Comprehending process diagrams in biology education

Kragten, M.

Publication date
2015
Document Version
Final published version

CHAPTER 2

DIAGRAMMATIC LITERACY IN SECONDARY SCIENCE EDUCATION*

Students in secondary Science education seem to have difficulties with understanding diagrams. The present study focused on explanatory factors that predict students' difficulties with process diagrams, i.e., diagrams that describe a process consisting of components that are related by arrows. From 18 compulsory national Biology exams of secondary school pre-university students all process diagram tasks (n = 64) were included in corpus. Features of the task, student, and diagram were related to the difficulty of that particular task, indicated by the cohort mean exam score. A hierarchical regression analysis showed main effects for (1) the cognitive task demand, (2) the familiarity of the components, and (3) the number of components in a diagram. All these main effects were in the expected direction. We also observed interactions. Within the category of tasks with a high cognitive demand, tasks about a diagram of which students have low prior content knowledge were more difficult than tasks about a diagram of which students have high prior content knowledge. Tasks with a high cognitive demand about a diagram with familiar arrows were, surprisingly, more difficult than tasks with a high cognitive demand about a diagram with unfamiliar arrows. This latter finding might be attributed to compensation for task difficulty by the large number of components in the diagrams involved. The final model explained 46 percent of the variance in exam scores. These results suggest that students have difficulties (1) with tasks that require a deeper understanding when the content is new, (2) with diagrams that use unfamiliar component conventions, and (3) with diagrams that have a small number of components and are therefore probably more abstract.

1. INTRODUCTION

Diagrams are effective learning tools (Winn, 1991). They can help build a mental model and can make abstract ideas more concrete by triggering learners to use their spatial skills. Research shows that diagrams support learners' self-explanation (Ainsworth & Loizou, 2003), inference generation, and information integration and that they do reduce comprehension errors (Butcher, 2006; Cromley, Snyder-Hogan, & Luciw-Dubas, 2010). Larkin and Simon's (1987) comparison of sentence processing and diagram processing models shows how clustering and placement of components in diagrams make it easier to find information and to use it effectively. Once the first piece of relevant information has been found in a diagram, it is very likely that the next piece will also be found. However, many studies report that students have difficulties with diagram interpretation (Bowen & Roth, 2002; Chittle-

* Kragten, M., Admiraal, W., & Rijlaarsdam, G. (2013). Diagrammatic literacy in secondary science education. *Research in Science Education, 43*, 1785–1800. doi:10.1007/s11165-012-9331-0

borough & Treagust, 2008; Guthrie, Shelly, & Kimmerly, 1993; Kindfield, 1993; Mathai & Ramadas, 2009; Schönborn, Andersen, & Grayson, 2002).

Interpreting diagrams is a key competence for students in mastering many of the biological theories. Bezemer and Kress (2008) showed that images are increasingly prominent as carriers of meaning in students' textbooks. In Biology course books, process diagrams explain processes like protein synthesis, immunology, photosynthesis, cellular respiration, compound cycles, etc. (Bowen & Roth, 2002; Campbell & Reece, 2002). In the present study we will study these biological process diagrams, defined as dynamic diagrams that describe a biological process consisting of components that are related by arrows.

Students are often not explicitly trained in how to interpret diagrams. They should be made aware of the conventions of diagrammatic representations and the scope and limitations of each diagrammatic mode (Gilbert, 2005; Winn, 1991). Ainsworth (2006) states that learners who are first presented with a novel representation must understand how it encodes information and how it relates to the domain it represent. It makes sense to provide opportunities for students to actively investigate diagrams and to build up adequate strategies for interpretation. Designers of such a program cannot rely on information on the difficulties students have in tasks that require an interpretation of biological process diagrams, as a thorough investigation of students' difficulties on the interpretation of these specific types of diagrams has not been performed yet.

The present, exploratory, study fills this gap in knowledge, examining three factors (i.e., the task, student, and diagram), and the extent in which these explained variances in difficulty, indicated by the cohort mean exam scores of students on diagram tasks.

2. THEORETHICAL FRAMEWORK

In the next three sections we will present research on features of the task, student, and diagram, and their effect on learning and performance.

2.1 Task features

A classic way of classifying the level of cognitive demand of a task is by using Bloom's Taxonomy (Bloom, Krathwohl, & Masia, 1956). This taxonomy distinguished six categories: knowledge, comprehension, application, analysis, synthesis and evaluation. Krathwohl (2002) revised the original taxonomy into a two-dimensional framework, with the dimensions knowledge and cognitive processes. Crowe, Dirks, and Wenderoth (2008) developed a 'Blooming Biology Tool' that can be used to assess the level of difficulty on tasks for several domain specific skills.

