这样的评估使得 EA-Coach 可以追踪学生在使用我们的系统时的知识和元认知水平的进化。两种模式使用的是同一个模型结构（预测、评价）。

## 2.1 EA-Coach 的用户模型



图表 2 EA-Coach 用户模型的片段

EA-Coach 的用户模型对应一个动态贝叶斯网络，其中的一部分展示在图 2 中。网络的主干由表示学生正在解决的问题的步骤节点和导致这些步骤的领域规则组成（参见图 2 中的 Step 和 Rule 节点），同时也包括元认知倾向建模的两个节点（参见图 2 中的 EBLCTend 和 MinAnalogyTend）。对于每个被建模的解决问题的步骤，用户模型网络也包括例子对于 APS 过程的影响节点（参见图 2 中的 slice t），如下：（1）相似性节点，捕获目标问题和例子之间的相似度；（2）复制节点，用来捕获学生通过参考例子获得相应解题步骤的概率；（3）EBLC 节点，用来捕获学生通过 EBLC 例子自我解释相关的规则的概率；（4）视线节点，表示学生是否看过相应的例子步骤。当学生在 EA-Coach 界面上产生一个解题步骤的时候，系统解题步骤和例子-视线信息放入模型作为证据，随后系统更新模型对于学生推理过程的理解（复制或者使用 EBLC 的自我解释）。举例来说，在图 2 中，slice t，问题/例子的相似度越高，复制的概率就越大，EBLC 的概率越小，学生也就学习了相应的规则。我们需要注意 EA-Coach 模型带宽比较低，比如，关于学生是否或者是怎么使用 EBLC 来自我解释的唯一的显式的信息是学生是否看过例子窗口的相关的解题步骤和/或者学生的后续解题条目。

评估 EA-Coach 系统我们发现，一般来说 EA-Coach 系统鼓励学生沉浸在最小类比和 EBLC 的元认知行为目标中。然而，评估也显示了有些学生需要比现有系统提供的更多的显式脚手架。因此，我们一直致力于设计这种支持。因为影响和元认知在学习过程中扮演着重要的角色，我们正在探索如何将情感支持加入到现有的 EA-Coach 系统中来，同时丰富现有的元认知支持水平。为了使得新的脚手架对于每位学生是量身定做的，模型如何获取必须的信息面临着挑战，同时，也要保留 EA-Coach 的交互自由的自然属性。

# 3 实验：用户研究

我们研究的目的是探索使用传感器获取的信息对用户关于情感和推理模式的高层次建模的效用。这里，我们将分析放在从眼动仪获取的数据上。实验被试有 15 名大学生，他们不是在上第一年的物理课程就是在高中上过物理课程，但是从来没有接受过高等物理课程。这也是我在另一个研究中[17]，所采用的策略。这背后的合理性在于被试有物理基础，但是又不是物理方面的专家，使得我们获取的数据更稳定。被试一些（1）由于他们的参与获得报酬（5 个被试），还有一些（2）获得了课程的加分（10 名被试）。

每个学生的会话分开进行。在一次会话中，首先给参与者介绍 EA-Coach 的用户界面，然后调整眼动仪，最后使用 EA-Coach 来解答两道牛顿第二定律的问题。对于每个问题，被试者在解决的时候有权参考由 EA-Coach 提供的例题。问题和例子之间的相似度之前已经计算过，所以对于每个问题，被试接收到和现有问题更相似的例子，而不是和其他问题相似的例子。通过提供两种相似度（高和低），我们希望最大化被试表现出范围更广的影响和推理行为的机会。问题和相似度的顺序已经全面平衡过了。被试被告知每道题他们有 60 分钟的时间，但是想停就停。

被试使用 EA-Coach 解题的同时，Tobii T60 眼动仪捕捉他们的凝视信息。这个眼动仪完全整合在一个 17 英寸的显示器中，完全不具有侵扰性，所以从参与者的角度来看，它就像一个规矩的电脑屏幕。校准眼动跟踪，参与者被要求专注于电脑屏幕上一系列的 16 像素点，这一过程花了大约一分钟。我们也使用一组非侵入式传感器捕获其他被试生理数据，但是这些数据还在分析中，并没有在这里呈现（这些传感器包括一个测量皮肤电导传感器的手镯，一个压力鼠标和一个放在受试者椅子上的压力垫，参见 21）。

为了获取被试在实验过程中推理和感受信息，我们要求被试大声地用语言表达他们的想法和感受（参见 22），在[23]中，这种方法被扩展到包括影响。被试的语言数据，和他们的眼动数据，以及在用户界面的交互信息都被 Tobii 系统记录在视频中；EA-Coach 将所有的交互信息记录为普通的文本日志。

