

代数应用题项目生成的结构分析方法*

杨向东

(华东师范大学课程与教学研究所, 上海, 200062)

摘要 自动化项目生成是近年来兴起的测量领域, 是一种以项目认知加工理论为基础的原则性项目设计(principled item design)模式。其中, 如何在项目认知模型基础上, 通过任务结构分析的方式系统全面的鉴别和提取任务特征是一个关键环节。基于已有文献中代数应用题的命题分析法、网络语言分析法、关系-函数分析法、任务分析地图等四种结构分析方法, 研究探索了能够服务于自动化项目生成的代数应用题任务结构分析方法。该分析表明, 前三种方法分别对应于个体解题过程需要形成的三种中介表征, 即问题陈述背后的命题表征、事件时空关系的情境模型、以及变量间数量关系的问题模型, 第四种方法从过程角度分析了问题解决的认知需求。然而, 要实现项目生成的特征提取需求, 尚需对现有四种方法所揭示问题特征的心理现实性、特征提取的系统性和完备性、任务领域的适用范围、以及不同方法的整合等问题开展进一步研究。

关键词 自动化项目生成; 代数应用题; 项目认知模型; 任务结构分析方法

分类号 B841

1 引言

近几十年来, 心理与教育测量学正在经历一场研究范式的重要转变(Whitely, 1983; Embretson, 2004; Mislevy, 1993)。这一转变的核心是对理论驱动测量模式的倡导和研究(Lohman & Ippel, 1993)。相应的, 旨在整合测量建构的认知理论与心理计量学技术的研究在过去几十年中大量涌现(Bejar & Yocom, 1991; Bouwmeester & Sijtsma, 2004; Embretson, 1994, 1998; Kyllonen & Christal, 1990; Irvine & Kyllonen, 2002; Wilson, 2013)。理论驱动的测量模式强调通过理论分析和实证研究, 建立测验任务反应机制与测验建构内在表征之间的对应关系(Embretson & Gorin, 2001), 确保对个体进行推断和解释的效度, 以及后继决策或干预措施的成效。近年来兴起的测量领域, 如自动化项目生成(Embretson & Yang, 2007)、认知诊断测验(Leighton & Gierl, 2007; Rupp, Templin, & Henson, 2010)等, 都是建立在这一思想基础之上的。

不过, 并非所有的自动化项目生成都同等程度的依赖于理论驱动的测量学模式。不同理论流派和技术路线可以从算法化生成程度(level of algorithmic generation)和建构操纵深度(level of construct manipulation)两个方面加以分析。早期Hively等人(Hively, Patterson, & Page, 1968)有关算术题的项目格式法(item form), 和Bormuth(1970)有关文本理解的语句转换法(sentence transformation)具有很高的算法化程度, 但是缺乏对项目反应机制的研究, 无法实现对测量建构的系统操纵。后续许多项目生成研究, 如Enright等人(Enright, Morley, & Sheehan, 2002)对数学项目生成的研究, 在已有认知理论基础上提出影响项目解决的若干项目特征, 但这些特征只能指导后继项目生成, 无法实现算法化。而近年来逐渐流行的两种项目生成模式, 即项目模型法(item model approach)(Bejar, 2002; Bejar & Yocom, 1991)和认知项目设计系统法(cognitive item design system approach)(Embretson, 1998; Embretson & Yang, 2007)则较好的实现了建构操纵深度和算法化项目生成的结合。和项目模型法相比, 认知项目设计系统法对测量建构的操纵更为系统和深入。

收稿日期: 2013-05-29

* 国家自然科学基金(31171000)和上海市教育科学2011年度重点项目(A1117)资助。

通讯作者: 杨向东, E-mail: xdyang50@hotmail.com

项目生成的认知设计系统法要求建立任务特征与项目解决过程中关键认知变量的对应关系,并标定它们对项目测量学特征的影响力度。这样,项目设计便可以通过系统变换任务特征,实现对项目解决所需认知成分及其负荷的操纵,从而生成具有指定测量学特征的项目(Embretson & Yang, 2007)。其中,如何鉴别和提取项目的任务特征是一个关键环节。自动化项目生成是一种基于原则的项目设计(principled item design)模式。任务特征的鉴别和提取也需要建立在系统性和原则性的任务结构分析方法上,尽可能减少研究者的主观判断。此外,为了实现对测量建构的系统操纵,结构分析方法还要与项目解决的认知模型相契合,以保证所提取的任务特征与项目解决过程中的关键认知变量能够建立关联,为后继认知研究和项目生成奠定基础。所以,指向项目生成的任务结构分析不同于一般性的分析方法。它不仅要有助于任务特征的识别以及相关变量的提取,还要能够促进对个体问题解决认知过程的理解(Hall, Kibler, Wenger, & Truxaw, 1989)。

因此,针对具体任务类型开展项目生成研究时,需要找到或形成一种(或几种)能够满足上述要求的问题结构分析方法。代数应用题作为具体领域的复杂认知任务,一直是认知研究的重要领域(Hinsley, Hayes, & Simon, 1977; Koedinger & Nathan, 2004; Mayer, Larkin, & Kadane, 1984)。近年来,随着自动化项目生成研究的深入,对代数应用题的项目生成研究也正在逐步展开(Arendasy & Sommer, 2007; Daniel & Embretson 2010; Deane & Sheehan, 2003)。本研究试图在认知项目设计系统法指导下,分析已有研究,探寻能够用于自动化项目生成的代数应用题结构分析方法,对后继项目生成研究奠定基础。

2 代数应用题的问题解决模型

在英文文献中,代数应用题又称为代数故事题(algebra story problem)(Mayer, 1981)。该类题目通常以具有人物、事件和情节的故事形式呈现。案例1给出了这类问题的一个样例。

案例1:小丁丁步行去少年宫。他平均每分钟走75米。小丁丁走了8分钟后,爸爸骑车追赶,他每分钟比小丁丁多走120米。爸爸几分钟后在途中追上小丁丁?

在本研究中,代数应用题被视为是以文本形式呈现的、嵌套在特定现实情境中的各种变量之间的某种量化结构(杨向东, 2013)。

在综合已有研究的基础上,Sebrecht等人(Sebrecht, Enright, Bennett, & Martin, 1996)提出了一个代数应用题问题解决的认知加工模型(见图1)。按照这一模型,代数应用题的解决过程由转译(translation)、整合(integration)、计划(planning)、监控(monitors)和执行(execution)五种认知活动构成。当面对问题时,个体首先根据问题陈述进行各种转译,比如借助于语言知识来理解不同陈述的语义(表层转译);利用日常经验建立问题情境中不同人物、对象或事件的时空关系(情境转译);从问题陈述中析取变量、确定变量取值、以及形成局部的变量间数量关系(量化转译)等等。转译后形成的各种问题元素继而整合成一个对问题结构的整体表征。它既需要整合转译阶段的语义内容、情境结构和数量关系,也需要整合当前内容与长时记忆中被激活的问题图式、情境模型和数学公式等。合理的问题表征和产生问题解决方案(比如形成解题算式)有着直接关系,但这种关系并不是确定性的。研究发现,即使有证据表明个体形成了合理的问题表征,但仍有可能不知如何解决(Koedinger & Nathan, 2004)。此时,个体需要根据已有表征,确定一种或多种问题解决策略,比如倒推(unwind)、猜测检验(guess-and-test)、先分部后整合(decompose-and-combine)等(Koedinger & Nathan, 2004; Sebrecht et al., 1996)。对于复杂问题,个体不仅要制定策略实施的计划,还要决定监控策略,确保内在问题表征和外在问题表征(画图、列算式或方程等)是一致的。在此基础上,个体根据既定问题表征和解题计划,形成问题解决的外在表征形式(如画出相应示意图或列出解题方程式),进行运算或代数操作以求出答案,并按照既定监控策略来确认这一过程是否与原有表征相一致。对于有既定问题图式或非常简单的问题,上述认知活动的外在表现是非常简化的。通常只会观察到个体阅读问题,继而列出算式并解答。而对于陌生的复杂问题,个体有时会通过外在表征(比如在阅读问题陈述的过程中逐步形成示意图,或者先列出分步算式,再整合成综合解题方案等)的形式来降低认知负荷,以确保问题解决过程的顺利进行。