Guthrie et al. (1993) investigated the cognitive processing of several displays, e.g. graphs, pictures and static diagrams, under two different task conditions, i.e., local and global searches. For local searches only a part of the display needs to be explored and the cognitive demand is specific and precise: Local searches demand discrimination between categories. Global searches involve the interpretation of the entire display and found to be more difficult than local searches. The latter effect is in line with the cognitive load theory (Sweller, 1994). Winn (1993) concludes that different search goals lead to different search patterns while readers look at diagrams. Goals set search strategies and effective search strategies requires knowledge of content and of the conventions of diagrams.

We will incorporate the findings of the presented studies about task features into the present study's definition of cognitive task demand.

2.2 Student features

A student's familiarity with the design of a diagram might be an important factor for learning and understanding (Carlson, Chandler, & Sweller, 2003). In this context, familiarity refers to the likeliness that students have a cognitive schema available that supports the interpretation of the diagram. Previous research suggests that familiarity with the spatial arrangement of the components and with the component type of a diagram has a positive effect on performance on learning tasks. Winn (1982) reversed the order of an evolutionary dinosaur diagram which decreased performance. In another study (Winn, 1988), participants performed worse on recall tasks when components were presented as unfamiliar icons rather than as labeled boxes. Winn, Li, and Schill (1991) found that participants who are familiar with the terms and conventions of family tree diagrams solved kinship problems faster. They conclude that familiar diagrams do not require the computation of solutions from rules and that diagrams should be used with students who already possess some knowledge of the terms and conventions.

Winn and Sutherland (1989) examined the effects of varying the familiarity and number of elements in maps and diagrams on the recall of lists of elements and their location. They found no main effect for familiarity, but high-ability participants used more helpful strategies like chunking and story construction than low-ability learners when presented an unfamiliar diagram. They conclude that there is some evidence that differences in performance are related with students selection of helpful or non-helpful strategies.

Prior knowledge of the content and the processes presented in the diagram might also affect student performance (e.g., Cook, Wiebe, & Carter, 2008) . Hegarty and Just (1989) focused on similar graphical displays as Mayer and Gallini (1990) and found that individual differences in prior knowledge affected diagram comprehension. Compared to participants with a low prior knowledge, participants with a high prior knowledge were more capable of locating the relevant information in a diagram and extracting information more selectively. Winn (1993) argued that prior knowledge facilitates analysis of a diagram in two ways. First, knowledge schemata inform readers what to look for and where the information is most likely to be found. Secondly, prior knowledge provides a structure in which information found can be organized. Lowe (1996) found that the difference between interpretation processes, i.e., the search for information and the interpretation and construction of weather maps, was affected by the level of prior content knowledge. An eye fixation study (Canham & Hegarty, 2010) complements these findings by concluding that information selection improves after instruction about relevant meteorological principles.

2.3 Diagram features

A biological process diagram consists of components that are related by arrows: Arrows can express many relations, such as pointing or connecting, sequence, change over time, path, or manner of movement or forces (Heiser & Tversky, 2006). The function of an arrow in a process diagram is usually conveyed by a label.

An early study by Holliday, Brunner, and Donais (1977) compared the effectiveness of two types of process diagrams that illustrated water, oxygen, nitrogen, and carbon dioxide cycles. In one diagram type, the components were text boxes with labels; in the other, they were iconic drawings with labels. In a number of retention and comprehension tasks the iconic drawings were found to lead to a significantly better performance, but that this effect was limited to low-ability students. Winn and Sutherland (1989) found that low-ability students remembered lists and locations of labeled elements better when they were shown as drawings rather than as squares. These results are supported by the dual coding theory of Paivio (1990) which suggests that the presentation of two modes of information, i.e., verbal and visual, promotes effective use of working memory.

The number of components and arrows in process diagrams can vary extensively. Cognitive load theory (Sweller, 1994) predicts a high intrinsic cognitive load if interaction between many elements, i.e., components, must be learned. Limited working memory makes schema acquisition, which is a major learning mechanism, difficult when multiple elements interact. Winn (1988) found that the number of components in a diagram had a significant effect on the recall of names and positions.

In their landmark study on diagram design principles, Mayer and Gallini (1990) found that students, mainly with low prior-knowledge levels, performed better on conceptual recall and problem solving questions when the illustration contained labels that describe the parts and steps of how mechanical devices such as a pump and a gear system works. Mayer and Moreno (1998) argue that when text and diagrams are separated, learners have to read some portion of the text and then maintain it in their working memory while reading the diagram. This places higher cognitive demand on working memory and increases the possibility that, because of working memory limitations, some information will be lost.

Winn (1991) presents a framework for learning from maps and diagrams which focuses on pre-attentive perceptual organization. In this framework, configuration and discrimination of components are fundamental when perceiving diagrams. The meaning of a diagram cannot be interpreted before the components have been configured and discriminated.

Configuration is the spatial relationship among components in a diagram. It determines which components appear to form clusters, the sequence in which components are processed, and which components receive the most attention. Clustering of components has been found to be effective for memorization tasks. In Biology, clustering often distinguishes certain components as functional groups or it is used to depict a specific location of a sub-process. Discrimination is the ease with which a component can be distinguished from another. In diagrams, discrimination can be supported by using many characteristics like color, shape, and size.