## 3.1 数据预处理：编码记录

为了调研用户的情感、推理和生理反应之间的相关性，我们需要研究提供以上两种事件的数据。为了获得数据，我们首先转录视频文件，包括被试的动作，说话方式和事件发生的时间戳。然后我们设计了一种识别推理（比如自我解释）和影响（比如开心）的方案。

方案中的推理部分基于我们之前的一个研究[17]。我们将如果被试表达关于与物理有关的领域特定规律的结论的表达方式编码成自我解释。被试表达了问题和例子之间的关系但是没有迹象显示除了类比其他的推理方式的说话方式被编码为类比。最终，我们加入了一个“其他推理”编码，因为我们想要捕捉用户表达一些推理的时候，虽然由于太浅了不能被分类为自我解释，但是确包含了高于明显的问题/例子的比较的类比。注意，自我解释是高层次的创造性的推理活动，和正向的学习产出相关；通过类比的推理相比之下缺少一些学习；同样的，在我们的定义中，“其他推理”和自我解释相比缺少一些建设性的推理。

情感编码：			
编码	#	描述	样例语句
积极	68	被试表达和开心、兴奋相关的积极情感	“我做对了，这让我很高兴”“噢，这令人

			很兴奋啊”“万岁”“我感觉很好”
消极	69	被试表达和挫败相关的消极情感	“我快疯了”“上帝啊,这太令人愤怒了” “不,不对!”“真讨厌!!”
羞愧	20	被试表达出差愧和懊悔的感情	“我感觉自己像个傻瓜一样”“我失败了...不好意思用了那么长时间”
困惑	29	被试表达出困惑	“我感觉很困惑”“可能它是让我画水平线...我不能理解”
<b>推理编码</b>			
<b>编码</b>	<b>#</b>	<b>描述</b>	<b>样例语句</b>
自我解释	39	被试解释或者澄清了一个物理相关的概念	“因为它在加速,所以我知道合力不等于零”“它变成零是因为没有 x 轴部分”
类比/复制	180	被试比较了问题和例子或者直接复制,但是并没有表现出附加的推理	“a 是我的鼠标加速块”
其他推理	106	被试表达出一些比较浅的推理,既不是自我解释也不是单纯的类比	“在这张图片上,它水平拉伸...然后...90加 40..130?”

编码中的情感部分是全新的,基于数据分析的好几个迭代来增加编码的鲁棒性。刚开始我们计划开发细致的影响分类(比如开心,兴奋和愤怒)。然而,被试有时候会明显地表达一种情感(比如“我很开心”和“我很生气”),有时候也会通过短语比如“不”和“万岁”来表达情感。在后者中,情感的一般导向也就是积极和消极,从语气和用的词可以清楚地获得。(比如,“不”用来表达消极影响),但是它不能精准地表达模糊不清的东西。因此,我们将影响的分类放宽,这样积极的编码就包括被试表达出兴奋、开心或者只是感觉很好。消极编码包括明确地表达出生气或者沮丧,或者使用负面的说话方式,比如“该死”,这个和沮丧相关。

上面描述的编码方案由第一位作者应用来将语句数据分类,将需要的视频文件返回。总体上,一共有 186 个影响编码和 325 个推理编码。

## 3.2 结果

就像上面提到的一样,这里我们专注于从眼动仪中获取的数据,详细就是瞳孔的反应数

据。考虑到被试的瞳孔尺寸基本线有所差异，我们使用 Z-scores 来正则化瞳孔尺寸（也就是， $\text{normalized pupil value} = (\text{original pupil value} - \text{mean pupil size}) / \text{standard deviation}$  参见[8]）。然后将每个编码的语言和正则之后的眼动追踪数据，以及 EA-Coach 的日志信息通过标准化时间戳将三种信息结合起来（转码文件，EA-Coach 日志，眼动追踪日志）。

为了分析这些数据，我们刚开始打算依赖重复测量方差分析和双样本 t 检验作为近似，也就是，依赖于问题中的自由变量的层级（method A, within-subjects 分析）。另外一个可选技术方案使用 one-way ANOVA（method B, between-subjects 分析）。每个方法都有自己的限制。A 方法可以容忍数据稀疏，因为不是所有的被试者都会表现出所有的影响和/或者推理。这样的话，样本的就会变小，从而降低了可信度，提高了 type 2 错误的概率（也就是，当用户实际上表现出某种情感的时候却不能找到）。方案 B 可选，就像我们在[8]中所做的一样。然而，数据集显示和特定编码相关的点不是相互独立的，这就增加了 type 1 错误发生的概率（也就是，当情感并不存在的时候，却找到了）如果使用方法 B 的话。考虑到上述因素，我们决定两种分析都实施，组合这两种方法的发现结果。下面我们将会从和影响有关的发现开始展示结果。