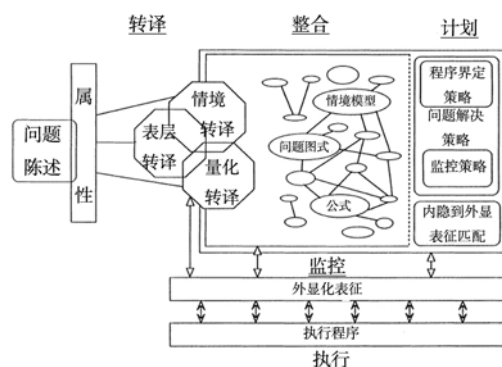


图1 代数应用题问题解决的认知加工模型

(资料来源: Sebrect et al., 1996)

上述模型表明, 代数应用题解决过程是一个复杂的认知活动, 会涉及到语言理解、心理表征、常识性知识、数学知识或技能等不同方面。在理解问题阶段, 个体需要形成三种不同的问题表征(Nathan, Kintsch, & Young, 1992), 即命题表征(propositional representation)、情境模型(situation model)和问题模型(problem model)。在解题阶段, 个体根据问题表征中的变量关系进行各种认知操作, 不断获取所需中介结果, 直至最终答案的获得。在这一过程中, 问题解决的认知需求主要受两方面影响: (1)每个解题环节需要同时考虑的元素数量及其整合复杂性, 和(2)获取最终答案所需的中介结果数量及其复杂性(Stillman, 1996)。因此, 在选择或构建代数应用题结构分析方法时, 需要考虑该方法能否揭示个体在问题解决过程中所形成的不同表征的特征, 以及影响问题解决认知需求的因素。

3 代数应用题的结构分析方法

已有研究文献中存在多种不同的数学应用题结构分析方法。其中较为成熟、运用较为广泛的包括命题分析法(propositional analysis)、网络语言分析法(network language analysis)、关系-函数分析法(relational-functional analysis)、任务分析地图法(task analysis map)等。这些方法关注的角度以及涉及到的问题表征水平各不相同。

3.1 命题分析法

命题分析最初被用来分析文本的语义内容(semantic content)。每一个命题表示一个观念单位(idea unit), 包括一个关系概念(relational concept)

以及一个或多个论元(argument) (Kintsch & van Dijk, 1978)。论元可以是一个概念, 也可以是一个嵌入式命题(embedded proposition)。论元在命题中既可以是施动者(agent), 也可以是受动者(object)或目标(goal)。例如, 句子“约翰和玛丽都喜欢数学老师”可以分解成如下三个命题¹。

P1 (和: 约翰, 玛丽)

P2 (喜欢: P1, 老师)

P3 (修饰: 老师, 数学)

其中, 命题P1作为一个嵌入式命题, 成为命题P2的一个论元, 在该命题中承担着施动者的语义角色。命题可以分为三种类型, 即谓语命题(predicate proposition)、修饰命题(modifier proposition)和连接命题(connective proposition) (Turner & Greene, 1977)。P1、P2和P3分别属于连接命题、谓语命题和修饰命题。

以命题为基本单位, 文本的语义结构可以表征为一系列彼此相连的命题, 即所谓的文本基(text base)。文本基包括文本中所有明确表达的命题和为了连贯而从文本中推断的命题。不同命题之间通过论元重叠形成参照连贯性(referential coherence)。通过这种连贯性, 文本基的微观结构可以表示成一个相互连接的命题网络图(参见Kintsch & van Dijk, 1978)。一般而言, 处于命题网络图上的命题往往是形成文本连贯性的核心命题, 成为宏观命题(macroposition), 形成文本的“主旨(gist)”。

数学应用题语言陈述的理解过程被认为符合文本理解的一般原则(Kintsch & Greeno, 1985)。因此, 命题分析方法或许可以作为数学问题语义结构的一种分析方法。例如, 案例1可以被表征为如下的一系列命题。

P1 (去: 小丁丁, 少年宫)

P2 (修饰: P1, 步行)

P3 (速率: P1, 75 米/分钟)

P4 (比...晚: P6, P1)

P5 (持续时间: P4, 8 分钟)

¹不同的命题表示方式仍然存在很多细节上的不同(见Bovair & Kieras, 1985; Kintsch & van Dijk, 1978; Turner & Greene, 1977)。本文的命题表示方式在 Bovair & Kieras (1985)的基础上做了改动。

P6(追赶:爸爸,小丁丁)

P7(修饰:P6,骑车)

P8(比...快:爸爸,小丁丁)

P9(速率:P8,120米/分钟)

P10(追上:爸爸,小丁丁)

P11(持续时间:P10,?分钟)

这11个命题构成了案例1的文本基。

通过不同命题之间论元的参照关系,可以构建案例1的命题结构图(见图2)。案例1中,“爸爸追上小丁丁”是该问题的主题,因而P10成为处于最上位的宏观命题。与该命题直接相关联的命题有三个(P11、P1和P6)。其中P11是个修饰命题(P10作为一个论元嵌套在该命题中),描述的是P10的时间特征(“几分钟后P10可以发生”)。P1和P6两个命题与P10各共享一个共同的论元(P1与P10共享“小丁丁”;P6与P10共享“爸爸”),分别描述了两个施动者各自的行动或状态(“小丁丁去少年宫”和“爸爸追赶”)。围绕P1和P6的是两个施动者各自的行动方式——“步行”(P2)或“骑车”(P7)、行动速度——“75米/分钟”(P3)或“每分钟快多少米”(P8),以及时间关系——“小丁丁出发几分钟后爸爸出发”(P4)。P5与P9作为两个修饰命题,分别限制了P4的时间特征(“8分钟”)和P8的数量特征(“每分钟快120米”)。此外,命题P8和P3之间用虚线相连,表明虽然在文本中没有明确提及,两个命题之间具有潜在的参照连贯性。