In the current study, the relationship was examined between features of the task, student, and diagram on the one hand and difficulty of a diagram task on the other hand. We formulated the following hypotheses:

1) Task features: Cognitive task demand does explain differences in task difficulty.
2) Student features: Diagram familiarity does explain differences in task difficulty, and prior content knowledge does explain differences in task difficulty.
3) Diagram features: The presented studies on diagram features usually focus on the effect of task achievement when a single element in a diagram is changed, e.g. label or no labels (Mayer & Gallini, 1990); text boxes or iconic drawing (Holliday, Brunner, & Donais, 1977). The present study included a large number of diagrams that can vary on many elements from each other. This large variability makes it difficult to predict the effect of any specific element or combinations of elements on task difficulty so no hypotheses were formulated for diagram features.

We also expected that difficulty of a diagram task can be explained by the interaction of cognitive task demand on the one hand and student features and diagram features on the other. It seems reasonable to expect that differences in diagram features have a larger effect on difficulty when the task is more cognitive demanding than when the task is less cognitive demanding (e.g., Mayer & Gallini, 1990; Carlson et al., 2003). We also find it reasonable to expect that the effect of cognitive demand is larger when prior knowledge is relatively low than when prior knowledge is relatively high (e.g., Canham & Hegarty, 2010), and that the effect of cognitive demand is larger when familiarity is relatively low than when familiarity is relatively high (e.g., Winn et al., 1991). In other words, when the task is more cognitive demanding, the effect of student and diagram variability is more pronounced.

3. METHOD

3.1 Data collection

Data from the Dutch compulsory national exams were collected by Cito – National Institute for Educational Measurement and can be obtained for research purposes. For the present study, the student's scores per task on 18 Dutch Biology national exams (Table 2.1) from the period 2001-2009 were obtained, i.e., there were two exams each year, a regular and a retake exam. The participants of the obtained Dutch Biology national exams were secondary school pre-university students with an average age of 18 years old.

The nine cohorts, i.e. 2001-2009, can be assumed to be similar with respect to student composition. For the last three years of their study in secondary education, the students chose Biology as one of their major topics within their exam program for which they received 480 hours of training. The biology curriculum in the Netherlands is well described on a national level. All students in the Netherlands follow a similar curriculum with respect to the content, although teaching approaches may and will vary of course.

Table 2.1. Summary of the national Dutch biology exams and data obtained from the Cito

Year	Exam	Tasks in exam	N/M size	Diagram tasks selected
2001	1	40	728/2145[a]	2
2001	2	45	27/27[a]	2
2002	1	40	1784/10865	7
2002	2	44	180/180	2
2003	1	38	1846/11073	5
2003	2	41	198/198	4
2004	1	41	1814/11893	4
2004	2	40	282/282	0
2005	1	39	1940/12444	8
2005	2	37	258/258	4
2006	1	36	6264/12804	6
2006	2	41	418/418	3
2007	1	35	7153/13798	6
2007	2	38	392/392	1
2008	1	37	8070/15288	3
2008	2	37	442/442	2
2009	1	35	10624/17539	3
2009	2	40	471/471	2
	Total	704	42891/110517	64

Note. N = sample size; M = population size; 1 = regular exam; 2 = retake exam.
[a]These samples are relatively small compared with the other regular and retake exams in the dataset because this was a transition year to the new exam program

For this research, student's scores per task from the regular and the retake Dutch Biology national exams 2001-2009 were used as data. The data obtained from the regular exams were large simple random samples (Hoyle, Harris, & Judd, 2002) of the scores per student per question; the data from the retake exam contained the entire population. The exams included a total number of 704 tasks of which 64 tasks required direct interpretation of a process diagram. All 64 tasks, selected by the first author (i.e. a part-time high school Biology teacher with 10 years' experience), were included in the present study.

Each task focused on a specific problem but the answer could require several steps: Each correct step was rewarded with a point. In the exams the information of the maximum score a student could achieve was presented to them in the left margin of the task: The maximum score of each of the 64 included tasks varied from 1 till 4. From the student scores on these 64 tasks we calculated the cohorts' mean score for each task, linearly transformed to a range from 0-1. This score was used as the dependent variable as we understood it as indicating diagrammatic task difficulty, i.e., a low cohort mean score indicates a high diagrammatic task difficulty.

3.2 Explanatory variables

The explanatory variables are presented in Table 2.2. *Exam* was included in the analysis as a covariate as we wanted to control for the possible differences between the students in the regular exam and the students who participate in the retake exam. Students who register for the Biology retake exam do that for various reasons. They can have failed their regular Biology exam, but also can want to increase their grade for Biology to compensate for a low grade on another subject (e.g., Chemistry or French), or they need an excellent grade to be selected for fixed-number studies in higher education like Medicine.

3.3 Task features

The explanatory variable *Cognitive task demand* was based on the studies from Guthrie et al. (1993), Bloom et al. (1956), and Crowe et al. (2008). We defined two levels of *Cognitive task demand*: tasks with a 'low' and tasks with a 'high' cognitive demand.