### 3.2.1 关于情感的结果

为了调研瞳孔反应和影响之间的关系，我们计算了一段时间内被试表达了我们认为的有感情的反应的平均瞳孔尺寸。我们使用 5 秒时间间隔，从言语开始的点开始计算（这个临界值和相关工作中使用的是相似的，比如[24]）。

我们从 within-subjects 分析开始。和我们预期的一样，我们发现每个被试并没有表达所有类型的情感反应，这使得数据有所缺失。当我们将困惑或者羞愧两种感情加入我们分析中的情感编码的时候，我们只剩下 6 个表达了所有四种类型的情感的被试。因此，我们决定只在积极和消极情感上面实施我们的分析，因为只有这样才能获得多于 6 个数据点。这个分析包括 10 名学生，我们分别计算了每个学生的在积极情感状态和消极情感状态的平均瞳孔尺寸。双样本 t 检验发现影响对于瞳孔尺寸有显著影响效果（ $t(9)=2.294$ ,  $p=0.047$ ）：平均来看，和被试表达积极情感的时候，瞳孔尺寸在被试在表达负面情绪的时候较小。

上面提到 EA-Coach 系统在用户界面将正确的部分标绿色，错误的部分标红色来提供即时反馈。被试的很多影响反应都是他们在 EA-Coach 系统中产生的块相关，具体来说就是，和被系统标记为正确或者不正确相关。因此，我们想要调研条目的正确性（或者缺少）影响了被试的情感。我们比较了正确和错误的条目之后的 5 秒钟。我们并没有发现正确性对于瞳孔尺寸有显著的影响（ $t(14)=0.508$ ,  $p=0.620$ ）。就 between-subjects 分析而言，ANOVA 分析显示了影响对于瞳孔反应的显著影响（ $F(3,182)=4.057$ ,  $p=0.008$ ）。然后，我们使用 Bonferroni 方法进行事后两两比较，以确定情感反应相互之间有显著差异。唯一的表现出显著的差异一对是积极-消极影响（ $p=0.006$ ），消极影响的瞳孔尺寸比积极情感的小（-0.0913 和 0.3214）。这也帮我们确认了 within-subjects 的分析结果。

### 3.2.2 关于推理的结果

为了调查研究过程中瞳孔反应和被试推理之间的关系，我们计算了被试表现出我们在转码文件中确定的以下三种类型的推理（自我解释，类比，其他推理）的时间段内的平均瞳孔尺寸。在分析中，我们使用了 15 秒的时间跨度，从言语开始的点开始计算（这种临界值被发现可以消除自我解释和缺少投入[8]）。

我们从 **within-subjects** 开始。和情感分析的数据一样，每个被试并没有表达出所有类型的推理。有九个被试表现了所有类型的推理；对于每个学生，我们计算每种类型推理的平均瞳孔尺寸（自我解释，类比和其他推理活动）。由于推理有三个层次，我们实施了重复测量方差分析。结果显示推理对于瞳孔反应有显著的主效应 ( $F(2,8)=3.63, p=0.047$ )。考虑到时候比较推荐时在 **within-subject** 中使用，我们采用了[25]中所提出的方法，实施了成对比较来确认三种类型的推理之间的差异。我们发现瞳孔尺寸在自我解释的时候明显比在其他推理的时候大 (0.4071 和 -0.0661;  $t(9)=-2.382, p=0.04$ )。我们也发现对于自我解释：来说，瞳孔尺寸也比类比的大，但是这个结果不显著 (0.4074 和 -0.0210;  $t(9)=1.744, p=0.115$ )。其他推理和类比之间的差别不显著 (-0.0661 和 -0.0210;  $t(9)=0.395, p=0.702$ )。

和情感分析例子一样，我们想要调研我们的结果是不是由被试的解题条目有关，特别是这些条目之间的（或者缺少）相关性。在这个分析中，我们考察在正确的条目之前和之后的 15 秒，使用双样本 **t** 检验来研究两个变量之间的差异。我们并没有发现任意一个窗口对于正确性的显著影响（之前和之后）。