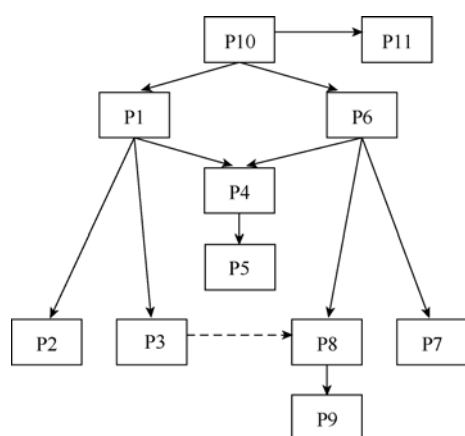


图2 案例1中不同命题之间的结构关系

3.2 网络语言分析法

网络语言分析法最早由Quintero(1981)提出,

Shalin & Bee(1985)加以完善。该方法认为,代数应用题的数学结构是由各种数量元素(quantitative element)及其关系构成的一个网络。Shalin & Bee(1985)区分了四种数量元素:广延元素(extensive element)、集中元素(intensive element)、差别元素(difference element)和因子元素(factor element)。其中广延元素是指事物的基本属性(primary property),如“长度、时间、体积”等,而集中元素指事物的二阶属性(secondary property),是两个基本属性之间的某种映射关系,比如“速度是单位时间内经过的距离”等。差别元素是指两个元素数量之间的加减比较²,如“爸爸每分钟比丁丁多走120米”。因子元素是指两个元素数量之间的乘除比较,如“长是宽的2倍”。四种元素都包括三个基本成分,即变量名称、变量取值和数量单位。例如,“丁丁每分钟走75米”表述了一个集中元素,其变量名称、变量取值和数量单位分别为“丁丁的速度”、“75”和“米/分钟”。在网络语言分析法中,每个数量元素用一个文本框来表示。不同类型的数量元素用文本框上方的不同形状来加以区分³。图3给出了不同类型数量元素的表征方式。

不同元素之间的关系构成了代数应用题中的数量关系。三个元素的关系组合构成一个三阶组合(triad)。图4给出了一个三阶组合的示意图。通常,总体量(whole quantity)放在三阶组合的上方,组成该总体量的两个部分(part)放在下方,以确保了从符号表征向数学代数式(算式)转换的一致性。例如,从图4中,可以推导出“虫的数量+鸟比虫多的个数=鸟的数量”,或者代数式“ $3+x=5$ ”,其中 x 代表鸟比虫多的个数。在实际运用中,总体量可以根据具体情况加以确定,如时间、速度和距离中的距离,三个数值中的最大值等。

² Hall et al.(1989)将差别元素和因子元素限定为两个广延元素之间的比较。这是不全面的。例如“爸爸每分钟比丁丁多走120米”实际上是两个集中元素(“爸爸速度”和“丁丁速度”)的比较。类似的情况也适用于因子元素。

³ 网络语言分析法在表示方式上也存在很多细节上的区别(Hall et al., 1989; Nesher & HersHKovitz, 1994; Sebrecht et al., 1996; Shalin & Bee, 1985)。本文综合了各种不同表示方式,力求全面反映该方法。

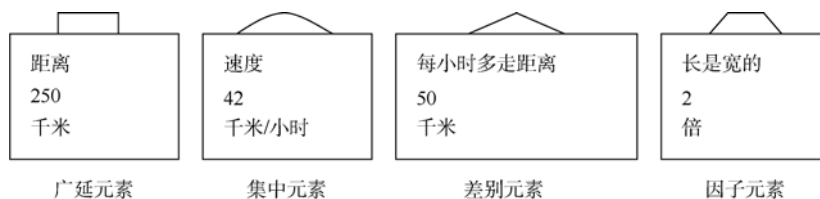


图 3 不同类型数量元素的表征方式

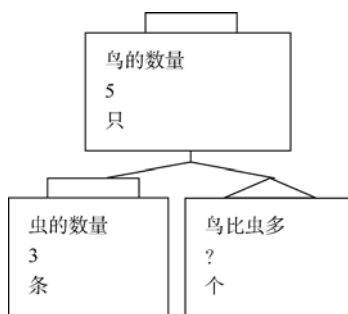


图 4 三阶组合的表征方式

复杂应用题是由不同三阶组合构成的。不同三阶组合之间存在 3 种结合方式：层级或嵌套型(hierarchical/nested)、共享整体型(a shared whole)和共享部分型(a shared part) (图 5 给出了三种组合方式的框架性示意图)。

上述基本元素提供了分析代数应用题变量结构的一种方法。利用该方法，案例 1 可以表征为图 6，其中粗线文本框表示所求未知量。该图表明，案例 1 共有 8 个数量元素，包括 5 个广延元素、2 个集中元素和 1 个差别元素。其中，“丁丁速度”、

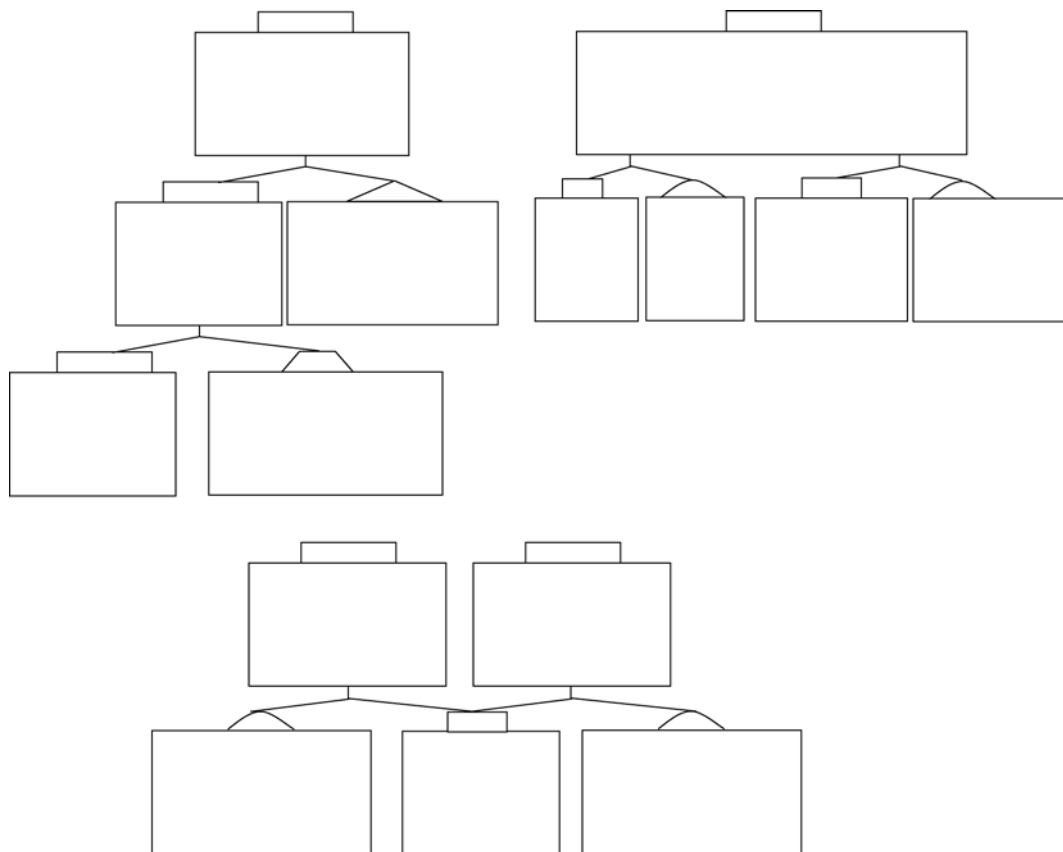


图 5 三阶组合的结合方式(左上：嵌套型；右上：共享整体型；中下：共享部分型)