We defined a 'low' *Cognitive task demand* when only a few elements needed to be explored. Once the relevant information was found in the diagram, finding a correct answer asked for little cognitive processing. This meant that this type of task only required information recall or understanding of a concept or terms. Examples of this task type in the dataset were summarizing the elements found, labeling elements, describing step-by-step a part of the process and/or some simple arithmetic like adding or subtracting amounts, such as the amount of energy transported to a compartment in an ecosystem.

A 'high' *Cognitive task demand* was usually a more global task; a larger part or the entire diagram needed to be explored. Once the selected information was found it must be processed in working memory and integrated (evaluated, inferred, compared, judged) with prior knowledge. The task might require a prediction of the most likely outcome given a new situation of the process or an interpretation of the data and a selection of the best conclusion. An example of a task with a 'high' cognitive demand from the regular Biology exam 2002 was: "Explain—solely on the basis of the diagram—that an increase of dead material can lead to an increase of the number of lemmings". This category also included 'productive' tasks, i.e., tasks where students had to add arrows or components to an existing diagram or create a diagram from a set of given components.

The tasks were coded for Cognitive task demand by two expert and experienced Biology teachers (Cohen's $\kappa = .87$, with a 95% confidence interval of $.63 < \kappa < 1.0$) and agreement was found on all items afterwards.

Table 2.2. The number of tasks (N) and task difficulty (cohort mean score and standard deviation) for the explanatory variables

Explanatory variable	*N*	Task difficulty *M*	Task difficulty *SD*
Exam			
Regular[a]	44	.57	.17
Retake	20	.44	.26
Task features			
Cognitive task demand			
low[a]	44	.60	.19
high	20	.36	.17
Student features			
Familiarity diagram components			
familiar[a]	47	.56	.22
unfamiliar	17	.45	.17
Familiarity diagram arrows			
familiar[a]	58	.53	.22
unfamiliar	6	.48	.13
Familiarity spatial arrangement			
familiar[a]	52	.55	.22
unfamiliar	12	.43	.15
Prior content knowledge			
high[a]	47	.55	.19
low	17	.48	.27
Diagram features			
Arrow type			
movement	21	.48	.18
step[a]	18	.60	.20
transfer	18	.49	.26
feedback	7	.56	.18
Component type			
iconic[a]	17	.58	.19
symbolic	27	.49	.23
text boxes	20	.54	.20
Clustered			
yes[a]	21	.52	.20
no	43	.53	.22
Process labels			
labeled[a]	24	.54	.20
unlabeled	40	.52	.22
Sequence			
linear[a]	30	.53	.22
cyclic	34	.52	.21
Components		9.27[b]	5.61
Arrows		11.00[b]	7.07

Note. [a]*Reference group.* [b]*This is the mean number of components and arrows.*

3.4 Student features

Familiarity of diagrams referred to the familiarity with components, arrows and spatial arrangement of a diagram. In the national Dutch Biology exams, diagrams might have specific conventions for components, arrows or spatial arrangement which a student is not familiar with. In such a case a student has to invest in understanding these conventions to perform on a task. For example, components might be combined in an unusual way or arrows might have a different meaning than in a conventional diagram about the same topic. There are also many examples of diagrams that were cross topic, e.g. nervous system and blood circularly system combined, or that used multiple conventions for components and arrows in one diagram.

The spatial arrangement of components in a diagram can be unfamiliar in various ways: the orientation might be right to left instead of left to right, bottom to top instead of top to bottom, linear instead of circular (or vice versa), etc. It could also be that components are clustered in an unusual manner. The familiarity of the components, arrows and spatial arrangement of a diagram were coded by the same two expert and experienced Biology teachers. Interrater reliability (Cohen's κ) for the familiarity of components, arrows and spatial arrangement were .86, with a 95% confidence interval of $.60 < \kappa < 1.0$, .82, $.48 < \kappa < 1.0$, and .87, $.63 < \kappa < 1.0$, respectively.

Prior knowledge of the content was defined to be 'high' if the information presented in the diagram was part of the curriculum. Although the Dutch Biology curriculum for pre-university education is well defined, it might, however, be that some topics that were not in the curriculum were considered to be part of the prior knowledge of the students. This variable was coded by the two teachers as well (Cohen's $\kappa = .86$, with a 95% confidence interval of $.60 < \kappa < 1.0$).

3.5 Diagram features

Arrows in a biological process diagram might have different meanings depending on the context, such as 'transport', 'causality', 'transformation', or 'feedback' interaction (Figure 2.1). Arrows that focus on 'transport' represent the flow of energy or matter from one component to another component. The arrows can be labeled with the amount of energy or matter that is transported per time or with the represented processes. Arrows that represent 'causality' describe step by step processes; a process in one component has an effect on the next component. The arrows represent the process and could be labeled with the name of the process. A sense of movement could be created by clustering, e.g. from the nucleus to the cytoplasm (see the 'causality' panel in Figure 2.1).