就 **between-subject** 的分析而言，ANOVA 分析发现推理对于瞳孔反应有显著的主效应 ( $F(2,322)=6.454, p=0.002$ )。然后我们使用 Bonferroni 事后成对比较来确定哪种类型的推理反应和其他的有显著的不同。结果表明平均而言，（1）瞳孔尺寸对于自我解释显著地比其他推理大 (0.2311 和 -0.0876;  $p=0.008$ )，（2）瞳孔尺寸对于类比显著地比其他推理大 (0.1195 和 -0.0876;  $p=0.006$ )。自我解释和类比之间的瞳孔尺寸没有显著的差异。

## 4 讨论和未来的工作

我们的结果显示瞳孔尺寸是提升用户模型的带宽的有前途的非侵入性的途径。就影响而言，**within** 和 **between** 被试的分析确定了和消极感情相比当被试表现积极感情的时候瞳孔尺寸明显更大。和小心控制的实验想对照，我们的被试并没有被诱导表达情感，但却作为和 EA-Coach 交互的时候的自然副产品被表达出来了。他们的情感反应影响瞳孔尺寸，这些信息可以被用户模型考虑进去，从而允许自适应式应用为用户的需求量身定做交互方式。考虑到心理学研究表明对于情感反应瞳孔尺寸会变大，我们的结果表明被试在我们的实验中经历的积极的情感比如兴奋比消极的情感比如沮丧更强烈。我们实验的上下文，也就是一个教学的环境，可能会影响个别的情感反应，所以更多的调研需要进行，来研究非教育环境如何影响瞳孔尺寸。另一个需要进一步研究的领域是如何测量情感。我们发现 **talk-aloud** 方法并不适合情感的精细区分。一般而言，如何测量情感是需要更多研究的关键挑战，但是到目前为止关于这个问题都没有全面的理解。

我们的研究发现了支持这样的事实：被试推理如何影响瞳孔尺寸。就像我们之前指出的一样，在精心控制的实验中，大的瞳孔反应和脑力负荷直接相关。我们比较了三种类型的推理：（1）自我解释，一种高度创造性的脑力活动，和（2）类比，包括问题/例子常量之间的和/或者从例子中复制的没有创造性的活动，和（3）其他推理。由于自我解释比其他两种类型的推理更具有建设性，它应该产出更大的瞳孔尺寸（就像[8]中所推荐的）。我们所实施的两种分析都真正地表明自我解释比其他推理的情况下的瞳孔尺寸更大。然而，我们并没有找出自我解释和类比之间的瞳孔尺寸的显著的不同。实际上，我们的 **between-subjects** 分析表明类比比其他推理产生的瞳孔尺寸更大，尽管这个结果没有被 **within-subjects** 分析确认，我们也不希望这样的结果。所有的分析都没有发现类比和自我解释之间的差别的一个原因可能是，类比的过程中可能真的需要脑力劳动，尽管它实际上是一种浅推理方式。我们在话语模型中看到被试整理问题/例子的实例（被试努力使用例子中的常量 **p** 来近似替换她的问题）。



这些困难可能增加了脑力劳动，因此会增加瞳孔尺寸。我们的结果显示，模型可能需要更多的信息来消除自我解释和类比性推理之间的模糊不清的部分。一种解决这个问题的一种方法可能是分析交互环境中的注意力模式：因为类比性分析需要问题/例子常量之间的比较，但是自我解释不需要，包括凝视信息可以消除自我解释和类比之间的模糊不清。

作为我们的下一步动作，我们计划实施更多的分析来进一步研究积极和消极情感的不同，并且确定导致这种差异的轻微的因素。另外一个相关的研究和探索情影响和认知之间的关系有关。有证据表明，被试在积极的情感状态下处理信息更好，所以分析这种情况是否或者怎么在我们的研究中发生一定很有趣。我们也计划分析眼动仪提供的数据的其他方面（比如注释和扫视）来探索它们如何影响用户模型。我们计划依赖我们在这个实验中和后续的分析中的发现来扩展 EA-Coach 的用户模型，考虑眼动数据，基于修订的用户模型来设计影响和附加的元认知支持。我们会继续评估这种支持如何影响辅导的教学效果和可用性。