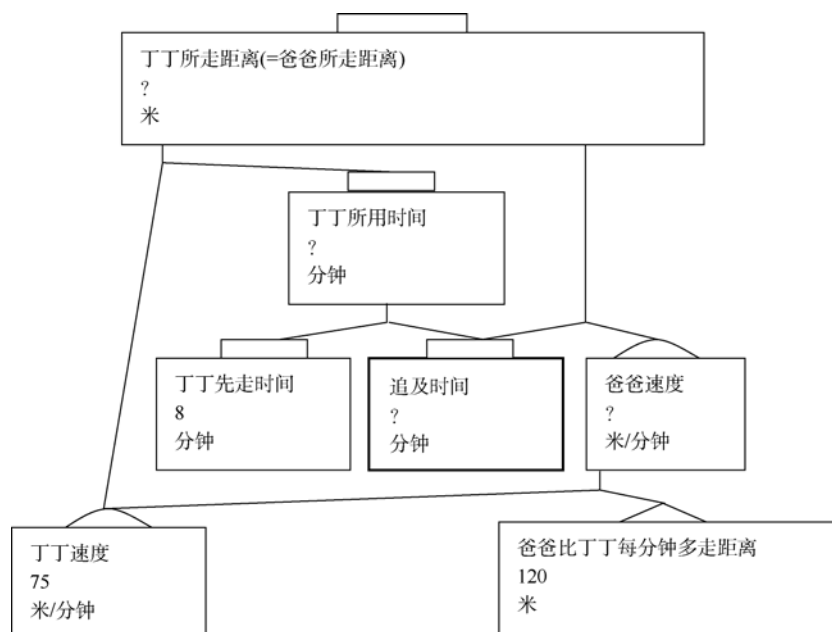


图6 案例1中不同数量元素的结构图

“丁丁先走时间”和“爸爸比丁丁每分钟多走距离”给出了明确取值,其它元素需要根据元素关系加以推导。从问题结构来看,“爸爸所走距离=丁丁所走距离”是联系问题中所有数量元素的核心。围绕这一核心关系,“丁丁速度”和“丁丁所用时间”、“爸爸速度”和“追及时间”分别与“丁丁所走距离”和“爸爸所走距离”联系起来。这种联系需要个体具备“速度 $\times$ 时间=距离”这一数学知识。此外,图中表明,爸爸所用“追及时间”同时也是“丁丁所用时间”的分量,“丁丁所用时间”是“丁丁先走时间”和“追及时间”的总量。“爸爸速度”需要从“丁丁速度”和“爸爸比丁丁每分钟多走距离”中推出。

网络语言分析法还提供了表征问题解决不同策略的可能性。图7给出了案例1在另一种解决策略下的结构示意图。尽管“爸爸所走距离=丁丁所走距离”仍是联系问题中所有数量元素的核心,图7求取“丁丁所走距离”的路径和图6不同。图7中,“丁丁所走距离”被分解为两个部分,即“丁丁先走距离”和“丁丁后走距离”。两个距离共享一个共同元素“丁丁速度”,但是不同的时间。改变策略导致图7用“丁丁先走距离”和“丁丁后走距离”两个变量取代了图6中的“丁丁所用时间”,因而增加了问题结构中的变量数(图7现在有9个数量元素)。新近研究表明,网络语言分析法适用范围

较广,除了概率或统计类问题之外,几乎所有类型的代数应用题都可以用该方法加以分析(卫芸, 2011)。

3.3 关系-函数分析法

关系-函数分析法由Hall等人(Hall et al., 1989)首先提出, Yerushalmy & Gilead (1999)进一步发展完善。该方法主要用于分析代数应用题的情境结构(situational structure)。在关系-函数分析法中,问题情境被视为是特定时空下彼此关联的各种事件。事件描述了特定时空内施动者的某种活动及其结果。例如,案例1包含两个事件,即丁丁在一段时间内步行到少年宫和爸爸在一段时间内骑车追赶。两者都可以刻画为施动者在一段时间内移动了某段距离。因此,时间和距离可以作为刻画该事件的基本维度(dimension)。活动的不同维度用直角坐标系作为分析框架(见图8的时间和距离)。

在直角坐标系内,不同事件的起终点可以表征为该事件在不同基本维度上的特定位置(place)。图8中,A和D表示丁丁被爸爸追上时所用时间的起点和终点,B和D则表示爸爸追赶小丁丁所用时间的起点和终点。事件在特定维度上的起点和终点定义了该维度上的一个特定区间(segment)。例如,AD这一区间定义了爸爸追上丁

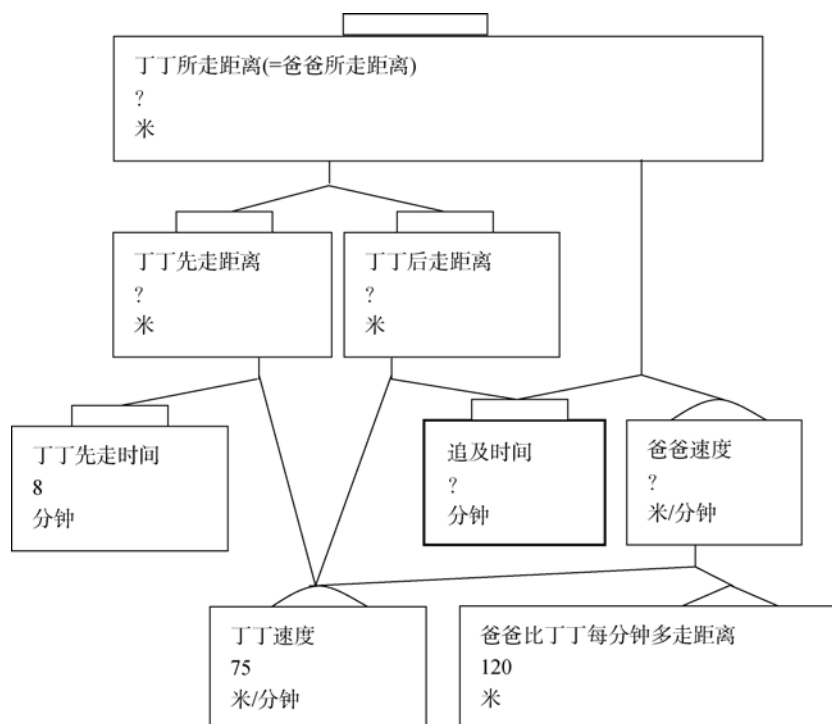


图7 另一种问题解决策略下案例1的数量元素结构图

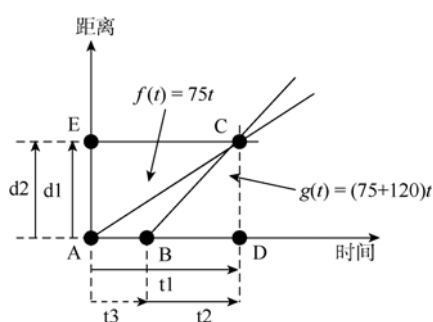


图8 案例1的问题情境结构

丁所用时间。在同一个维度上,不同事件之间的关系可以通过它们在该维度上对应区间的空间关系来理解和表征。同一维度不同区间属于共线关系(collinear)。两个区间的共线关系可以分为一致性(congruent)和相邻性(adjacent)两种情况。其中一致性又可以分为完全一致性和部分一致性。案例1中,丁丁所走距离(d_1)和爸爸所走距离(d_2)这两个区间具有相同的起点和终点,并且在每一点上都是重叠的,因而属于完全一致性关系。相比之下,时间维度上的两个区间 t_1 (丁丁所用时间)