The 'transformation' arrow type describes the transformation of the compartment itself, e.g. a chemical compound. The arrows represent the chemical reactions and were sometimes labeled with the process name (e.g. hydrolysis) or the responsible enzyme (e.g. amylase); sometimes the arrows also depicted movement from one location to another by using clustering.

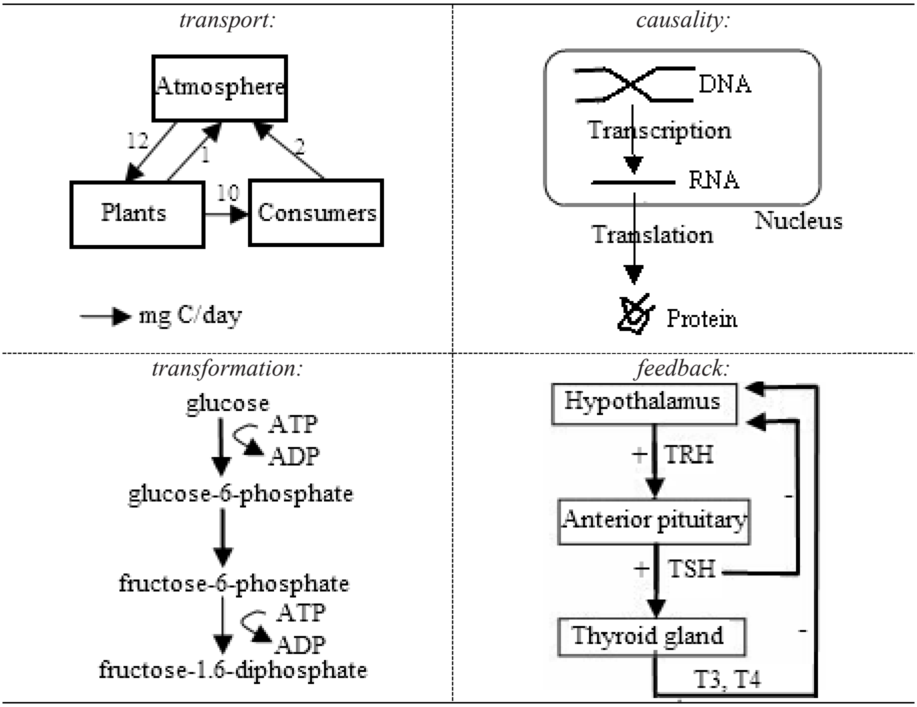

Figure 2.1. Arrow types and exemplary diagrams. The presented diagrams are exemplary and where not part of the actual dataset. Some examples of topics in the actual dataset with these arrow types are: transport: food webs, diffusion; causality: nervous system, immune system; transformation: assimilation/dissimilation; feedback: homeostasis, climate.

The 'feedback' arrow type was typically used in diagrams explaining hormonal processes. Compartments of hormone diagrams represented the glands or sometimes the hormone itself and the arrows represent production of the hormone and the type of feedback, i.e. inhibition or stimulation, of the hormone on the production of itself or other hormones. The arrows were often labeled by a plus or minus sign representing the direction of the feedback. Feedback arrows are also common in climate models, e.g. the effect of the carbon dioxide emission on sea level rise or photosynthesis.

Finally, the components of a diagram were classified as 'text boxes', 'symbolic' or 'iconic'. If some components were spatially separated from others to form a functional group or to depict a location then we considered clustering to be present in the diagram. Diagrams with arrows that were labeled directly or in indirectly by using a legend, e.g. a letter/number on the arrow corresponding to a legend item, were categorized as labeled. If a diagram could only be read in a single direction then the diagram was categorized as linear instead of cyclic. Diagram features were very distinct and were therefore only coded by the first author.

3.6 Data analysis

A hierarchical multiple regression analysis was performed with task difficulty as the dependent variable, features of the task, student, and diagram as explanatory variables and exam (regular vs. retake) as covariate. The categorical variables were converted using dummy coding; the reference groups (Table 2.2) were selected by several criteria. First, the group had to be well defined, and preferably be at the upper or lower boundary and contain a sufficient number of cases (Hardy, 1993). We also pre-analyzed if the selection of the reference group would leave out any interesting significant main effects and interaction effects in the regression model. These criteria led to a selection of reference groups that assured that no significant effects missed in the final regression model. The continuous variables, i.e. *Compartments* and *Arrows*, were centered around their means to reduce problems with multicollinearity (Frazier, Tix, & Barron, 2004).

The interaction terms were created by multiplying the explanatory variables. The variables entered the analysis in two steps. The first step included the covariate *Exam* and the main effect *Cognitive task demand* in the first block and all other main effects in the second block of the model (stepwise method). Only the significant main effects were kept in the regression model. The interaction between the *Cognitive task demand* and the other explanatory variables was first inspected visually by plotting the simple slopes of each combination. In the second step the most promising interaction terms and the accompanying main effects were added to the model one by one using the enter method, keeping only the significant interactions effects and their accompanying main effects in the final model. Significant interaction effects were analyzed post hoc by a *t*-test for simple slopes as suggested by Aiken and West (1991).