## 5 参考文献

1. VanLehn, K.: Student modeling. *Foundations of Intelligent Tutoring Systems*, 55 – 78 (1988)
2. Klein, J., Moon, Y.: This computer responds to user frustration: Theory, design, results, and implications. *Interacting with Computers* 14, 119 – 140 (2000)
3. Aleven, V., Koedinger, R.: An effective metacognitive strategy: learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science* 26(2), 147 – 179 (2002)
4. D' Mello, S.K., Picard, R.W., Graesser, A.C.: Towards an affect-sensitive autotutor. *IEEE Intelligent Systems* 22(4), 53 – 61 (2007)
5. Bursleson, W.: *Affective Learning Companions: Strategies for Empathetic Agents with Real-Time Multimodal Affective Sensing to Foster Meta-Cognitive Approaches to Learning, Motivation, and Perseverance*. Ph.D thesis, MIT (2006)
6. Gluck, K., Anderson, J.: Cognitive architectures play in intelligent tutoring systems? In: *Cognition and Instruction: Twenty-Five Years of Progress*, pp. 227 – 262 (2001)
7. Qu, L., Johnson, L.: Detecting the learner's motivational states in an interactive learning environment. In: *12th International Conference on Artificial Intelligence in Education*, pp. 547 – 554 (2005)
8. Conati, C., Merten, C.: Eye-tracking for user modeling in exploratory learning environments: an empirical evaluation. *Knowledge Based Systems* 20(6), 557 – 574 (2007)
9. Chi, M., Bassok, M., Lewis, M., Reimann, P., Glaser, R.: Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science* 13, 145 – 182 (1989)
10. Marshall, S.P.: Identifying cognitive state from eye metrics. *Aviation, Space, and Environmental Medicine* 78, 165 – 175 (2007)
11. Van Gerven, P.W.M., Paas, F., Van Merrinboer, J.J.G., Schmidt, H.G.: Memory load and the cognitive pupillary response in aging. *Psychophysiology* 41(2), 167 – 174 (2004)
12. Vo, M.L.H., Jacobs, A.M., Kuchinke, L., Hofmann, M., Conrad, M., Schacht, A., Hutzler, F.: The coupling of emotion and cognition in the eye: Introducing the pupil old/new effect. *Psychophysiology* 45(1), 130 – 140 (2008)
13. Partala, T., Surakka, V.: Pupil size variation as an indication of affective processing. *Int. Journal of Human-Computer Studies* 59(1-2), 185 – 198 (2003)



14. Iqbal, S., Zheng, X., Bailey, B.P.: Task-evoked pupillary response to mental workload in human-computer interaction. In: CHI 2004 extended abstracts on Human factors in computing systems, pp. 1477 – 1480 (2004)
15. Schultheis, H., Jameson, A.: Load in adaptive hypermedia systems: Physiological and behavioral methods. In: Adaptive hypermedia. Interacting with Computers, pp. 225 – 234 (2004)
16. Conati, C., Muldner, K., Carenini, G.: From example studying to problem solving via tailored computer-based meta-cognitive scaffolding: Hypotheses and design. *Technology, Instruction, Cognition and Learning (TICL)* 4(2), 139 – 190 (2006)
17. Muldner, K., Conati, C.: Evaluating a decision-theoretic approach to tailored example selection. In: IJCAI 2007, 20th International Joint Conference in Artificial Intelligence, pp. 483 – 488 (2007)
18. VanLehn, K.: Analogy events: How examples are used during problem solving. *Cognitive Science* 22(3), 347 – 388 (1998)
19. VanLehn, K.: Rule-learning events in the acquisition of a complex skill: An evaluation of cascade. *The Journal of the Learning Sciences* 1(8), 71 – 125 (1999)
20. Muldner, K.: Tailored Support for Analogical Problem Solving. Ph.D thesis, University of British Columbia (2007)
21. Dragon, T., Arroyo, I., Woolf, B.P., Burleson, W., el Kaliouby, R., Eydgahi, H.: Viewing Student Affect and Learning through Classroom Observation and Physical Sensors. In: Woolf, B.P., Aïmeur, E., Nkambou, R., Lajoie, S. (eds.) ITS 2008. LNCS, vol. 5091, pp. 29 – 39. Springer, Heidelberg (2008)
22. Ericsson, K., Simmon, H.: Verbal reports as data. *Psychological Review* 87(3), 215 – 250 (1980)
23. Craig, S., D’Mello, S., Witherspoon, A., Graesser, A.: Emote aloud during learning with autotutor: Applying the facial action coding system to cognitive-affective states during learning. *Cognition and Emotion* 22(5), 777 – 788 (2008)
24. Van Gerven, P., Paas, F., Van Merriënboer, J., Schmidt, H.: Memory load and the cognitive pupillary response in aging. *Psychophysiology* 41(2), 167 – 174 (2001)
25. Cardinal, R., Aitken, M.: ANOVA for the Behavioural Sciences Researcher. Routledge, London (2006)
26. Mauss, I., Robinson, M.: Measures of emotion: A review. *Cognition & Emotion* 23(2), 209 – 237 (in press)
27. Levens, S., Phelps, E.: Emotion processing effects on interference resolution in working memory. *Emotion* 8(2), 267 – 280 (2008)