和 t_2 (爸爸追击时间)属于部分一致性关系, t_3 (丁丁先走时间)和 t_2 属于相邻性关系。

事件在不同维度上的对应区间还具有方向性。方向性代表了事件在不同基本维度上发展变化(时间历程、位移顺序、增量序列等等)的实质意义。区间方向性用箭头表示(见图8)。例如, t_1 对应的带箭头线段表示时间区间 AD 从丁丁出发开始算起,到丁丁被爸爸追上时截止。共线关系和方向性相结合,使得我们可以推论 t_2 的起点是 t_1 区间中的某一点,继而推断爸爸出发时间是在丁丁出发后的某个时间点。关系-函数分析法规定所有的时间区间都有共线关系,且具有相同的方向。但对其他基本维度而言,同一维度的不同区间除了共线关系外,还存在方向性的不同。图9给出了案例2的问题情境结构示意图。其中 t_1 和 t_2 分别代表甲乙两车所走时间, d_1 和 d_2 分别代表甲乙两车所走距离。 t_1 和 t_2 为一致性共线关系,而 d_1 和 d_2 具有共同起点但方向相反,为相邻性共线关系。

案例2 两列火车同时从一个车站出发,相背而行。甲车每小时走100公里,乙车每小时走120公里。几小时后两列火车相距880公里?

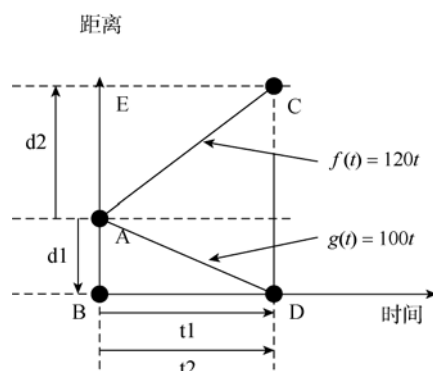


图9 案例2的问题情境结构

前面提到,集中变量反映了事物两个基本属性之间的某种映射关系,如“速度是单位时间内经过的距离”。在关系-函数分析法中,集中变量被表征为联接不同坐标轴的二维斜面(two-dimensional incline)。斜面斜率的变化反映了不同维度之间对应关系的不同。例如在图8中,AC给出了丁丁所走时间(t_1)与其所走距离(d_1)之间的对应关系。AC的斜率即是丁丁步行速度(75米/分钟)。它定义了 t_1 和 d_1 之间的一个函数关系(functional relation),即 $d = f(t) = 75t$,表征了 d_1 随着 t_1 变化而变化的速率。

维度、位置、区间、斜面构成了关系-函数分析法的基本要素,提供了分析代数应用题情境结构的一种语言形式。不仅如此,该方法还可以用来界定代数应用题情境结构的可能空间。例如,综合考虑行程问题中不同事件在两个基本维度(时间和距离)上对应区间之间的空间关系,可以将该类问题的情境结构分为四类:一致性距离区间和一致性时间区间、一致性距离区间和相邻性时间区间、相邻性距离区间和一致性时间区间、相邻性距离区间和相邻性时间区间。这四种类型加上区间方向性(相同或相反),可以刻画更多的问题情境类型。方向性可以通过距离区间(如图8和9)或者通过函数符号来表示。函数符号的正负变化(对应于斜面斜率的正负方向)可以表征方向性:两个都是正号,两个都是负号,以及一个正号和一个负号。三种函数符号的组合和上述四种区间空间关系相结合,可以形成12种不同的情境结构(Yerushalmy & Gilead, 1999)。每一种情况都对应于一类具有相同情境结构的代数应用题。实际上, Hall 等人(1989)曾运用该方法尝试分析代数应

用题,发现适用于 Mayer (1981)代数应用题分类框架中的多数问题类型。

3.4 任务分析地图法

任务分析地图法由 Chick 等人(Chick, 1988; Chick, Watson, & Collis, 1988)提出,本来是利用 SOLO 系统(Biggs & Collis, 1982)划分不同认知水平的一种视觉辅助手段。后来,经过多位研究者修改完善(Collis & Watson, 1991; Stillman, 1996; Watson & Mulligan, 1990),该方法逐渐成为一种可以分析数学问题结构的方法。与前面几种方法不同,该方法侧重于对问题解决过程的分析。

该方法将个体问题解决过程分为问题线索(cue)、过程(process)或概念(concept)、结果(result)三个部分。其中,问题线索包括问题所求未知数和问题提供的数量元素或相关概念,用符号▲表征。在解决问题的过程中,个体运用相关的概念或过程(表征为●)选择和加工不同的数量元素,推导出一个或多个问题解决的中介结果,并导致问题最终结果的形成(表征为■)。

运用该方法,案例1的问题解决过程可以用图10表示⁴。图中连线表示了数据元素在概念或过程应用下产生问题中介或最终结果的路径。可以看出,根据问题所求未知变量以及提供的数量元素,如“先走8分钟”或“行走速度75米/分钟”等,个体多次运用“相加”和“距离=速度×时间”等概念或公式,求出问题解决所需的各种中介变量或关系式。此外,正确解决问题需要用到问题陈述中蕴涵的一个基本概念,即在爸爸追上小丁丁的过程中,两个人行走的路线和距离都相同。这是形成最终关系式的核心概念。该图还表明,形成最终关系式需要求取4个中介变量(具体数值或相应表达式)。因此,任务分析地图法较为明确的表征了问题解决所需的数据来源、知识基础、解题策略及其中介环节。从任务特征的角度来看,该方法非常清晰的表明了问题中数量元素个数,解题过程中涉及的过程或概念类型和数量,以及中介结果数量等等。