4. RESULTS

The results of the hierarchical multiple regression analysis with task difficulty as the dependent variable are summarized in Table 2.3. Model 1 includes the main effect from the variable *Cognitive task demand*. This variable explains 30% of the variance, $F(1, 62) = 14.18$, $p < .001$. In the second block of the analysis the significant main effects, *Familiarity diagram components* and *Components*, and the accompanying main effects of the significant interaction effects, *Familiarity diagram arrows* and *Prior content knowledge*, were added to the model. Model 2 explains 37% of the variance.

The significant interaction effect, *Cognitive task demand x Prior content knowledge*, was added to the third block of the analysis. Model 3 explains 43% of the variance, $F(6, 57) = 7.68$, $p < .001$. In the fourth block of the regression analysis the interaction effect *Cognitive task demand x Familiarity diagram arrows* was added to the final model which now explains 46% of the variance, $F(7, 56) = 7.59$, $p < 0.001$.

Table 2.3. Model summary of the multiple regression analysis with task difficulty (cohort mean score) as the dependent variable

Model	Variable(s) included	Adjusted R^2	F	p
1	Cognitive task demand	.295	14.184	< .001
2	Familiarity diagram components, Components, Familiarity diagram arrows, Prior content knowledge	.373	7.234	< .001
3	Cognitive task demand x Prior content knowledge	.426	7.682	< .001
4	Cognitive task demand x Familiarity diagram arrows	.456	7.589	< .001

The final regression model is presented in Table 2.4. The covariate *Exam* was not a significant predictor (t = 1.39, p = .17). The variable *Cognitive task demand*, with 'low' as the reference group, was found to be a significant predictor (B = -.175, p = .003) of task difficulty. This means that tasks with a 'high' cognitive demand showed lower mean exam scores and can be considered to be more difficult, compared to tasks with a 'low' cognitive demand.

Table 2.4. Final model of hierarchical regression analysis for variables predicting task difficulty (cohort mean score) on diagrammatic tasks (N = 64)

Variable	B	$SE\ B$	t	p
(Constant)	.578	.044	13.042	.000
Cognitive task demand	-.175	.056	-3.115	.003
Exam	.063	.045	1.391	.170
Familiarity diagram components	-.140	.052	-2.684	.010
Components	.009	.004	2.104	.040
Familiarity diagram arrows	-.063	.103	-.612	.543
Prior content knowledge	.094	.062	1.499	.140
Cognitive task demand x Prior content knowledge	-.248	.095	-2.617	.011
Cognitive task demand x Familiarity diagram arrows	.291	.145	2.008	.050

Note. Components and Arrows are centered at their means. Only unstandardized B is reported because β is not properly standardized in equations that include interaction terms and are thus not interpretable (Frazier et al., 2004).

The explanatory variables *Familiarity diagram components* (B = -.140, p = .010) and *Components* (B = .009, p = .040) also were significant predictors. Note that 'unfamiliar' is the reference group for the variable *Familiarity diagram components* and that *Components* is centered to its mean score. This means that tasks about diagrams with 'unfamiliar' components showed lower mean exam scores and can be considered to be more difficult, compared to tasks about diagrams with 'familiar' components. The positive effect for the number of *Components* showed that the more *Components* a diagram had, the higher the mean exam scores. This indicates that the task was less difficult if there were more *Components*.

The upper panel of Figure 2.2 presents the simple slopes for the significant interaction effect *Cognitive task demand x Prior content knowledge*, (B = -.248, p = .011). Post hoc probing of the interaction by a t-test for simple slopes shows that the interaction between a 'high' *Cognitive task demand* and *Prior content knowledge* is significant, $t(16) = 2.14$, $p < .05$. The interaction between a 'low' *Cognitive task demand* and *Prior content knowledge* is not significant, $t(40) = 1.32$, $p > .05$. This means that the combination of a task with a 'high' cognitive demand and 'low' *Prior content knowledge* showed a lower mean exam score, indicating that the task was more difficult than a combination of a task with a 'high' cognitive demand and 'high' *Prior content knowledge*.

The interaction effect *Cognitive task demand x Familiarity diagram arrows*, (B = .291, p = .050) was added to the regression model in the fourth step. Figure 2.2 (lower panel) presents the simple slopes for this interaction effect. Post hoc probing of the interaction by a t-test for simple slopes shows that the interaction between a 'high' *Cognitive task demand* and *Familiarity diagram arrows* is significant, $t(16)$ = 2.15, $p < .05$. This means that the combination of a 'high' *Cognitive task demand* and an 'unfamiliar' arrow type showed higher mean exam scores than the combination of a 'high' *Cognitive task demand* and a 'familiar' arrow type: This indicates, surprisingly, that tasks with a 'high' cognitive demand were less difficult with an 'unfamiliar' arrow type than with a 'familiar' arrow type. The interaction between tasks with a 'low' cognitive demand and *Familiarity diagram arrows* is not significant, $t(40) = .69$, $p > .05$.

5. DISCUSSION

The aim of this study was to examine variables that could explain the difficulty of a diagrammatic task. In this study, the cohorts' mean exam score on a diagrammatic task was used as an indicator of task difficulty. It was hypothesized that features of the task and student explained differences in task difficulty; no hypothesis was formulated for diagram features. We also hypothesized that difficulty of a diagram task can be explained by the interaction of cognitive task demand on the one hand and student features and diagram features on the other. In the result section, we reported significant main and interaction effects for features of the task, student, and diagram.

The *Cognitive task demand* explained most of the variance in task difficulty. The familiarity of the components was also related to task difficulty. Diagrams that contain 'familiar' components showed higher mean exam scores and were therefore seen as less difficult. We found no interaction effect between *Cognitive task demand* and *Familiarity diagram components*. This means that the effect of 'familiar' components on task difficulty is not different for tasks with a 'low' and a 'high' cognitive demand. This result seems to be consistent with research by Winn and Sutherland (1989) and Winn et al. (1991) who also focused on familiarity of components. Their experiments were performed in a controlled setting in which students got a limited amount of time to complete the tasks, with a limited variation in diagram presentation.

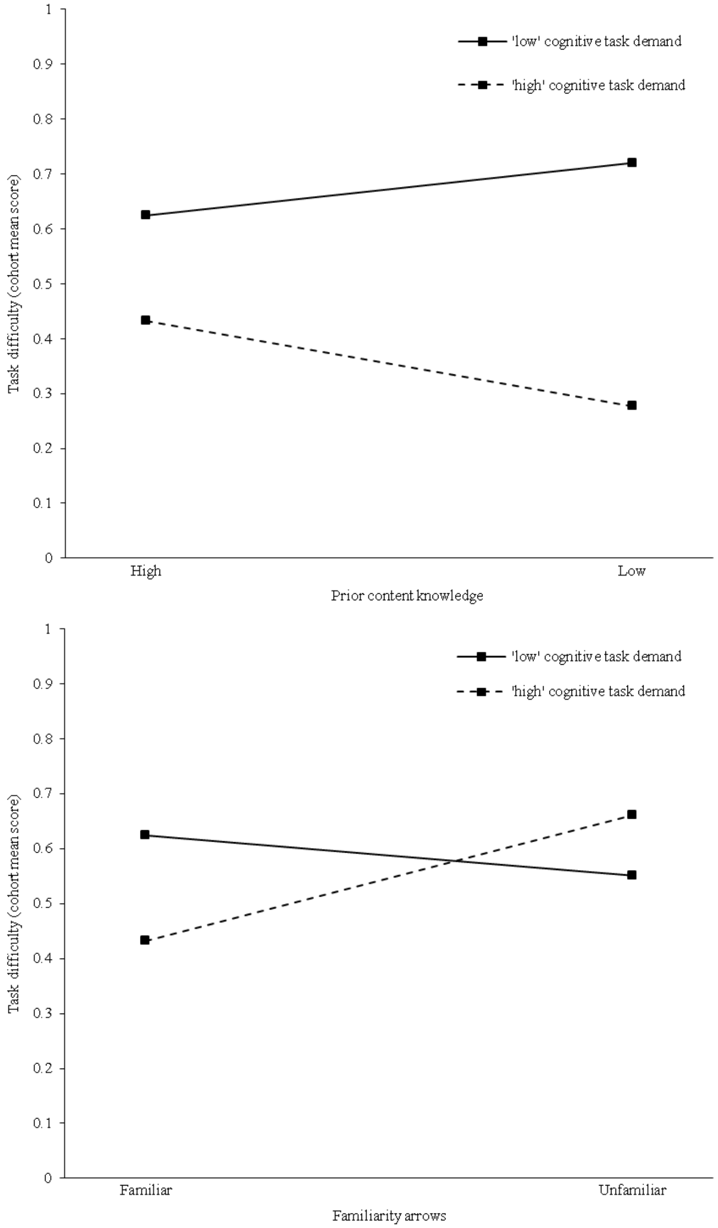

Figure 2.2. Plot of the significant interaction effect between Cognitive task demand and Prior content knowledge (top panel) and Cognitive task demand and Familiarity diagram arrows. The presented lines are the simple slopes and were calculated as suggested by Aiken and West (1991).

Moreover, they used memorization and response latencies as indicators for performance. The present study however demonstrated that in an ecological valid setting familiarity is also an important design issue for various diagram tasks. This conclusion is in line with the findings of Carlson et al. (2003) who suggest that diagrams, from a cognitive load perspective, facilitate understanding when they activate familiar schemas of the learner.