任务分析地图法也可以揭示同一问题的不同解决途径(Watson & Mulligan, 1990)。如前所述,

⁴不同研究者所使用的表示方式有所不同。本文在 Watson & Mulligan (1990)的基础上略加修改。

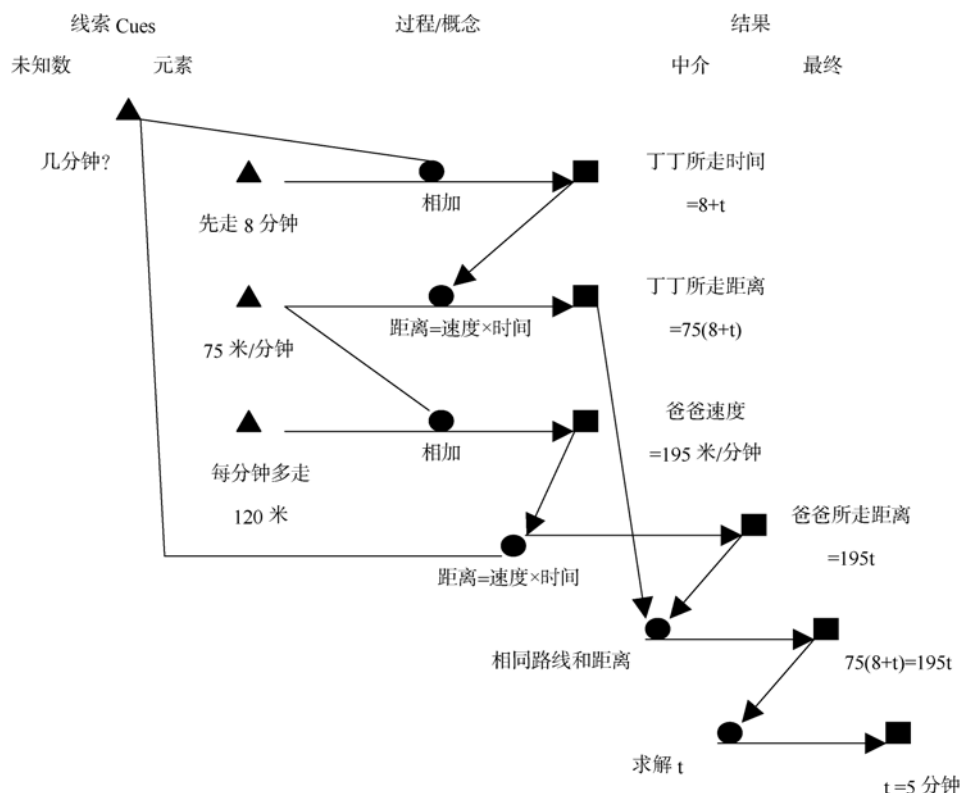


图 10 案例 1 的任务分析地图

案例 1 存在多种问题解决的策略。图 10 给出的只是其中的一种。刻画不同策略需要改变图 10 中数量元素的结合方式, 所运用的概念或原理, 以及得出最终结果所需产生的中介结果等。不仅如此, 该方法还可以用来分析个体实际的问题解决过程, 揭示不同个体在解决代数应用题时所知觉的问题线索、采用的概念和原理以及解题策略, 并可以从中分析他们可能出现的错误观念和策略, 发现和鉴别不同反应类型等(Stillman, 1996; Watson & Mulligan, 1990)。

4 项目生成视角下对上述方法的分析和研究启示

本研究试图站在理论驱动的认知项目设计范式下, 以代数应用题认知加工模型为基础, 探寻适合于代数应用题项目生成的、能够系统的、基于原则的(principled)提取任务特征的问题结构分析方法。该研究表明, 对应于认知加工模型中的语义结构、情境模型和问题模型等不同中介表征, 以及问题解决过程中的认知需求, 存在不同的基

于原则的问题结构分析方法。然而, 从项目生成的角度, 这些方法还只是提供了一个可能性。要实现项目生成特征提取的需求, 尚需对这些方法的心理现实性、系统性和完备性、适用范围以及如何整合等问题开展进一步研究。

4.1 问题结构分析的心理现实性

结构分析的心理现实性(psychological reality)是指通过某方法所揭示出的问题特征是否表征了个体在解决问题时真正经历的状态或过程。本研究中, 命题分析法、关系-函数分析法和网络语言分析法分别对应于问题陈述背后的语义结构、情境结构和变量结构, 而任务分析地图则揭示了个体解题过程中的路径、每个环节需要整合的变量、以及中介结果。从这个角度讲, 似乎这四种方法与问题解决认知模型之间具备了形式上的对应性。

但是, 心理现实性不仅仅限于形式上的对应, 还涉及到各种方法所揭示的特征是否与认知理论所预期的效应相一致。比如, 对案例 1 的命题分析表明, “爸爸追上小丁丁”是该问题最为核心的

宏观命题,而“小丁丁去少年宫”和“爸爸追赶(小丁丁)”是与该命题直接关联的两个核心命题。其他命题都是描述这三个命题的持续时间、行动状态等具体特征的微观命题。按照文本加工理论(Kintsch & van Dijk, 1978),宏观命题体现的是问题陈述的主旨,比微观命题在语义理解中加工程度更高,因而更容易为被试提取和回忆。新近研究证实了这一预测(马永鑫, 2013)。研究选择了12种不同类型的代数应用题随机呈现给40名初中学生,并采用命题分析法对每个代数应用题的语义结构进行了分析。后继回忆结果表明,跨越不同类型,学生对宏观命题的回忆率显著高于微观命题。类似的,Shalin & Bee (1985), Nesher & HersHKovitz (1994)等学者对网络语言分析法所揭示的代数应用题深层数学结构的心理现实性也做了实证研究,均在一定程度上验证了预期假设。但在一般意义上,仍需要对各种分析方法的心理现实性进行系统全面的研究。

4.2 问题特征分析的系统性和完备性

Hall 等人(1989)区分了问题分析的两种水平,即生成性(generative)和预测性(predictive)。生成性分析方法采用某种符号系统,描述了问题解决需要满足的各种限制条件。例如,要能够根据问题陈述列出相应的代数式,掌握表达代数式所需各种符号(如 x 代表未知数、 $+$ 代表加法运算等等)和限制(如加减乘除的运算等级、系数与未知数相乘的规定等)。但是没有回答“学生为什么或如何生成这一特定的代数式(Hall et al., 1989, p. 226)”。而要实现这一目的,就需要能够表征问题解决中介状态、选择特定解题策略或结果的限制条件。如果这些表征与个体实际认知活动高度一致,该方法就实现了预测功能。

依照这种观点, Hall 等人(1989)认为网络语言分析方法仅具有生成功能,而不具有预测功能。该方法明确表征了问题中的元素名称、类型、取值和单位,借助于元素类型限定了元素间可能的数量关系和可能经历的中介运算,以及通过连贯的网络实现从问题元素向解题代数式(或算式)的转换等。但是,该方法没有提供个体是如何从问题陈述中提取出这些元素来的,没有指出元素间的局部关系(一个三阶组合内)和整体结构(问题模型)是如何形成的。相比之下,关系-函数法提供了网络语言分析没有解答的信息。在案例1中(见图

8),个体基于常识能够意识到该问题描述了两个事件,即“丁丁步行去少年宫”和“爸爸骑车追赶丁丁”。在距离维度上,两个事件是一致性共线关系,且方向相同,在时间维度上,两个事件是部分一致性共线关系,且方向相同。这两点是个体形成如下两个推断的基础:(1)“爸爸追上丁丁”那一刻(点),两人所走距离是相同的,(2)“爸爸追赶丁丁”所用时间同时也是丁丁所用时间的一部分。这两个推断是个体解决该问题的必要基础,也是个体构建变量间关系的关键所在。出于这一原因, Hall 等人认为关系-函数法具有预测性功能。是否如此尚有待进一步验证。