The number of *Components* was negatively related to the task difficulty; the more *Components* in a diagram the less difficult a task was. This suggests that adding more *Components* to a diagram makes it easier to interpret. This might be because the information in a diagram becomes more concrete if more *Components*, i.e., bits of information, are added: The diagram becomes less ambiguous and comprehension errors are less likely.

There was no interaction effect of *Cognitive task demand* and the number of *Components*, suggesting that this effect was not different for tasks with a 'low' and a 'high' cognitive demand. From a cognitive load (Sweller, 1994) point of view it might be expected that an increase in the number of *Components* might be more difficult with a 'high' *Cognitive task demand*, i.e., when the entire diagram needs to be examined. Not observing this effect in this corpus might be due to the fact that most diagrams are 'familiar' and that no real new information had to be learned. Note that this corpus is taken from central compulsory exams, in testing practice it is unusual to include learning of new items. There might be a third-order effect between *Cognitive task demand*, number of *Components* and *Prior content knowledge* but there was not enough data available to confirm this.

Post hoc analysis of the interaction between *Cognitive task demand* and *Prior content knowledge* revealed a significant difference in task difficulty for tasks with a 'high' cognitive demand and *Prior content knowledge* levels. There was no main effect for *Prior content knowledge* and there was no interaction between *Prior content knowledge* and a 'low' *Cognitive task demand*. This latter result is somewhat different as the findings by Lowe (1992) who reported that meteorologists selected more relevant information on weather maps than non-meteorologists. The introduction of a diagram about a new biological topic to students with a fair amount of biological background is probably not similar to presenting abstract and domain specific weather maps to non-meteorologists. It is suggested here that at least some of the students' knowledge about conventions in process diagrams which facilitate an analysis of the diagram might be easily transferred to biological diagrams about unknown topics.

The interaction effect between tasks with a 'high' cognitive demand and *Prior content knowledge* seems to confirm the conclusions of Winn (1993) that the absence of well-constructed knowledge schemata makes interpretation of a diagram more difficult. This finding raises some concerns for learning new scientific topics by diagram representation, especially when considering that images become increasingly important as carriers of meaning (Bezemer & Kress, 2008).

Post hoc analysis of the interaction between *Cognitive task demand* and *Familiarity diagram arrows* revealed a positive effect on task difficulty for tasks with a 'high' cognitive demand when the diagram has an 'unfamiliar' arrow type. This effect might seem a bit peculiar at first glance. But a closer look showed that two of the three diagrams with an 'unfamiliar' arrow type had many more *Components* (i.e., 23 and 24 components) than average (M = 9.27, SD = 5.61). This increase in concreteness by these number of *Components* might compensate for the 'unfamiliarity' of the arrows. No other main effects or interaction effects (i.e., besides the number of *Components*) where found with the explanatory variables which described the diagrams' features. This obviously does not imply that previous research on the effect of diagram features like *Component type* (Holliday et al., 1977), *Clustering* (Winn, 1991), *Process labels* (Mayer & Gallini, 1990), etc., is disapproved. It merely suggests that the heterogeneous composition of the diagrams in our dataset probably makes it more difficult to find any significant effect as many factors might interact. It must also be mentioned that many of the other studies were about diagrams which might were not similar to biological process diagrams, e.g. a bike pump and a car brake (Mayer & Gallini, 1990), a family tree diagram (Winn et al., 1991), or a weather map (Lowe, 1996; Canham & Hegarty, 2010).

The results from the present study provide relevant insights into the design of a training of students in the use of process diagrams in Biology. We conclude that a training program should: (1) include strategies for encoding diagrams with unfamiliar components, (2) focus on the interpretation of abstract (i.e., highly conceptual with a minimal amount of context information) diagrams, and (3) facilitate students in learning how to gain a deeper understanding of diagrams that contain new information. We suggest that a training on meta-cognition which involves self-explaining (Bielaczyc, Pirolli, & Brown, 1995) and thinking about diagrams on a meta level, e.g., what conventions are used, how diagrams can be constructed, what information (the designer of) the diagram is trying to communicate, etc., might be an effective approach. Another approach would be to show students several videos of novices who interpret diagrams successfully or unsuccessfully as a form of modelling. The students observe and have to compare and evaluate task behaviour; this kind of training has been done earlier within various domains, e.g., written composition (Raedts, Rijlaarsdam, Van Waes, & Daems, 2007), and has proven to be effective.

Finally, this study adds to previous research on the role of task demand (e.g., Guthrie et al., 1993), prior knowledge (e.g., Lowe, 1996; Cook et al., 2008), familiarity with conventions (e.g., Carlson et al., 2003), and design features (e.g., Mayer & Gallini, 1993; Canham & Hegarty, 2010) in learning from external representations. The present study reveals what factors contribute to the difficulty of the tasks, but not how these factors interact during processes of understanding. We still need a more thorough understanding of which problems are urgent and relevant when improving students diagram interpretation skills. Foci of further research on the interpretation of process diagram could be, for example, differences in students' search efficiency and solving strategies and the role of spatial ability.