4.3 不同分析方法的整合

从项目生成的视角,需要整合上述方法,实现对代数应用题解题过程及中介状态的完整表征,是后续研究的重要问题。实现不同方法合理整合的基础依然是代数应用题的认知加工模型。根据该模型,个体在面对问题时,首先是解码问题陈述,形成语义层面的理解。在这一过程中,命题的构建主要基于问题文本和个体的语言知识。问题文本是指在问题陈述中明确表达的内容。个体语言知识是指个体如何分析句法,以及如何基于句法分析提取陈述背后的基本观念等。但在这一过程中,个体的世界知识和对问题情境的经验也具有非常重要的作用。事实上,比较图2和图8可以看出,命题结构图揭示了案例1的三个宏观命题,即“爸爸追上小丁丁”,“小丁丁去少年宫”和“爸爸追赶(小丁丁)”。这和关系-函数分析法所揭示的两个事件是相同的。这说明两种方法在宏观层面所揭示的结构上是一致的。但是,两种方法在微观层面上存在不同。关系-函数法突出了问题情境中事件在不同维度上的关系,而命题分析法突出了围绕宏观命题,对施动者行为状态、方式或数量特征的修饰或限制。从项目生成的角度看,一个关键的问题是,给定某种问题情境中事件的时空关系,是否只有一种对应的命题结构,还是存在多种命题结构?如果是后者,不同命题结构的影响是什么?

类似的,网络语言分析法可以揭示问题陈述背后的变量结构。通常,变量结构相同但问题情境不同的问题被称为同形异态问题(isomorphic problems) (Eysenck & Keane, 2000)。研究发现,即使都是同形异态问题,但问题主题背景不同时,

个体正确识别率会有很大差异。比如,当两个问题都属于工作问题(work problem)时,个体识别出它们属于同形异态问题的正确率为76%,而当两个问题都属于混合问题(mixture problem)时,相应的正确率只有38%(Reed, 1987)。此外,即使数学结构和问题情境都相同,问题陈述方式的不同也会导致解题正确率上。例如,案例3和案例4除了问题陈述不同之外,在数学结构和问题情境方面完全相同。

案例3 有5只鸟和3条虫。假设这些鸟都来抢虫吃,每个鸟都想吃到一条虫。有几个鸟会吃不到?

案例4 有5只鸟和3条虫,鸟比虫多几个?

但研究发现,绝大多数幼儿园大班儿童和一年级小学生能够答对案例3(幼儿园大班:83%;一年级学生:100%),但是只有约五分之一的大班儿童和约半数一年级学生答对案例4(幼儿园大班:17%;一年级学生:64%)(Hudson, 1983)。它表明在某些情况下,问题情境的呈现方式本身存在某些启动特征(enabling feature),能够深刻的影响个体对问题结构的理解和对问题解决策略的选择,进而影响问题解决结果。这些都表明,数学结构无法解释个体问题解决复杂性的全部。因此,如何借助于这些方法剥离问题认知复杂性的不同来源及其影响,这是后续研究需要关注的。

4.4 结构分析方法的适用范围

整合不同的结构分析方法,涉及到不同方法的适用范围是否一致。假如图1所描述的代数应用题认知模型适用于整个任务领域,那么理想状态下,整合后的问题结构分析方法也应该适用于整个领域。这样,才能形成领域范围的项目设计原则(item design principle)以及一系列的项目结构(item structure)。就目前初步研究的情况来看,命题分析方法和网络语言分析方法的适用范围是比较广泛的(卫芸, 2011; 马永鑫, 2013),而其他两种方法尚有待进一步研究。

本文在自动化项目生成的认知设计系统法这样一种理论模式下,虽然只是就代数应用题这一任务类型的项目特征鉴别和问题结构分析开展了研究,但是该研究此处的讨论适用于当前自动化项目生成研究所面临的普遍问题。目前,自动化项目生成越来越成为测量学中比较活跃的研究领域(Alves, Gierl, & Lai, 2010; Gierl & Haladyna,

2012)。自动化项目生成所研究的任务类型也从20世纪末期智力测验中的各种过程密集型(process-intensive)项目逐渐过渡到学业成就测验中的知识密集型(knowledge-intensive)项目。但是,应该看到,许多研究虽然都强调了在项目认知模型基础上提取任务特征,也针对具体任务类型提出了许多对任务测量学特征有影响的设计变量,但是对这些特征或变量的心理现实性、系统性和完备性方面的理论论证和经验验证尚有待深入。多数研究还停留在借助于个别项目例证特定任务类型自动化项目生成的理论可行性上,较少在项目认知模型所涵盖的整个任务领域探索算法式项目生成的可能性,及其对认知模型背后测量建构的操作广度和深度。本研究是面向这一方向的初步探索,也希望有更多的项目生成研究关注这些问题。唯如此,自动化项目生成研究才有可能真正成为测量建构的认知理论和个别差异研究的中介桥梁,也才有可能真正实现用理论驱动的测量学范式解决现实测量问题的要求。

参考文献

- 马永鑫. (2013). *代数应用题的回忆实验研究*. 硕士学位论文, 华东师范大学.
- 卫芸. (2011). *代数应用题的类型、结构与分布研究*. 硕士学位论文, 华东师范大学.
- 杨向东. (2013). 代数应用题项目生成中的认知过程和任务特征分析. *心理科学进展*, 21, 175-189.
- Alves, C. B., Gierl, M. J., & Lai, H. (2010). *Using automatic item generation to promote principled test design and development*. Paper presented at Paper presented at the annual meeting of the American Educational Research Association, Denver, CO, USA.
- Arendasy, M., & Sommer, M. (2007). Using psychometric technology in educational assessment: The case of a schema-based isomorphic approach to the automatic generation of quantitative reasoning items. *Learning and Individual Differences*, 17, 366-383.
- Bejar, I. I. (2002). Generative testing: From conception to implementation. In S. H. Irvine, & P. Kyllonen (Eds.), *Generating items for cognitive tests: Theory and Practice* (pp. 199-218). Mahwah, NJ: Erlbaum.
- Bejar, I. I., & Yocom, P. (1991). A generative approach to the modeling of isomorphic hidden-figure items. *Applied Psychological Measurement*, 15, 129-137.
- Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic

- Press.
- Bormuth, J. R. (1970). *On the theory of achievement test items*. Chicago: University of Chicago Press.
- Bouwmeester, S., & Sijtsma, K. (2004). Measuring the ability of transitive reasoning, using product and strategy information. *Psychometrika*, 69, 123–146.
- Bovair, S., & Kieras, D. E. (1985). A guide to propositional analysis for research on technical prose. In B. K. Britton, & J. B. Black (Eds.), *Understanding expository text* (pp. 315–362). Hillsdale, NJ: Erlbaum.
- Chick, H. L. (1988). Student responses to a polynomial problem in the light of the SOLO Taxonomy. *Australian Senior Mathematics Journal*, 2, 91–110.
- Chick, H. L., Watson, J. M., & Collis, K. F. (1988). Using the SOLO Taxonomy for error analysis in mathematics. *Research in Mathematics Education in Australia*, 34–46.
- Collis, K., & Watson, J. (1991). A mapping procedure for analyzing the structure of mathematics responses. *Journal of Structural Learning*, 11, 65–87.
- Daniel, R. C., & Embretson, S. E. (2010). Designing cognitive complexity in mathematical problem-solving items. *Applied Psychological Measurement*, 34, 348–364.
- Deane, P., & Sheehan, K. (2003). *Automatic item generation via frame semantics: Natural language generation of math word problems*. Paper presented at the National Council on Measurement in Education Annual Meeting, Chicago, IL, USA.
- Embretson, S. E. (1994). Applications of cognitive design systems to test development. In C. R. Reynolds (Eds.), *Cognitive assessment: A multidisciplinary perspective* (pp. 107–135). New York: Plenum Press.
- Embretson, S. E. (1998). A cognitive design system approach to generating valid tests: Application to abstract reasoning. *Psychological Methods*, 3, 380–396.
- Embretson, S. E. (2004). The second century of ability testing: Some predictions and speculations. *Measurement*, 2, 1–32.
- Embretson, S. E., & Gorin, J. (2001). Improving construct validity with cognitive psychology principles. *Journal of Educational Measurement*, 38, 343–368.
- Embretson, S. E., & Yang, X. (2007). Automatic item generation and cognitive psychology. In S. Sinharay, & R. Rao. (Eds.), *Handbook of Statistics: Psychometrics* (pp. 747–768). Amsterdam, The Netherlands: Elsevier, B.V.
- Enright, M. K., Morley, M., & Sheehan, K. M. (2002). Items by design: The impact of systematic feature variation on item statistical characteristics. *Applied Measurement in Education*, 15, 49–74.
- Eysenck, M., & Keane, M. T. (2000). *Cognitive Psychology: A student's handbook*. (4th ed.). Hove, UK: Psychological Press.
- Gierl, M. J., & Haladyna, T. M. (2012). *Automatic item generation: Theory and practice*. New York, NY: Routledge.
- Hall, R., Kibler, D., Wenger, E., & Truxaw, C. (1989). Exploring the episode structure of algebra story problem solving. *Cognition and Instruction*, 6, 223–283.
- Hinsley, D., Hayes, J. R., & Simon, H. A. (1977). From words to equations—Meaning and representation in algebra word problems. In P. Carpenter, & M. Just (Eds.), *Cognitive processes in comprehension* (pp. 89–106). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hively, W., Patterson, H. L., & Page, S. H. (1968). A “universe-defined” system of arithmetic achievement tests. *Journal of Educational Measurement*, 5, 275–290.
- Hudson, T. (1983). Correspondences and numerical differences between disjoint sets. *Child Development*, 54, 84–90.
- Irvine, S., & Kyllonen, P. (2002). *Generating items for cognitive tests: Theory and Practice*. Mahwah, NJ: Erlbaum.
- Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review*, 92, 109–129.
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363–394.
- Koedinger, K. R., & Nathan, M. J. (2004). The real story behind story problems: Effects of representations on quantitative reasoning. *Journal of the Learning Sciences*, 13(2), 129–164.
- Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working memory capacity?! *Intelligence*, 14, 389–433.
- Leighton, J., & Gierl, M. (2007). *Cognitive diagnostic assessment in education: Theory and applications*. New York, NY: Cambridge University Press.
- Lohman, D. F., & Ippel, M. J. (1993). Cognitive diagnosis: From statistically based assessment toward theory based assessment. In N. Frederiksen, R. J. Mislevy, & I. I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 41–71). Hillsdale, NJ: Erlbaum.
- Mayer, R. E. (1981). Frequency norms and structural analysis of algebra story problems into families, categories and templates. *Instructional Science*, 10, 135–175.
- Mayer, R. E., Larkin, J. H., & Kadane, J. B. (1984). A cognitive analysis of mathematical problem-solving ability. In R. J. Sternberg (Eds.), *Advances in the psychology of human intelligence* (pp. 231–273). Hillsdale, NJ: Lawrence

- Erlbaum Associates.
- Mislevy, R. J. (1993). Foundations of a new test theory. In N. Frederiksen, R. J. Mislevy, & I. I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 19–39). Hillsdale, NJ: Erlbaum.
- Nathan, M. J., Kintsch, W., & Young, E. (1992). A theory of algebra–word-problem comprehension and its implications for the design of learning environments. *Cognition and Instruction*, 9, 329–389.
- Nesher, P., & Hershkovitz, S. (1994). The role of schemes in two-step problems: Analysis and research findings. *Educational Studies in Mathematics*, 26, 1–23.
- Quintero, A. H. (1981). The role of semantic understanding in solving multiplication word problems: Two step R E problems. In J. L. Schwartz (Eds.), *The role of semantic understanding in solving multiplication and division word problems* (pp. 21–36). National Institute of Education Final Rep. No. NIE-G-80-0144.
- Reed, S. K. (1987). A structure-mapping model for word problems. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 13, 124–139.
- Rupp, A. A., Templin, J., & Henson, B. (2010). *Diagnostic measurement: Theory, methods and applications*. New York, NY: The Guilford Publication Inc.
- Sebrechts, M. M., Enright, M., Bennett, E., & Martin, K. (1996). Using algebra word problems to assess quantitative ability: Attributes, strategies and errors. *Cognition and Instruction*, 14, 285–343.
- Shalin, V., & Bee, N. V. (1985). *Analysis of the semantic structure of a domain of word problems* (Tech. Rep. No. APS-20). Pittsburgh: University of Pittsburgh, Learning Research and Development Center.
- Stillman, G. (1996). Mathematical processing and cognitive demand in problem solving. *Mathematics Education Research Journal*, 8, 174–197.
- Turner, A., & Greene, E. (1977). *The construction and use of a propositional text base* (Tech. Rep. No. 63). Boulder: University of Colorado, Institute for the Study of Intellectual Behavior.
- Watson, J. M., & Mulligan, J. (1990). Mapping solutions to an early multiplication word problem. *Mathematics Education Research Journal*, 2, 28–44.
- Whitely, S. E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin*, 93, 179–197.
- Wilson, M. (2013). Seeking a balance between the statistical and scientific elements in psychometrics. *Psychometrika*, 78, 211–236.
- Yerushalmy, M., & Gilead, S. (1999). Structures of constant rate word problems: A functional approach analysis. *Educational Studies in Mathematics*, 39, 185–203.

Toward A Principled Structure Analysis Method of Algebra Story Problems for Item Generation

YANG Xiangdong

(Institute of curriculum and instruction, East China Normal University, Shanghai 200062, China)

Abstract: Automatic item generation is a principled approach to item design that is grounded on cognitive theory of item solution, of which a key step is to identify and extract task features systematically via principled structure analyses of the items. The current study aims to develop such a principled structure analysis method of algebra story problems that serves the purpose of automatic item generation by synthesizing four structure analysis methods that are currently available in the literature, that is, propositional analysis, network language analysis, relation-functional analysis and task analysis map. It is shown that task features extracted from the first three methods correspond to the three types of interim mental representations formed during algebra story problem solving, that is, semantic structure, situational model and problem model, respectively, whereas those from the fourth reveal the cognitive demands at various stages of problem solving. However, further studies are required to investigate the psychological reality, thoroughness, and the range of applicability for such methods before an integrated method of structure analysis of algebra story problems can be formulated for automatic item generation.

Key words: automatic item generation; algebra story problem; cognitive model; principled structure analysis method